

What is the “ADAM OTHER” Class of Datasets, and When Should it be Used?

John Troxell, Accenture Accelerated R&D Services, Berwyn, PA

As is well known by now, the CDISC ADaM team has defined four classes of ADaM datasets: ADSL (Subject-Level Analysis Dataset), BDS (Basic Data Structure), ADAE (Adverse Events Analysis Dataset), and OTHER. The ADAE class is soon to be generalized into the OCCDS (Occurrence Data Structure) class. The ADSL, BDS and ADAE/OCCDS structures are defined and widely used. However, the OTHER class is by nature relatively unspecified and mysterious. This paper explores philosophical and practical questions surrounding the OTHER class of ADaM datasets, including considerations for deciding when its use is appropriate, and when it is not.

INTRODUCTION

The ADaM standard has defined some standard data structures, yet it also permits datasets that do not have a standard structure. And of course, there exist analysis datasets that are not ADaM datasets at all. How do we make sense of all of this? And what is the right thing to do in a given situation?

To address these questions, this paper starts by reviewing the ADaM fundamental principles, and then the standard ADaM dataset classes and structures that embody the principles. The ADaM OTHER class of datasets is described, and contrasted to non-ADaM analysis datasets. The paper then addresses the question of when the ADaM OTHER class should be used, and when it should not be used.

This paper contains opinions of the author.

CDISC ADAM

The Clinical Data Interchange Standards Consortium (CDISC) Analysis Data Model (ADaM) is a model for observed and derived subject data enabling analysis and supporting review. ADaM has always been influenced by the needs of FDA statistical and medical reviewers, and is preferred by FDA. ADaM will be required by FDA for new studies no earlier than Dec. 18, 2016. ADaM has fundamental principles, standard data and metadata structures, standard variables, dataset and variable naming conventions, rules, etc.

ADAM FUNDAMENTAL PRINCIPLES

The first step in the evolution of the ADaM standard was the development of a philosophy represented in the ADaM fundamental principles. As described in the CDISC Analysis Data Model document, the fundamental principles are:

- Analysis datasets and their associated metadata must:
 - facilitate clear and unambiguous communication
 - provide traceability between the analysis data and its source data (ultimately SDTM)
 - be readily useable by commonly available software tools
- Analysis datasets must:
 - be accompanied by metadata
 - be analysis-ready.

FUNDAMENTAL PRINCIPLES HAVE PRACTICAL EFFECTS - A FEW EXAMPLES

The fundamental principles govern not only how one should approach the design and development of ADaM datasets and metadata in one's work. They also influence and are reflected in various aspects of the standard and its data structures. For example, “facilitate clear and unambiguous communication” is reflected in the requirement to include the date imputation flag when a date is imputed. It is also reflected in the rule that one should keep all the observed and derived records, not just those that are analyzed. Similarly, “provide traceability between the analysis data and its source data” is reflected in advice to include traceability-facilitating variables, e.g. SDTM --SEQ, when feasible. (Of course metadata traceability is always required as well). Another example is that the principle that analysis datasets “must be analysis-ready” means that virtually all data derivation between tabulation and analysis results is materialized in analysis datasets rather than performed on the fly during analysis results generation.

FOUR CLASSES OF ADAM DATASETS – BUT ONLY THREE STANDARD DATA STRUCTURES

There are four classes of ADaM datasets, but only three standard ADaM dataset structures:

1. Subject-Level Analysis Dataset (ADSL). The only member of this class is the dataset ADSL, whose standard structure is one record per subject.
2. Basic Data Structure (BDS). The BDS is a standard vertical structure that can handle so many situations and analyses that one might refer to it as the Swiss Army Knife of ADaM dataset structures. There are normally several BDS class datasets in a study.
3. Occurrence Data Structure (OCCDS). This soon-to-be-released standard structure is a generalization of the ADAE structure for adverse events, extending it to handle additional kinds of occurrences data such as data about medical history and concomitant medications. There may be several datasets of the OCCDS class in a study.
4. Other (OTHER). By definition, this class has no standard structure!
Note: the OTHER class is called ADAM OTHER in the CDISC controlled terminology for CLASS.

