HOW TO IMPROVE DATA INTEGRITY

Peter Boogaard worries about the integrity of data in the laboratory

In 2013, the US Food and Drug Administration (FDA) reported that laboratory processes, and deficiencies associated with laboratory controls, were ranked in the top three most frequent causes of 'observations' following FDA inspections. The same report also cited an increase of 50 per cent in warning letters related to data integrity.

During the annual meeting of the International Society for Pharmaceutical Engineering (ISPE) held in 2014 in Las Vegas, it was reported that the FDA had identified more than a dozen Indian pharmaceutical manufacturers who had problems with the data integrity practices at their facilities. That number is significant, since India and China account for 80 per cent of active pharmaceutical ingredient (API) production. Many of these companies in Asia are privately held and therefore not mandated to submit corporate information. In these countries labour is plentiful and cheap, so facilities tend to use paper-based manual processes, which increases the potential for inconsistencies in data and lack of data integrity.

Other regulatory bodies, including the European Medicines Agency, have made similar observations (Table 1). It is expected that this trend will continue to grow. Many international corporations externalise significant parts of their operations to cut costs, but it has been observed that the number of occasions is growing when poor data integrity practices outweigh the savings in cost.

The FDA reported an increase of 50 per cent in warning letters related to data integrity

Data integrity is currently one of the highest cited areas in regulatory observations. Yet, data integrity is not a new requirement. For years, the basic principles have been described in international good manufacturing practice (GMP) guidelines. In this article I will highlight the ways in which organisations, to their own benefit, can reduce data integrity inconsistencies within their operations. Before zooming in on the details, however, we need to set a baseline to ensure we have a common understanding.

Table 1: Laboratory data integrity observations

Alteration of raw, original data and records
Multiple analyses of assay with the same sample without adequate justification
Manipulation of a poorly defined analytical procedure and associated data analysis in order to obtain passing results
Backdating stability test results to meet the required commitments
Creating acceptable test results without performing the test
Using test results from previous batches to substitute testing for another batch

CONTEXT IS KING

FDA regulations define an electronic record as any combination of text, graphics, data, audio, pictorial, or other information represented in digital form that is created, modified, maintained, archived, retrieved, or distributed by a computer system[1]. Data integrity is the assurance that data records are accurate, complete, intact, and maintained within their original context, including their relationship to other data records.

In short, data integrity aims to prevent unintentional changes to information. It refers to maintaining and assuring the accuracy and consistency of data over its entire life-cycle, including the usage of any system which stores, processes, or retrieves data. The definition applies to data recorded in electronic and paper formats or a hybrid of both.

Ensuring data integrity means protecting original data from accidental or intentional modification, falsification, malicious intent (fraud), or even deletion (data loss). It is the opposite of data corruption! Data integrity and security are closely linked to the 21 CFR Part 11 for electronic records and electronic signatures, but also to other directives relating to data, regardless of format. ➤

MIND-SET CHANGE

In a data integrity-focused audit, the emphasis has moved away from providing information solely based upon a technical and scientific context, towards providing evidence that the final analytical results are not false. This holistic approach, based on the end-result, may require a different mind-set for certain organisations and requires a focused effort to prepare for this new approach. As regulators increase their focus on data integrity and reliability, auditors are examining with multiple regulations and standards in mind. These may include Pharmaceutical Quality/Current Good Manufacturing Processes[2] (CGMP), Good Laboratory Practices (GLP), Good Automated Manufacturing Practice (GAMP[3]), Good Clinical Practices (GCP) and the Application Integrity Policy (AIP) in addition to US FDA-recognised consensus standards.

According to the FDA, source data needs to be 'attributable, legible, contemporaneous, original, and accurate' (ALCOA) and must meet the regulatory requirements for recordkeeping. ALCOA+ refers to additional terms included by the European Medicines Agency on electronic data in clinical trials (Table 2). It is highly recommended to use this concept.

Table 2: Terms associated with ALCOA+

A	**Attributable**	Who performed an action and when? If a record is changed, who did it and why? Link to the source data
B	**Legible**	Data must be recorded permanently in a durable medium and be readable
C	**Contemporaneous**	The data should be recorded at the time the work is performed and date & time stamps should follow in order
O	**Original**	Is the information the original record or a certified true copy?
A	**Accurate**	No errors or editing performed without documented amendments
+	**Complete**	All data including repeat or reanalysis performed on the sample
+	**Consistent**	Consistent application of data time stamps in the expected sequence
+	**Enduring**	Recorded on controlled worksheets, laboratory notebooks or electronic media
+	**Available**	Available / accessible for review / audit for the life time of the record

INFORMATICS DATA JOURNEY

Data-intensive science is becoming far more mainstream in laboratories. It is best to take a pragmatic approach to what this means for laboratory operations. When samples are being analysed, several types of scientific data are being created in the laboratory. They can be categorised in three different classes (Figure 1).

