PharmaSUG 2025 - Paper DS-126

# ADaM Fundamental Principles vs. Rules: Which to Follow, When, and Why?

Sandra Minjoe, ICON PLC;
Mario Widel, IQVIA

## ABSTRACT

The two of us authors have been actively involved in the CDISC ADaM team a long time – one for more than 20 years, and the other for over 15 years. During that time, we have led and contributed to a variety of ADaM documents, and we are both authorized CDISC ADaM trainers.

In our years of working with ADaM, we've seen many different interpretations of ADaM documents – some interpretations are correct, others are not. This paper and presentation aim to address the confusion we have seen on ADaM Fundamental Principles vs. Rules, describe when each are needed, propose a hierarchy, and recommend how to handle compliance. We provide examples, focusing on tricky cases where there can be multiple correct options for implementation.

## INTRODUCTION

There are many documents that comprise ADaM:

- Analysis Data Model v2.1, released in 2009, and referred to here as "the model document". It "specifies the fundamental principles and standards to follow in the creation of analysis datasets and associated metadata"
- ADaM Implementation Guide (IG), has versions 1.0, 1.1, 1.2, and 1.3. It "specifies ADaM standard dataset structures and variables, including naming conventions. It also specifies standard solutions to implementation issues." ADaMIG includes standard structures ADSL and BDS.
- ADaM Structure Occurrence Data Structure (OCCDS v1.0) and ADaM Structure for Occurrence Data Implementation Guide (OCCDS v1.1) describe "the general data structure and content typically found in occurrence analysis. Occurrence analysis is the counting of subjects with a given record or term, and often includes a structured hierarchy of dictionary coding categories." OCCDS v1.0 and 1.1 supplement the ADaMIG, adding a third standard structure, OCCDS.
- CDISC ADaM Conformance Rules v4.0 and Guide use the content described in the model document, ADaMIG documents, OCCDS documents, and a few other documents not described above (such as the ADaMIG for Medical Devices), to create automatable compliance checks. All checks cite at least one of the contributing documents for the context of the rules.

Scattered throughout the model document, ADaMIGs, and OCCDS documents are rules and requirements to guide in data creation and validation. Fundamental principles are described in both the model document and the ADaMIGs.

## WHAT ARE THE ADAM FUNDAMENTAL PRINCIPLES?

The ADaM Fundamental Principles, shown in Figure 1, are described in the ADaM model document (v2.1) Section 3. Notice the word "must" is included. These are non-negotiable.

We can think of the fundamental principles are belonging to 2 categories: making your data useable, and making it understandable.

Let's consider the following questions:

- How is the data used to create analysis output?
- Where did data come from or how it was derived?

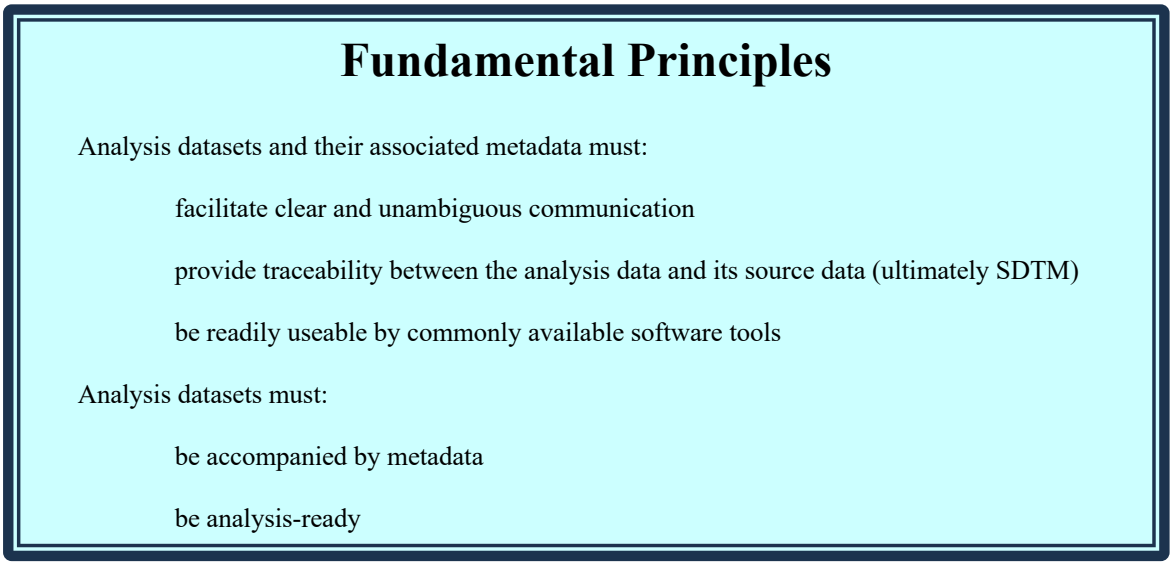


**Figure 1: Fundamental Principles of ADaM**

## MAKING YOUR DATA USABLE

Two of the five ADaM fundamental principles, shown in Figure 1, are about useability:

First, analysis datasets and their associated metadata must be readily useable by commonly available software tools. But what are "commonly available software tools"? Well, we know we must submit data to regulatory authorities in the SAS® v5 transport file format. That's because this version of the SAS transport file is in the public domain, meaning tools other than SAS (e.g., R, S-Plus, JMP) are able to use them. So by putting your data into SAS transport files, you've met the requirement of making it useable by commonly available software tools.

Next, analysis datasets must be analysis-ready. In the ADaM model (v2.1) it describes analysis-ready as "analysis datasets that have a structure and content that allows statistical analysis to be performed with minimal programming." It goes on to state that "No complex data manipulations such as transformations or transpositions are required to perform the supported analysis.", and that "This approach eliminates or greatly reduces the amount of programming required by analysts such as statistical reviewers." The point here is to do the bulk of the data preparation when creating the dataset, so that it is easy to use that dataset to create the analysis output.

Sometimes people interpret "analysis-ready" to mean "one proc away", but that isn't always the case. For example, a typical Adverse Event table has multiple levels of hierarchy reported, where you likely need one procedure to generate the body system counts, another procedure to generate the preferred term counts, and yet another procedure to derive the percentages. While this table needs multiple procedures, the ADaM analysis dataset needs to be as ready as it can be to plug into those procedures.

## MAKING YOUR DATA UNDERSTANDABLE

Three of the five ADaM fundamental principles, shown in Figure 1, are about making your data understandable.

First, analysis datasets and their associated metadata must facilitate clear and unambiguous communication. The point here is to make it easy for people to understand what you did. For example, metadata is often useful as a specification even before a dataset is written, and we often must double program our datasets from that specification to ensure quality. When two different programmers using those specifications are able to come up with the same results, then those specifications were clear and unambiguous. Now think about it from a regulatory agency reviewer perspective: is that data and metadata clear enough that I can understand what you did and recreate those steps myself? Because, at

the end of the day, when we submit data to an agency, we want them to be able to work with it as easily as possible and with as few questions as possible, in order to avoid delays in getting the product to market.

Next, analysis datasets and their associated metadata must provide traceability between the analysis data and its source data (ultimately SDTM). For example, if I see on a summary table that 3 subjects had an extreme result on one lab test, I might want to investigate who they were and what else might be going on. First, I can search through the ADaM data to find all the records with those extreme results. Typically, ADaM datasets don't contain all the variables from SDTM, but instead have the ones pertinent to the analysis. However, when those ADaM records have something as simple as the --SEQ variable included, I can then look in the source SDTM dataset for that sequence number and see all the rest of the information that wasn't brought over from SDTM into ADaM. Like the "facilitates clear and unambiguous communication" fundamental principle, traceability helps a reviewer find answers to their questions without having to slow down the review to ask the sponsor.

Last, analysis datasets must be accompanied by metadata. Have you ever inherited data with no documentation? Without knowing where something came from or how it was derived, that data is essentially useless. Metadata provides that context. Define-XML forces us to include content like origin and derivation methods, ensuring that the user of the data has what they need to understand what they're looking at.

### WHAT IS NOT INCLUDED IN THE FUNDAMENTAL PRINCIPLES

Looking back at Figure 1, there is no mention of standard ADaM content such as dataset structures and variable names. Although these are important features of ADaM, they do not rise to the level of fundamental principles. An ADaM dataset can be in a structure other than one of the defined standard ADaM classes. An ADaM variable can be named something other than one of the defined standard variable names.

The ADaM standard dataset structures and variable names were built on and designed to support the ADaM fundamental principles. Having standard structures like ADSL, BDS, and OCCDS, and standard variable name like PARAM and AVAL allow people to easily find relevant content, which makes the data useable and understandable. However, ADaM allows for lots of non-standard options, when the standard ones don't meet your analysis needs, and the fundamental principles ensure that even non-standard content is still ADaM.

## WHAT ARE THE ADAM RULES?

The ADaMIG and OCCDS documents contain many rules that must be followed to be considered ADaM. The ADaM Conformance Rules v4.0 took the text from the different versions of these two documents to develop about 700 official conformance rules, and it contains over 1900 rows. The first few rows are shown in Figure 2. Notice that the same rule can be repeated, once for each version of the cited reference where it applies.

Each row in Figure 2 describes what data it applies to and includes citations to the actual ADaM documents. For example, looking at the 3rd column in Figure 2, each rule often applies to only certain ADaM structures.