THE ADAM OTHER CLASS

In the author’s opinion, the purpose of the OTHER class is to support analyses that cannot be supported by the standard ADaM dataset structures that have been defined to date; so far, they are ADSL, BDS, and OCCDS.

Datasets of the OTHER class must:

- Follow the ADaM Fundamental Principles;
- Follow all ADaM dataset and variable naming conventions, etc. as much as possible.

Over time, the OTHER class shrinks as the CDISC ADaM Team defines new classes and structures. E.g., after ADSL, there were BDS (2009), ADTTE in BDS (2012), ADAE (2012), and OCCDS (2015).

However, the OTHER class is theoretically unlimited and grows each time someone creates an ADaM dataset to support an analysis that is not yet addressed by a standard ADaM structure.

NON-ADAM ANALYSIS DATASETS

A non-ADaM analysis dataset is simply an analysis dataset that does not follow the ADaM fundamental principles or does not follow all ADaM dataset and variable naming conventions, etc., as much as possible. Typical examples are legacy analysis datasets, or datasets that follow a proprietary standard.

Next, let’s look at how all of these kinds of datasets relate.

ADAM OTHER IN THE UNIVERSE OF ANALYSIS DATASETS

Figure 1 shows where ADaM datasets of the OTHER class fit within the universe of analysis datasets.

