**Paper SI12**

# ADaM Programming – The Good, the Bad and the Ugly

Pam Howard, ICON Ltd, Dublin, Ireland

## 1. ABSTRACT

**Delivering ADaM datasets right the first time can be challenging. However delays and errors in finalizing ADaM datasets are very costly in terms of customer satisfaction, quality and meeting timelines. This presentation will present an analysis done on sources of programming errors and a process improvement project. The strategy around resolving some of the common issues/root causes of ADaM programming deficiencies will be discussed along with examples. The roadmap of solutions will be outlined. The importance of ADaM standards & the governance of these changes to standards are also reviewed. Innovations around ADaM tools for both single studies and consistency across programs of studies will also be discussed.**

## 2. INTRODUCTION

CDISC is an organisation that focuses on data standards within the clinical trial arena. Foundational CDISC Standards are the basis of the complete suite of standards, supporting clinical and non-clinical research processes from end to end. Foundational Standards focus on the core principles for defining data standards and include models, domains and specifications for data representation. One of these foundational models is the Analysis Data Model (ADaM). The Analysis Data Model (ADaM) supports efficient generation, replication, and review of analysis results. The purpose of ADaM is to provide a framework that enables analysis of the data, while at the same time allowing reviewers and other recipients of the data to have a clear understanding of the data's lineage from collection to analysis to results. Whereas ADaM is optimized to support data derivation and analysis, CDISC's Study Data Tabulation Model (SDTM) is optimized to support data tabulation.

ICON's ADaM Centre of Excellence (iACE) is the umbrella strategy that focuses on ADaM delivery.The iACE approach to optimising the process around the production of SAS datasets is discussed in this paper.

## 3. WHY ADAM DATASETS?

ADaM datasets have increased in popularity especially due to the need for standardized datasets for regulatory submissions [(1)] implementing CDISC standards an industry best practice.

a. It is now an FDA requirement for NDAs, ANDAs, BLAs, and DMFs on studies that started after December 17, 2016. Sponsors must submit data in FDA-supported formats listed in the FDA Data Standards Catalog, which specifies the use of CDISC standards: SDTM, SEND, ADaM, Define-XML and Controlled Terminology.
b. However despite the regulatory requirements, approx.35% of studies that ICON is currently working on are not using ADaM standards. The reasons why some studies do not require ADaM datasets are as follows:
    - The study uses legacy client standards
    - Study of long duration that started some time ago
    - ICON is performing partial tasks such as DMC support that uses non-CDISC analysis data structures

## 4. WHO GENERATES ADAM DATASETS?

Looking at the three main ADaM datasets activities - specifications, generation, checking - the following roles are involved:

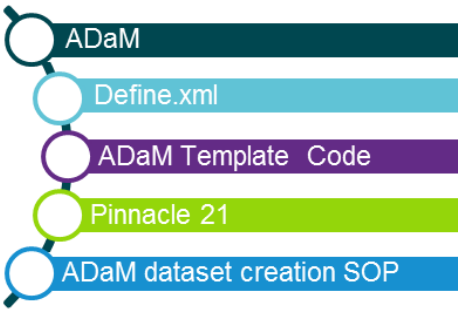
Table 1: ADaM datasets – roles and responsibilities

| | Specification | Development | Validation | Review |
|---|---|---|---|---|
| Project Statistician | ✓ | | | ✓ |
| Lead Programmer | ✓ | ✓ | ✓ | ✓ |
| Programming team | | ✓ | ✓ | |

Our analysis showed that approx. 85% of programmers were involved in ADaM dataset generation in the last 6 months.

In order to ensure that those performing the tasks have the knowledge, iACE has been offering foundational trainings that are mandatory for all programmers – these cover the main aspects of ADaM generation.

Table 2: Mandatory Foundational Trainings




The iACE subject matter experts contribute to and keep abreast of new standards & regulatory guidances, In addition to their training responsibilities; they are responsible for sharing and cascading emerging and newly released standards. In addition, colloquia, newsletters, and lunch & learn meetings focus on ADaM knowledge Example of topics covered would be 'Differences between ADaM IG versions 1.0 and 1.1', 'CFAST standards' etc.
ADaM dataset training is also a key component of the internal programming training schedule.

Table 3: A selection of ADaM related intern trainings


## 5. HOW MANY ADaM DATASETS ARE CREATED?

Measuring the number of unique ADaM datasets (# of ADaM datasets created once per study not counting any iterations or changes), between 600 to 900 datasets are produced per quarter.