Although lengthy, what is included in this list of rules is really just a subset of all the possible ADaM rules: it contains only rules that can be programmatically checked. As the CDISC ADaM Conformance Rules Guide states, "The conformance rules are machine-readable (i.e., programmable within computer software) and capable of being implemented by ADaM users." So while machine-testable rules are very helpful, they are not sufficient for determining whether data is ADaM-compliant. This is why manual review is so important and must be done in addition to automated review.

| Check Numbe | IG Versio | ADaM Structure Group | Message Type | Guide | Section | Item | Cited Guidance |
|---|---|---|---|---|---|---|---|
| 1 | 1.0 | ADSL | Error | Model v2.1; ADaM IG v1.0 | 6; 2.3.1 | | Model v2.1, Section 6: ADSL and its related metadata are required in a CDISC-bas trial even if no other analysis datasets are submitted.<br><br>ADaM IG v1.0, Section 2.3.1: ADSL and its related metadata are required in a CDI clinical trial even if no other analysis datasets are submitted. |
| 1 | 1.1 | ADSL | Error | Model v2.1; ADaM IG v1.1 | 6; 2.3.1 | | Model v2.1, Section 6: ADSL and its related metadata are required in a CDISC-bas trial even if no other analysis datasets are submitted.<br><br>ADaM IG v1.1, Section 2.3.1: ADSL and its related metadata are required in a CDI clinical trial even if no other ADaM datasets are submitted. |
| 1 | 1.2 | ADSL | Error | Model v2.1; ADaM IG v1.2 | 6; 2.3.1 | | Model v2.1, Section 6: ADSL and its related metadata are required in a CDISC-bas trial even if no other analysis datasets are submitted.<br><br>ADaM IG v1.2, Section 2.3.1: ADSL contains one record per subject, regardless of of the ADSL dataset is "Subject-Level Analysis Dataset." In a study, there is only LEVEL ANALYSIS DATASET", and its name is ADSL. |
| 1 | 1.3 | ADSL | Error | Model v2.1; ADaM IG v1.3 | 6; 2.3.1 | | Model v2.1, Section 6: ADSL and its related metadata are required in a CDISC-bas trial even if no other analysis datasets are submitted.<br><br>ADaM IG v1.3, Section 2.3.1: ADSL contains one record per subject, regardless of of the ADSL dataset is "Subject-Level Analysis Dataset." In a study, there is only LEVEL ANALYSIS DATASET", and its name is ADSL. |
| 2 | 1.0 | ALL:SDTM | Error | Model v2.1; ADaM IG v1.0 | 4.1.2; 3 | 4 (General Variable Naming Conventions) | Model v2.1, Section 4.1.2: Any ADaM variable with the same name as an SDTM v SDTM variable, and its label, attributes, and values cannot be modified. ADaM a known as "same name, same meaning, and same values." |

**Figure 2: ADaM Conformance Rules 4.0**

## HIERARCHY OF PRINCIPLES AND RULES

In trying to determine which is a higher priority, let's consider two different questions:

1. Can a dataset be called ADaM compliant if it meets all the fundamental principles of ADaM but breaks a conformance rule?
2. Can a dataset be called ADaM compliant if it meets all the conformance rules but breaks an ADaM fundamental principle?

Considering the first question, what might a dataset look like that fails an ADaM fundamental principle? Well, it might be unclear, not include traceability, not be in format (like a SAS v5 transport file) that your tool can use, not have accompanying metadata, or require a lot of processing to get to the analysis results. Is your data ADaM if it isn't useable or understandable?

What about a dataset that breaks one of the 700 conformance rules? Consider the following examples:

1. You might have a dataset where you transposed some BDS data to get visits as columns, so that it can be easily used for your analyses. In that case, you might still have PARAM, but you wouldn't have an AVAL or AVALC. Here the data doesn't follow a BDS structure. Is that wrong? Of course not - that's what class ADaM Other is for! However, if the tool uses the fact that you have a variable called PARAM to imply it is of structure BDS, then it would run the BDS checks on it and complain that you are missing AVAL or AVALC.
2. You might have a good reason to break a rule. For example, some CDISC Therapeutic Area User Guides (TAUGs), including Prostate and Breast Cancer, and some FDA documents make use of variable PARQUAL, even though doing so currently breaks an ADaM BDS rule. Since you probably want to provide data in the structure that the reviewers prefer, as described in those TAUGs and FDA documents, in this case it seems reasonable to break the ADaM rule.
3. You might have "messy" data that breaks a rule. For example, a collected start date might be after an end date. If it is determined that this discrepancy will not be "fixed" in the data, then you have broken a rule.