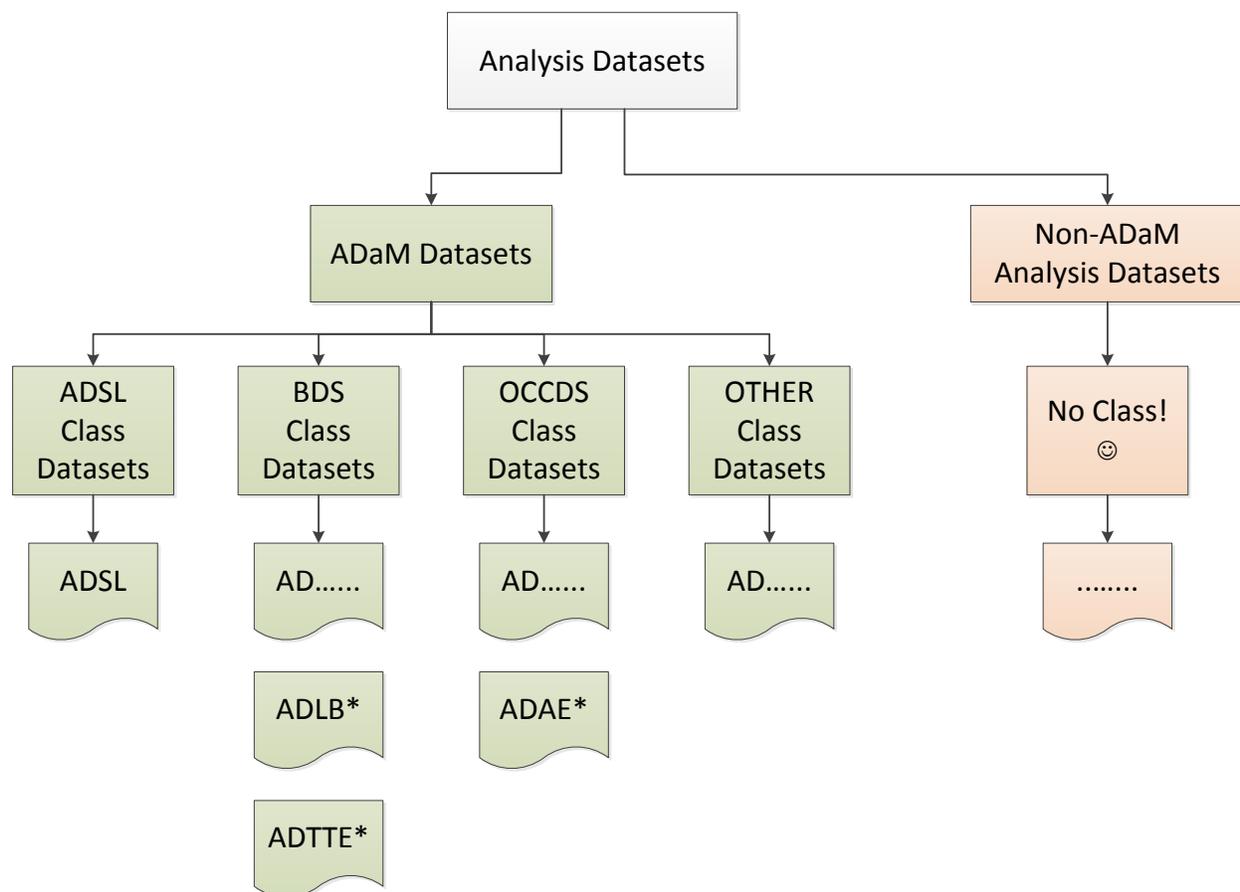


Figure 1. ADaM OTHER in the Universe of Analysis Datasets

In the figure, dots in a dataset name represent characters that are specified in an actual dataset name. An asterisk after a dataset name means that it is an example name of a dataset of the class.

WHAT HAVE WE SEEN SO FAR?

- Not all analysis datasets are ADaM datasets.
- “ADaM datasets” is not a synonym of “analysis datasets”.
- Non-ADaM analysis datasets are not ADaM datasets of the OTHER class.
- There are standard ADaM data structures (ADSL, BDS, and OCCDS, so far).
- The ADaM OTHER class does not have a standard structure.

CAN I USE THE OTHER CLASS TO DO WHATEVER I LIKE?

Question: “I really don’t like the BDS vertical data structure. I prefer a horizontal structure. Can I make a horizontal dataset like I am used to and call it an ADaM OTHER class dataset?”

Short Answer: “Probably not!”

Long Answer:

The author would answer as follows. ADaM is a data standard. Therefore, a standard data structure should be used if it is capable of supporting the analysis. If it were optional to use the standard data structure, then ADaM would not be a data standard.

FDA wants sponsors to use the ADaM data standard, and will soon require them to use it.

In the author’s opinion, when a BDS dataset can support the analysis, a BDS dataset should be used. Similarly for ADSL and OCCDS. The purpose of the OTHER class is to support analyses that cannot be supported with a standard ADaM data structure. The OTHER class is not a loophole to drive a truck through.

Use an ADaM standard data structure when a standard data structure is capable of enabling the analysis.

Question: “Darn. Well then, when can I use OTHER?”

Let’s look at four situations where it may make sense. There may be others.

FOUR SITUATIONS WHERE A DATASET OF THE OTHER CLASS MAY BE NEEDED (THERE MAY BE MORE)

1. NO STANDARD STRUCTURE CAN ENABLE THE ANALYSIS

If ADSL, BDS and OCCDS cannot support the analysis, then the OTHER class can be used. Of course, one must follow ADaM fundamental principles, naming conventions, etc.

An example of an analysis that cannot be supported by a standard ADaM data structure is Multivariate Analysis of Variance (MANOVA). More than one dependent variable is needed on the same row.

Another example is a correlation matrix. The variables to be correlated must be on the same row.

Note: A horizontal OTHER Class dataset can often be created by transposing a BDS dataset, using BDS PARAMCD for the analysis variable names (but this is not always as simple as it may at first sound due to data irregularities etc.).

