Paper PP04

Using Meta Data to drive standards, speed, consistency and quality in reporting

Shafi Chowdhury, Shafi Consultancy Limited, U.K.
Daniel Boisvert,Biogen, U.S.A

ABSTRACT:
Meta data, meta data, it is all around us with the promise that one day it will make our lives easier. One day we will be able to push that button and the meta data that was so time consuming to populate will give us everything we ever wanted. This paper will show how company standards, therapeutic area and project standards, together with a small number of selected meta data can help us get closer to the dream. It may not write the report for us (yet!), but it can generate all the reporting programs in a standard structure that are easy to run and maintain over time. They will follow the agreed standards, and be quick to produce and validated. Meta data finally delivering on the promise to drive standards, speed, consistency and quality in reporting.

INTRODUCTION

Current modern process includes much data driven activity, but ends effectively at dataset creation. Why does it end with an MDR? There is a big opportunity to learn from the success of the upstream data structuring and apply it to the analysis. It is obvious that an ADaM dataset is more use than a PDF of an ADaM dataset but when we talk about a SAP, table shells, analysis programs or the analysis outputs it is hard for us to understand how to move from document to data. However, this move is possible and the results position us much better for the future.

There is tremendous value in knowing what was planned to be analyzed, what was analyzed, how it was analyzed, what the results are, where they are stored and where they are used and reused. All of these have internal operational efficiency possibilities and outside of the department and for the good of clinical research opportunities.

Timelines for analysis continue to get squeezed and the budget for creating them tends to stay stable or reduce, yet the amount of work to produce the analysis has increased. As more requirements for traceability, reproducibility, auditability, and also new data standards and different types of data come in, the time dedicated to analysis decreases. In our current models we see added steps to our already full days. Adding these to our current process is not the answer, as the questions asking of us have no relation to when the process was first developed. Instead we need to understand the current state as our baseline and build a technology stack and processes that work to produce exactly what we need.

There is no benefit to building a cutting edge analysis platform all in one shot. By the time it has been built either the technology will be outdated or the scope will have again adjusted. Also

something has to be said for the change management. Innovation takes time, and its adoption takes longer. Instead we have to take an incremental approach finding pieces of file based information and converting them to data. This paper will talk about the pros and cons of each and their relative cost (time). The order of adoption is relatively unimportant, as each company will have different priorities. The importance is to carefully move all information into structured, reusable, findable, analyzable information.

STRUCTURED APPROACH

Benefits of meta data is achieved over time, as more data is generated, the more useful the data becomes. Therefore it is important to ensure that the structure of the meta data is one that can survive the test of time. If the structure needs constant updates then the system becomes more burdensome as those providing the data will also need to perform some kind of updates, and it will very quickly fall out of favour.

Structured protocol

Pros:

- Critical information - the foundation of the entire clinical trial
- Many parts of the protocol can be re-used
- Aids in findability for secondary research
- This can save both time and effort, and therefore lead to a reduction in lead time for trials.
- Structured protocol off the shelf tools exist

Cons:

- Protocol is developed in co-operation with many different groups, and for a structured approach to be used, this would require a major company wide investment strategy.

Structured SAP

Pros:

- Second only to a structured protocol in terms of importance.
- Information from the SAP is used and reused in many different documents, including the CSR, and more informally in instructions for programmers.
- Critical information can therefore be structured for reuse.
- What will be analyzed can then be later found and reused.
- This could be used to efficiently speed up SAP writing, improving consistency, quality and reduce time.

Cons:

- As this is a regulatory document, and therefore subject to strict scrutiny, care needs to be taken that the document can be easily updated for each study.
- The users are also not necessarily data people and may find it cumbersome.

- Although there are systems available within some companies, there are also no well-known commercial products available.
- Creating a custom system is costly

Structured Table shells

Pros:

- Standard table shells, knowing what was the planned analysis on every trial
- Easy to find analyses already performed
- Can easily see variables, data subsets used in the analysis
- Allows data to be prepared in advance
- Structured shells allow shells to be more flexible, if data connected to it is pre-defined

Cons:

- Requires analysis to be pre-planned
- Need to change habit of requesting changes at a later time
- No well-known commercial products exist.