So sometimes it is OK to fail conformance checks. This is precisely why the Analysis Data Reviewer's Guide (ADRG) has a section to document your conformance check findings and provide an explanation.

In other words, content that is missing any of the fundamental principles, even if you passed the ADaM content through a tool that tested for all the conformance checks and there were no errors found, would not be ADaM. However, meeting all the fundamental principles but failing a conformance check may not mean that the data isn't ADaM. This means that the fundamental principles are of a higher order than the conformance checks. It also means that breaking a rule to meet regulatory needs is OK, but breaking a fundamental principle is not.

## CHECKING FOR COMPLIANCE

### FUNDAMENTAL PRINCIPLE COMPLIANCE

Two of the principles, that there is associated metadata and that the data is provided in a form that is usable (for now that is SAS v5 transport files) can be automated, but there are no conformance checks that content is clear, includes sufficient traceability, or does not require a lot of processing to get to the analysis results. In fact, the CDISC ADaM Conformance Rules Guide states "thoroughness and clarity of metadata cannot be tested by a machine-readable algorithm but are necessary to enable the traceability that ADaM requires."

If we can't test by an automated rule, what can we do? We must perform these checks manually! Manual review is always required, in addition to automated review, and is especially necessary for the fundamental principles that content be clear, traceable, and analysis ready.

### COMPLIANCE AND SUBMISSION

For any issue found with automated checking, you must review and determine whether the issue should be fixed or explained. When a rule is broken that is intentional (for example, using PARQUAL to meet FDA needs), then it should be explained in the ADRG. When a rule is broken that is not intentional (for example, a typo in controlled terminology), then it needs to be fixed. Additionally, because fundamental principles cannot be broken, any issues found with these must be fixed. This is why checks should be run early and often, so you don't discover unintentional issues just before database lock (or, even worse, after lock), when you would feel pressure to explain what actually needs to be corrected.

It's worth noting that it can cause suspicion when data is too clean. It was mentioned in a conference presentation that it can be a red flag when a review agency receives data that passes all checks. It brings up questions like "Why is it so clean?" and "Did they modify the data just to avoid checks firing?" Because we work with real data, some imperfections are to be expected. Automated tool messages are not necessarily a bad thing, because it gives you an opportunity to explain something unique about your data.

## CONCLUSION

ADaM has both Fundamental Principles and Rules, each serving a different purpose. Fundamental Principles are the priority – they are "fundamental" after all – and cannot be broken. Rules are often based on dataset structure, and there are times when rules can be broken. Some rules, especially those which are tied to fundamental principles, cannot be broken. While automated tools can help with checking compliance of both rules and fundamental principles, manual review is also required.

## REFERENCES

**CDISC ADaM documents** are available at https://www.cdisc.org/standards/foundational/adam. Referenced in this paper are the following documents:

- Analysis Data Model v2.1
- ADaM Implementation Guide v1.0, 1.1, 1.2, and 1.3
- ADaM Structure Occurrence Data Structure v1.0
- ADaM Structure for Occurrence Data Implementation Guide v1.1
- CDISC ADaM Conformance Rules v4.0 and CDISC ADaM Conformance Rules Guide v4.0

**CDISC TAUGs** are available at https://www.cdisc.org/standards/therapeutic-areas/published-user-guides.

**FDA Study Data Standards Resources** documents are available at https://www.cdisc.org/standards/therapeutic-areas/published-user-guides.

**Analysis Data Reviewers Guide Package** is available at https://advance.hub.phuse.global/wiki/spaces/WEL/pages/26804660/Analysis+Data+Reviewer+s+Guide+ADRG+Package.

Related papers:

- Making an ADaM Dataset Analysis-Ready: https://www.lexjansen.com/pharmasug/2022/DS/PharmaSUG-2022-DS-095.pdf.
- Traceability: Some Thoughts and Examples for ADaM Needs: https://www.lexjansen.com/pharmasug/2018/DS/PharmaSUG-2018-DS01.pdf.
- When Should I Break ADaM Rules?: https://www.lexjansen.com/pharmasug/2021/DS/PharmaSUG-2021-DS-080.pdf.
- Why Are There So Many ADaM Documents, and How Do I Know Which to Use?: https://www.lexjansen.com/pharmasug/2020/SS/PharmaSUG-2020-SS-081.pdf.

## RECOMMENDED READING

**CDISC ADaM documents** are available at https://www.cdisc.org/standards/foundational/adam. Additional CDISC ADaM documents that might be useful but which are not referenced in this paper are:

- ADaM Examples of Traceability v1.0
- ADaM Metadata Submission Guidelines v1.0

## CONTACT INFORMATION

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

Sandra Minjoe
Sandra.Minjoe@iconplc.com

Mario Widel
MHWidel@gmail.com