ICON classifies ADaM datasets into simple safety and complex safety and efficacy datasets. **Complex analysis datasets** are those datasets with either complex structure requiring significant manipulation of input data and/or with large volumes of derivations e.g. exposure datasets for multi-drug dosing combinations, subject-level datasets with multiple baseline variables and/or subgroup definitions (such as ADSL). The number of complex datasets is estimated based on prior experience of the therapeutic area and/or dependent on the study design. The number of complex datasets may be revised (up or down) once the analysis dataset specifications have been reviewed and approved. As an example the following datasets may be defined as complex: ADSL, ADEX, ADLB, and efficacy datasets such as ADTTE.

Table 4: Categorisation of ADaM datasets

| SIMPLE SAFETY | Complex SAFETY | Efficacy |
|---|---|---|
| Does not fit into the complex category | Datasets with either complex structure requiring significant manipulation of input data and/or with large volumes of derivations | Datasets with efficacy parameters |

Looking at the breakdown of datasets types, on average 20% to 30% of datasets are classified as efficacy with the rest as safety datasets of various complexities.

## 6. PROCESS IMPROVEMENT

This section will go through some of the key steps taken in an effort to improve the process of producing ADaM datasets, using some of the lean six sigma toolkit.

**SIX SIGMA TOOLKIT: DMAIC**

When considering a project aimed at improving an existing process, Six Sigma uses a project methodology consisting of 5 phases under the acronym **DMAIC**.


- ***D**efine* the process, the voice of the customer and their requirements, and the project goals, specifically.
- ***M**easure* key aspects of the current process and collect relevant data; calculate the 'as-is' Process Capability.
- ***A**nalyse* the data to investigate and verify cause-and-effect relationships. Determine what the relationships are, and attempt to ensure that all factors have been considered. Seek out root cause of the defect under investigation.
- ***I**mprove* or optimize the current process based upon data analysis using techniques such as design of experiments to create a new, future state process. Set up pilot runs to establish process capability.
- ***C**ontrol* the future state process to ensure that any deviations from the target are corrected before they result in defects. Implement control systems such as statistical process control, production boards, visual workplaces, and continuously monitor the process. This process is repeated until the desired quality level is obtained.

### 6.1 MEASURING OUTPUT VARIABILITY IN TERMS OF ADaM DATASETS

Output variability is important as it is the indicator as to the baseline state of the process, Also this Is the measure which is used to determine whether process improvements have worked i.e. whether the process improvements have resulted in a decrease in output variability and also whether the targets in terms of performance have been met.
There are a number of parameters that could be used to measure output in terms of ADaM dataset programming. Ideally the customer should set the target in terms of over and/or under performance but most customers are reluctant to articulate under-performance targets or set targets such as no programming errors in all deliverables in all studies.

Examples of output measures that could be used for the ADaM programming process are as follows:

Table 5: A selection of outputs from ADaM dataset creation

| Process Output | Measure | Defect (examples) |
|---|---|---|
| Datasets | # of major/minor programming errors (per deliverable) | 1+ major programming errors and/or 10+ minor programming errors (major/minor as per agreed definition) |
| | % of deliverables not fit for purpose (i.e. with at least one defect) | >5% |
| Documentation | # of critical/major audit findings | >0 critical and/or >1 major finding |
| Cost | Average cost/dataset | |
| Time | # days deviation from agreed delivery plan (per deliverable) | >1 day late |
| | % of deliverables on (or before) time | ≤90% |
| Customer satisfaction | Meets/exceeds expectations (e.g. rated on 5 point scale) | <85% of studies rated as meets/exceeds expectations |

In terms of this process improvement effort, the focus was the measurement of ADaM datasets 'errors'. These errors are collected via a system either as changes noted by the programming validator, or changes noted by the statistician, or the client. These error rates along with the types of errors are being analyzed within the department quarterly.

### 6.2 DESIGN PHASE: VISUALISING THE PROCESS of ADaM DATASET CREATION

A process mapping tool called **SIPOC** was used during the design phase to help visualize the full scope of a process and what influences it in terms of:

- **S**uppliers
- **I**nputs
- **P**rocess
- **O**utputs
- **C**ustomers

Table 6: Typical SIPOC diagram

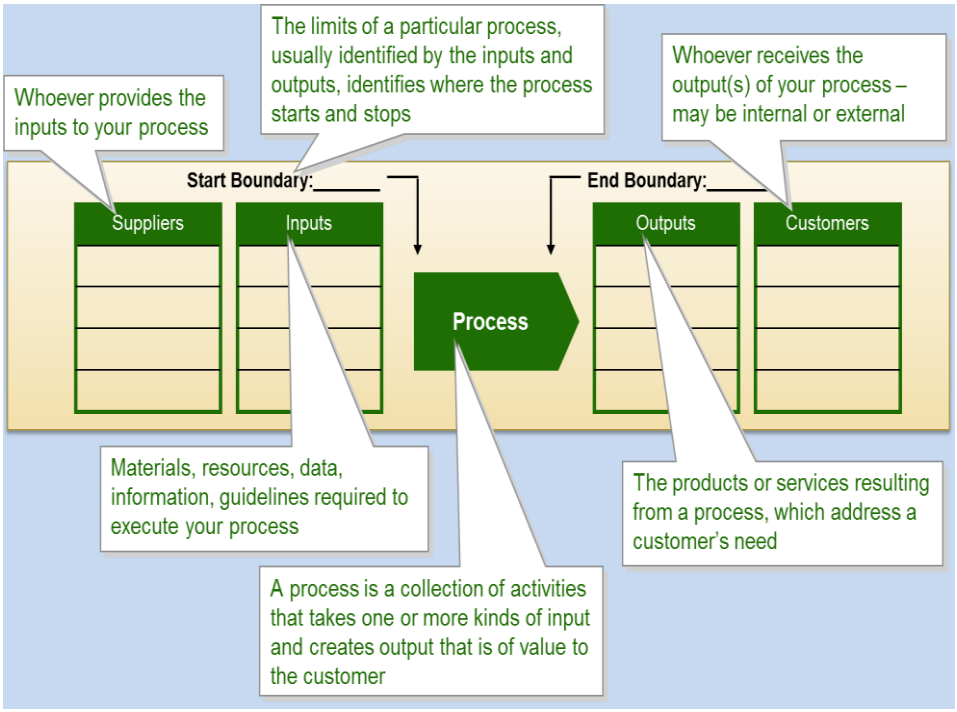


Table 7: SIPOC ADaM datasets

| Suppliers | Inputs | Process | Outputs | Customers |
|---|---|---|---|---|
| • DM<br>• Clients/sponsor<br>• Biostatistics<br>• Clinical<br>• Patients | • Protocol<br>• Data<br>• SAP<br>• ADaM dataset specifications<br>• TLF mocks<br>• ADaM template code<br>• Programming rules | • Training<br>• People<br>• Program development & testing<br>• Program validation<br>• Code review<br>• Lead Programmer review | • Datasets<br>• Programs<br>• Logs<br>• Documentation | • Biostatistics<br>• Medical Writing<br>• Clients<br>• Regulatory authorities<br>• Patients |

When considering how to improve the process, all of the elements above need to be considered

### 6.3 ANALYZE: ROOT CAUSE

In the analyze phase of process improvement the goal is to analyse the data and to investigate and verify cause-and-effect relationships. Root causes of defects/errors are usually divided into 6 possible categories