Structured Analysis Programs

Pros:

- As there is standardised data, using standard analysis programs can significantly improve quality and consistency across trials
- Can be automated based on derivation rules
- Easy to see which version of rules used and which studies have used these rules

Cons:

- Data driven, identification of data issues
- Handling of data issues can vary due to trial specific needs

Structured Analysis Results

Pros:

- Results can be created once and reused as needed
- Reduces QC burden
- Results could be updated automatically as data updates
- Full traceability back to original data and method used to create result is preserved

Cons:

- Many home grown efforts over time have not been successful
- No well-known existing commercial products

Getting started

As standardised analysis dataset are already generated by companies using various methods, generating structured shells that are pulled in together with structured analysis results to produce the tables required is the most effective approach as a starting point. There are two key details that are required for this to be successful:

- List of analysis to be displayed
- Shells specifying which analysis are to be displayed and how

Once the list of analysis to be displayed is determined, then functions can be programmed to produce the analysis. These functions will store the results in a pre-specified structured dataset that can then be selected as and when required. This means for example the number of patients in the treated analysis set is only derived once, but the value will be used in many different table shells and may be displayed in a hundred different tables.

The meta data associated with each shell will specify which function is to be used for each of the different parts of the display template. This will ensure that when this shell is called, the pre-calculated results are then simply displayed in the appropriate places. No further derivations are performed at this stage.

This ensures there is full traceability. One can see which shell is used for a particular table, the functions used to populate that table, and the structured analysis datasets where those numbers are generated from. As there is traceability from the analysis dataset to the original data, the link is therefore complete from the raw data to the individual values within a table.

How far do we go?

We can reach a place where we have:

- Structured table shells
- Structured analysis datasets
- Structured results datasets

This then allows table shells to be created where the user only needs to select what data they want to show and what analysis they want to do with it. The results are then automatically derived using the pre-defined functions, and displayed by code that places the selected results in the appropriate place within the table shell.

Now:

- Develop table shells
- Write programs
- QC output
- Perform thousands of iterations of this process to generate the different outputs.

Future:

- Decide what summaries to present
- Derive those once using pre-programmed functions into a structured dataset
- Use the values as and when needed in each table shell

Aim: analysis should be done as you design the shell

This approach means that we replace programs and use functions, create structured analysis results based on standard data and use these results directly to produce the tables.

Couple things that are slowing us down:

- Submitting programs to FDA
    - Multiple scenarios where sponsors have updated their analysis code to take out macros, update macros to produce readable code.
    - More than 80% of the outputs do not require complex statistics – why are we submitting the program that is doing a count? or a %? or univariate more than to document the selection criteria?
    - Submitting programs means that we are not allowed to automate the simple analysis through modern practices.
- Job security
    - No one wants to talk about it, but there is a feeling of dread when we talk about automating the simple analyses. What are programmers going to do? We have to get over this
- Technology
    - While the technology exists to create these tools, there is not enough of a demand for technology companies to invest.

        - Pharma industry is so SAS heavy that mentioning another technology gets squashed because “all of our expertise” is in SAS.
        - Understanding and accepting a paradigm where there are no physical programs that create the analysis is a scary concept when we are so used to the dataset->program->output concept. Web applications generally skip the program piece (as a separate file that just creates one output file, and has a reporting module that uses methods to generate the output).
        - Misunderstood requirements about software validation, reproducibility and traceability (all currently understood only in a SAS similar paradigm) limit creative people from gaining true efficiency in this process.

CONCLUSION

To repeat the process that has got us here is one way of thinking, and to try and find new ways of doing the same thing is another approach. However, to make a leap instead of just a step or a jump, we need to aim to go further than we think we can achieve.

May be it is time to stop thinking about programs, logs and outputs. Let us think about the analysis, and drive the delivery from that approach. What data do we have? What analysis do we want to perform? Show the results in the most appropriate manner and standardise that display template. This then takes us away from the mundane, time consuming efforts of managing output and shells, and gives us more time to look at innovative ways to analyse the data and present those results.

CONTACT INFORMATION

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

Daniel Boisvert
Biogen
daniel.boisvert@biogen.com

Shafi Chowdhury
Shafi Consultancy Limited
Uxbridge
United Kingdom
Email: shafi@shaficonsultancy.com
Web: www.shaficonsultancy.com