However, it must be said that hypothesis testing involving true statistical multivariate analysis is very rare in late-stage clinical trials. Virtually all hypothesis testing in clinical trials involves univariate analysis (with e.g., one analysis variable on the left hand side of the model statement).

The BDS works great for univariate analyses (e.g., ANCOVA, Cochran-Mantel-Haenszel, most Time-to-Event analyses, and many others). The BDS does support inclusion of all explanatory variables, such as treatment, covariates, etc.

2. “LISTING” VIEW

Often, someone wants or needs to see analysis variables or values across timepoints side-by-side. This can be a perfectly legitimate need of statistician, clinician, or reviewer. An ADaM dataset of the OTHER class, or even a non-ADaM dataset, can be created to support eyeball review of data.

If such a dataset follows ADaM fundamental principles etc., it can be a member of the ADaM OTHER class.

However, in the author’s opinion, such a dataset should not be used as the input dataset to the statistical analysis if ADSL, BDS, or OCCDS would suffice.

3. PRE-ADSL DATASET

It is sometimes helpful to create a dataset in which are assembled facts to be used when deriving ADSL variables, for example, the per-protocol population flag PPROTFLL. One might for example need to ascertain or derive:

- Was the subject compliant enough with the prescribed therapy?
- Did the subject take certain disqualifying concomitant medications?
- Did the subject violate dietary restrictions?

This pre-ADSL dataset could be almost exactly like a BDS dataset, with a parameter PARAM for each of these facts; but if so it probably won’t be BDS-compliant, because for example, it probably won’t have planned treatment, and certainly it won’t have variables copied from ADSL.

If it follows ADaM fundamental principles etc., it can be a member of ADaM OTHER class. In the author’s opinion, a pre-ADSL dataset shouldn’t be used for analysis. If these facts are needed for analysis, create ADSL variables to contain them based on the pre-ADSL dataset.

4. INTERMEDIATE DATASET

An intermediate dataset is a dataset that is created based on precursor data that is itself a precursor to another dataset. Intermediate datasets can often be of the BDS class. An example of an intermediate dataset that does not follow one of the ADaM standard data structures is an EX-plus dataset that might be needed for statistician eyeball review (see case 2, “Listing View”, above). This dataset might be comprised of

- SDTM EX
- Plus some SDTM EX supplemental qualifiers pivoted onto the dataset as columns
- Plus some columns copied from ADSL
- Plus some derived variables

If it follows ADaM fundamental principles etc., it can be an ADaM OTHER class dataset.

Note: Renaming EXTRT to PARAM and some other variable to AVAL does not make this a BDS dataset! In the author’s opinion, this dataset should not be used for the statistical analysis.

A compliant BDS dataset may be derived from this intermediate dataset, to contain parameters summarizing aspects of exposure for analysis.

INVALID REASONS FOR OTHER CLASS DATASETS

In order for ADaM to function as a true data standard, the author believes that ADaM standard data structures should be used when they are capable of supporting an analysis. The ADaM OTHER class is not some kind of giant loophole in the ADaM data standard. Invalid reasons for creating an ADaM OTHER class dataset include:

- It’s going to cost us a lot to change to the standard ADaM structures
- I want to do it a different way.

CONCLUSION

CDISC ADaM has rules, but can still express the science. There is plenty of room for expression within ADSL, BDS, and OCCDS. Designing ADaM datasets is a creative process.

The ADaM OTHER class is available for use when there is a need that can’t be met by the standard ADaM dataset structures. Don’t be afraid of it, but use it only when truly necessary.

ACKNOWLEDGMENTS

The author would like to thank Sandra Minjoe, his colleague on the CDISC ADaM Team and at Accenture, for suggesting the general topic of this paper and for her comments.

RECOMMENDED READING

- CDISC ADaM standard documentation. Available at www.cdisc.org/adam

CONTACT INFORMATION

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

Name: John Troxell
Enterprise: Accenture Accelerated R&D Services
E-mail: john.troxell@accenture.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.