Table 8: Root cause analysis

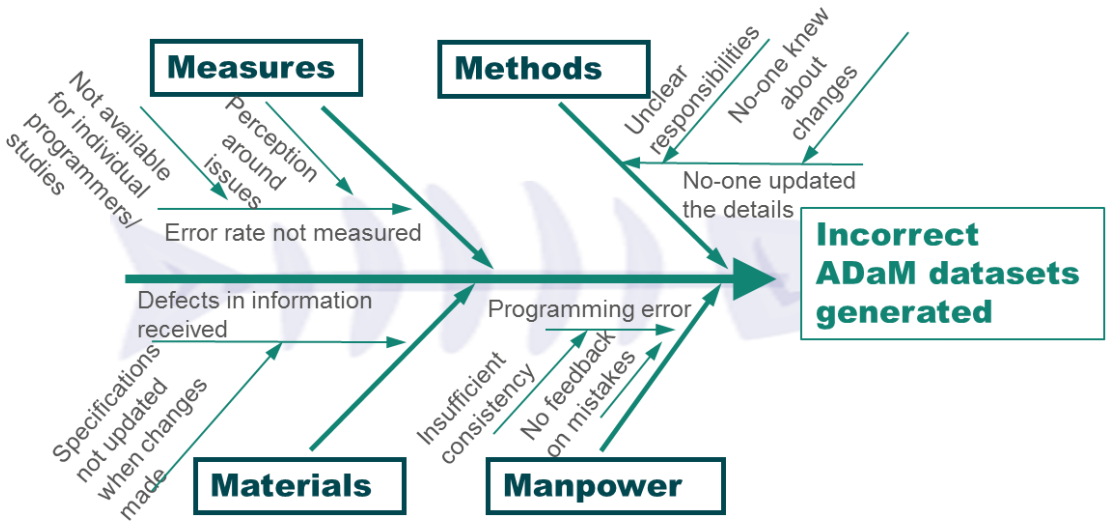
| | | |
|---|---|---|
| **Manpower** | **People, Staffing, Organization, Skills, Management** | • **Are people trained with right skills?**<br>• **Is there person to person variation?**<br>• **Are people over-worked?** |
| **Mother Nature** | Environment, Work Environment, Market Conditions, Regulatory Environment | • Is the workplace safe and comfortable?<br>• Are outside regulations impacting the business?<br>• Does the company culture aid the process? |
| **Materials** | Parts, Supplies, Forms, Information, Inputs | • Are parts, forms or supplies obsolete?<br>• **Are there defects in the inputs/specifications?**<br>• **Is information stable throughout process?** |
| **Methods** | Procedures, Policies, Documentation, Management Systems | • **How is the work performed?**<br>• Are procedures correct?<br>• Are procedures enforced? |
| **Machines** | Tools, Software, Technology, Equipment, Facilities | • Is equipment reliable? Properly maintained?<br>• Is there sufficient capacity?<br>• Are software platforms compatible? |
| **Measures** | Judgment, Evaluation, Measurement Units/Methods/Devices | • **Is data good enough? Is data readily available?**<br>• **Is data open to interpretation or bias?**<br>• Do different people judge inputs and outputs the same way? |

In order to determine the most important causes of variation, a tool called 'Fishbone analysis' is used to analyse the root cause of defects.

Wikipedia defines 'Fishbone analyses as

> *To break down (in successive layers of detail) root causes that potentially contribute to a particular effect. Ishikawa diagrams (also called **fishbone** diagrams, herringbone diagrams, cause-and-effect diagrams, or Fishikawa) are causal diagrams created by Kaoru Ishikawa (1968) that show the causes of specific events.*

Table 9: Fishbone analysis of the root causes of error in ADaM dataset creation




Initial root cause analysis of improving the process relating to ADaM datasets generation showed the following areas that needed to be focused on

**Programmer to programmer variation**
**Defects in terms of the specification**s

### 6.4 IMPROVE PHASE: IMPLEMENTING PROCESS IMPROVEMENTS

The following solutions were implemented.

#### 6.4.1 ADaM template code

**Problem:** Programmer to programmer variation
**Solution: ADaM template code**
**Hypothesis:** The use of ADaM template code would minimize errors

**ADaM template code**

Whilst ADaM dataset structures are well defined and standard, ICON recognized the need to accommodate study to study variations. With this in mind, ICON decided to create a comprehensive suite of template programs which also allow flexibility for study to study variation.

The goal of iACE in creating ADaM template code was that this would increase efficiency & productivity & decrease cycle times. The goal was to create code that would require minimal modification from study to study.

Some of the key characteristics of the template code

- Is based on ADaM V2.1, ADaM IG V1.1 and ADaM OCCDS V1.0.
- Includes mainly safety datasets:
- ADaM Template code provides all ADaM required variables and ICON standard mock variables at minimum
- For variables derivation, wherever applicable ADaM Template code follows ADaM IG rules. For example TRTSDTM derived with Date time of first exposure to treatment for a subject in a study.
- Some derivations show wide variety in most studies. For example: Date imputations, the ADaM Template – common rules were implemented

ADaM template code follows a uniform structure with clear comments that are sufficient to understand the code. Code blocks, shown below, make code review and modifications/changes easier when maintenance needs to be performed.

- Header section
- Setup section
- Define Macro section
- Data section
- Study specific code block
- Analysis section
- Report section
- Cleanup section

Analysis of ADaM datasets created in 2017 shows that ADaM template code covered many of the most frequently produced ADaM datasets

Table 10: Most frequently produced ADaM datasets since beginning of 2017

| Top 10 ADaM datasets | Covered by ADaM template code |
|---|---|
| ADAE | X |
| ADSL | X |
| ADLB | X |
| ADCM | X |
| ADEX | |
| ADVS | X |
| ADMH | X |
| ADEG | X |
| ADTTE | |
| ADPE | X |

### 6.4.2 Standard DIAMOND macros

**Problem:** Programmer to programmer variation
**Solution:** Standard reporting macros
**Hypothesis:** The use of standard macros and utilities would minimize errors


