Paper TT02

How can we integrate new technology and maintain data integrity?

Shafi Chowdhury, Shafi Consultancy, Uxbridge, United Kingdom

ABSTRACT

When a 14 year old teenager produced for me in an hour a complex figure that I think would take most SAS programmers at least a day, should we be worried? Or should we be excited that there are new things we can learn? Should we be happy about the new opportunities and the new paths they may lead to?

Uninhibited and unburdened by our ingrained ways, left to their own devices, the young generation are able to solve problems quickly, efficiently and correctly. Although we are embracing software like R, too often it still feels like the hobby of an enthusiast within our team, who uses it in presentations here and there. Is it time to bring everything into the fold and find them a place within our process – be it R, Python or any of the many other tools around? This paper will explore how programmers can embrace new technology without affecting delivery integrity.

INTRODUCTION

The purpose of this paper is to discuss some of the factors that may limit our usage of new technology in data analysis and how we can overcome them. SAS® has been the key tool for analysing clinical trial data for over 30 years, and this will continue. However, there are many new tools available for analysis, and although we use them, and some, like R are becoming more common, they are still used on the fringes of analysis as opposed to the primary tool. Where they are used, they are often used in a separate environment, almost secondary.

As we move into the brave new world, embracing technology and all it has to offer is critical. One of the key components of a regulated environment is data integrity. Maintaining data integrity as we analyse all the various data to compile the final report is essential to the validity of the report. Where the data came from, how it was analysed, confirmation that the result produced is accurate are all key factors in data integrity. Having a process to maintain these in the background when new analysis tools are introduced, and delivering this quickly is key to increasing the usage of new analysis tools in final deliverables.

DATA INTEGRITY

Data integrity relates to the completeness, consistency and accuracy of the data throughout the lifecycle of the data, from its creation/collection to manipulation, processing and maintenance. This integrity is what gives credibility to any results and conclusions drawn from the analysis of the data. This integrity is what allows the results to be trusted.

To demonstrate this integrity, the data should be attributable, legible, contemporaneous, original or a true copy, and accurate (ALCOA). To maintain data integrity, systems and processes used throughout the life cycle of the data should support users to detect errors in the data, and be validated for their intended purpose.

ATTRIBUTABLE

It should be possible to determine who generated a piece of data and when. Although this can be manually documented, increasingly having an audit trail to digitally document this in a validated system both reduces the administration work and increases reliability. Good documentation practice also recommends that who changed or added new data and when is also documented, although the who can be an alias.

LEGIBLE

The data should be readable and permanent, i.e. it should not be possible to accidentally change the original data and It should be possible to determine by whom and when a piece of data was generated or updated.

CONTEMPORANEOUS

The actual date and time of when a piece of data is generated or changed should be documented. The time should not be retrospective and all changes should be in chronological order to confirm reliability of data.

ORIGINAL

The medium where the data was first recorded is important to keep, i.e. the primary data in primary storage. In this respect, traceability as the data is copied across environments and analysed in different systems should be documented so that traceability back to the origin is maintained.

ACCURATE

The data must be error-free, complete and reflect the original data collected. To ensure this is true, all changes to the original data must be documented i.e. who made the change, when, why and what change was made, If data is changed programmatically, then program details should be documented and the log file saved to confirm that the program worked as expected.

KEY CHALLENGES TO NEW TECHNOLOGY

New technology brings with it a level of uncertainty, and it should. However, once the system has been tested and validated for its purpose then the uncertainty is reduced if not completely removed. Provided the system is also maintaining an audit trail and follows guidelines such as CFR 21 Part 11, then the system can be used for data analysis with more assurance. For submission and other analysis for regulatory GxP systems (various Good Practices used in the development and maintenance) should be used.

TIME

The validation part of the process is often the most time consuming. Even GxP tools have to be validated before they can be used. Hence the usage of new technology and tools is often limited to presentations and exploratory analysis. At times the time taken to implement a new system runs into years, and this leads to the environment in which it is being used moving faster than the tool, i.e. it no longer provides many of the benefits that was originally intended. If new technology can be integrated into everyday work quickly, also for use with testing and running pilots, then their full benefit can be realised quicker.

PROCESS

Although it is not said in this way, “we have always done it like this” sits comfortably behind many decisions of new technology. Although the tools may differ, often we are using them in the same way to produce the same output, whether that is datasets or tables, listings and figures. Programs are written to produce analysis datasets, and then programs are written to use these to produce summary tables, figures and listings for the trial report, pharmacokinetics report, regulatory responses and so on, each time, ensuring that there is a program, log and an output file to ensure data integrity.

This process has not changed in over 30 years, and given that those in charge have grown up with this process, and there is a reluctance to take unnecessary risk, it is unlikely to change in the near future. However, the new starters in the industry have a very different thought process. They did not learn to use the internet, it was always there. They are not burdened with old processes that were born of a different time with different barriers. They think differently, and perhaps more space should be provided to broaden this thinking.