- Set of modular macros.
- Each macro is designed to perform a specific function and when logically grouped together, support each step of development of an output :
  - Environment settings (global macro-variables)
  - Intermediate analysis (summary statistics, frequencies,…)
  - Creation of the report dataset for the output
  - Creation of the output
- To be used by development programmer only to create standard outputs

### 6.4.3 Supplemental dataset checks

**Problem:** Programmer to programmer variation
**Solution:** Supplemental dataset checks
**Hypothesis:** Additional internal compliance checks developed by iACE coupled with checks done by external vendor software would minimize non-compliance issues

### 6.4.4 Patient walk through

**Problem:** Measuring defects in terms of the specifications
**Solution:** Patient walk through
**Hypothesisis:** Biostatistician reviews analysis datasets via patient walk through would be an extra check of the validity of the ADaM datasets

- Patient Walkthrough (PWT) used for stat review of ADS
  - A step-by-step check of the variables/values in the Raw datasets (SDTM or equivalent) contributing to a derived variable in the ADS to ensure that these are consistent, on a per-patient basis
- Focus is on checking implementation/programming of key derivations and algorithms
- Scope of review is pre-specified in QC Plan
  - Based on SAP and ADS specifications
- Lead Statistician is responsible for writing the QC Plan which defines the review parameters

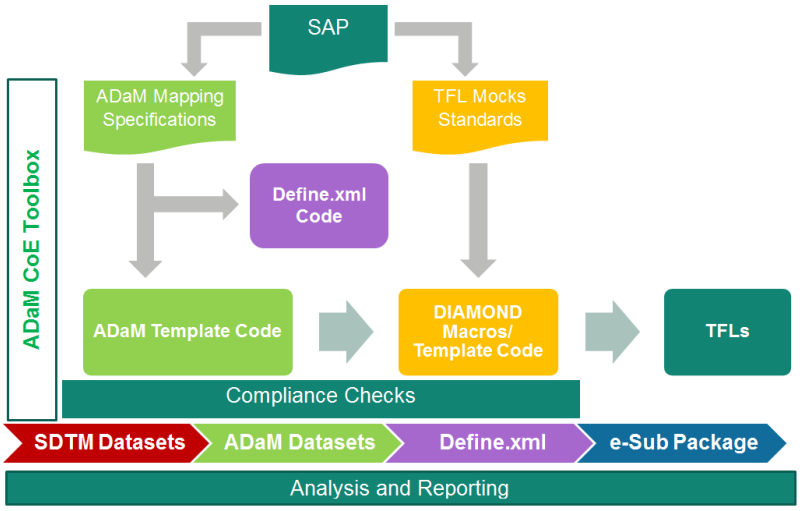
### 6.4.5 Other process improvement

Other process improvements were implemented such as improvements in the specifications of datasets (e.g., ADaM specification peer review), working with DM in improving their checks in terms of delivery of the SDTM datasets, supplemental training, and individualized feedback on errors etc.
iACE has also developed a standard ADaM dataset specification, pre-populated in alignment with the ICON ADaM template code and standard mock TFLs.

In addition, iACE has developed a simple utility that converts the study ADaM specification to shell datasets, that are then checked by P21 and the ICON customised utilities'.

The graphic below details the ADaM generation ecosystem in terms of tools.

Table 10: iACE toolbox




## 7. CONCLUSION

At the start of the paper, some of the basics & metrics around ADaM dataset creation were outlined. Following this, steps taken to improve the 'defect' rate of ADaM dataset creation were examined. The use of Fishbone analysis to determine the root cause of errors was reviewed and the solutions implemented around these root causes were examined. Optimising the production of ADaM datasets is an iterative process and requires ongoing analysis of the output variation and the root causes.

## 8. REFERENCES

(1) http://www.appliedclinicaltrialsonline.com/fda-binding-guidance-pivotal-milestone-cdisc-standards. FDA Binding Guidance: A Pivotal Milestone for CDISC Standards. Dec 01, 2016. By Barrie Nelson. Applied Clinical Trials. Volume 25, Issue 12

## 9. ACKNOWLEDGMENTS

Many thanks to all in the B&P department who contributed to providing metrics and reviewing this paper.

## 10. RECOMMENDED READING

Recommended reading lists go after your acknowledgments. This section is not required.

## 11. CONTACT INFORMATION

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

Pam Howard
ICON Clinical Research Plc
South County Business Park
Leopardstown
Dublin 18
Ireland
Work Phone: +353 1 2912000
Email: pam.howard@iconplc.com
Web: www.iconplc.com