IT INFRASTRUCTURE

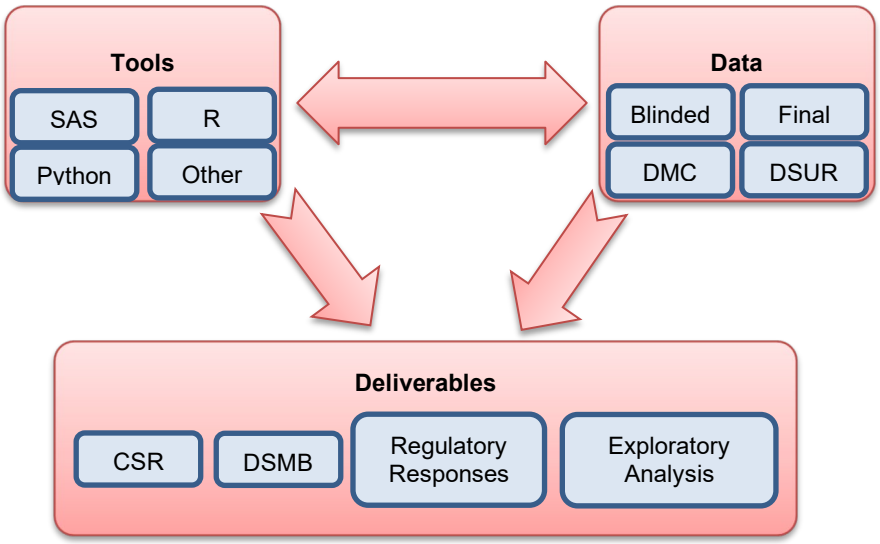
Given the need for data integrity, processes have been devised and IT infrastructure was deployed to meet those requirements and restrictions. The importance of data integrity has not diminished, and IT infrastructure has to take into account the risk from human errors, transfer errors and cyber threats while providing validated data. However, as more and more systems are moving to using cloud computing, both private and public, this is then allowing easier access to new tools for data analysts with remote access.

TRANSFERS BETWEEN SYSTEMS

Data used to be copied from one environment to another so that different tools can be used for analysis. The various output produced were then copied back to a central area where all output were combined into the deliverable required. Each of these transfers adds risk to data integrity.

SUCCESSFUL INTEGRATION OF NEW TECHNOLOGY

Re-usability of data by different tools is one of the strongest forms of data integrity. Therefore having an environment where different tools can be installed and use and contribute to the same pool of data, means the user can select multiple tools for the same delivery.



SOLUTION

An environment where programs using different tools can be stored and submitted from the deliverable area, where the output is subsequently returned into the deliverable area, e.g. CSR (Clinical Study Report), means that there are no manual data transfers to use the different tools and no manual copy of individual outputs into one central location for them to be combined. This is a huge step for data integrity. There is no question of whether the same snapshot of data is used, or if the output was changed after the program was submitted, there is also no question of date/time in one output being different from another because they are coming from different servers in different locations!

NEW AND NOT SO NEW TECHNOLOGY

R has been widely used for more than 15 years, it is not new, yet its validation status is still limiting its usage within the pharmaceutical industry. Python is relatively new in the industry and is considered to be very good for big data, however, it is very resource intensive and the number of programmers widely available is limited. There are other analysis tools available like SPSS, Tableau, Power BI and Apache Spark that can be used in the pharmaceutical industry as they are used in others. However, until the enthusiasts are allowed to show the established teams what is possible, the industry remains slow and instead pulls the new starters back to how it has been done over the years.

CONCLUSION

Systems are there to help the users produce what they need to deliver, and this should be done quickly and with integrity so that any results produced are credible. Systems should be intuitive and easy to use, and new systems should provide actual measurable benefit to the users over their existing systems. Having a system from which programmers can submit SAS, R and Python programs, or perhaps NONMEM and other tools that are used by different groups that use the clinical data, will ensure that the user is focused on their delivery, not where they need to copy the data to in order to use a different tool to perform an analysis.

Using R for a small number of analysis is becoming quite common, while the majority are produced in SAS. Having a system that allows the users to do this without risking data integrity is key for the pharmaceutical industry. There are other tools also available and more new tools will come into the industry, but those in the know must be allowed to experiment with these to see if they really are better, and where this is proved to be so then those new tools should be integrated quickly into the fold and be available for others to use.

The aim should be to have an environment where it is easy to integrate new tools, even if they are used for deliverables not requiring GxP tools, so that users can test them and see if they add benefit. Based on that experience, decisions can be taken as to whether to move those tools to be GxP compliant and used for more deliverables. This will ensure that new technology can be tested quickly and successful tools can be integrated into the system for everyone to use in a very short time frame.

REFERENCES

Data Integrity and Compliance With Drug CGMP Questions and Answers Guidance for Industry, FDA, Dec 2018

CONTACT INFORMATION

Contact the author at:
Shafi Chowdhury
Shafi Consultancy Limited
The Charter Building
Charter Place
Uxbridge
Middlesex
UB8 1JG
United Kingdom

Email: Shafi@shaficonsultancy.com
Tel: +44 7702 887 219

Web: www.shaficonsultancy.com