Automating the capture of metadata is a very effective way of maintaining data integrity. Self-documenting processes capture metadata automatically, without human interaction. They eliminate transcription errors and avoid unnecessary retyping of data. In a recent survey, 32 per cent of respondents stated that data integration in a paperless laboratory will eliminate manual entries and data transfer. Automating this process will also eliminate hybrid systems – a combination of a paper and electronic system. Hybrid systems significantly increase integrity challenges, since both systems need to be synchronised in a consistent way. Hybrid systems are no longer recommended.

SINGLE POINT OF TRUTH FOR (META) DATA

To avoid challenges to the integrity of data, it is crucial to have one copy (the master copy) of the data or information. Master data is a single source of data used across multiple systems, applications, and processes. To achieve a single point of 'truth', it is necessary to understand the key differences between spreadsheets and databases. The perception that a spreadsheet can act as a database is fundamentally wrong. The primary function of a spreadsheet is to manipulate, calculate, and visualise data, while a database's primary function is to store and retrieve data in a structured manner.

A spreadsheet has serious drawbacks when used for data storage: it cannot enforce relationships; there are no multi-user capabillities; and it offers neither data validation nor protection against data corruption. There is a risk of misuse of spreadsheets in the laboratory. The worst nightmares are caused by the sort function and

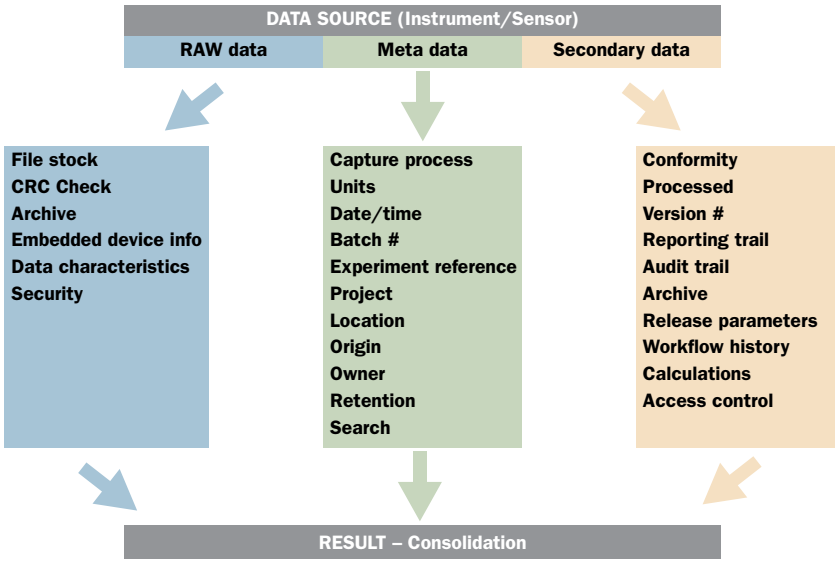

Figure 1: Different classes of data are created in the laboratory

Saknakorn/Shutterstock.com

When a smartphone captures a photo or film, it automatically includes systematic metadata with the object. Examples include capture and storage of GPS location, weather conditions, personal condition such as heartbeat – without adding an additional manual entry to the data object. Similar developments are expected to be introduced in the scientific community. Modern balances may automatically include temperature and humidity when transmitting the weight.

Table 3: GMP Regulatory Requirements for Data Integrity	
Instruments must be qualified and fit for purpose	**[§211.160(b), §211.63]**
Software must be validated	**[§211.63]**
Any calculations used must be verified	**[§211.68(b)]**
Data generated in an analysis must be backed up	**[§211.68(b)]**
Reagents and reference solutions are prepared correctly with appropriate records	**[§211.194(c)]**
Methods used must be documented and approved	**[§211.160(a)]**
Methods must be verified under actual conditions of use	**[§211.194(a)(2)]**
Data generated and transformed must meet the criterion of scientific soundness	**[§211.160(a)]**
Test data must be accurate and complete and follow procedures	**[§211.194(a)]**
Data and the reportable value must be checked by a second individual to ensure accuracy, completeness and conformance with procedures	**[§211.194(a)(8)]**

Derived from the laboratory data integrity definition and the applicable 21 CFR 211 GMP regulations – FDA's Focus on lab Data integrity – Bob McDowall Part 1

the copy/paste function losing all associated metadata links, creating a data integrity nightmare.

REDUCE AND SIMPLIFY WORKFLOW COMPLEXITIES

Simplifying scientific processes will significantly reduce data integrity problems. While this industry is trying to harmonise scientific processes, other regulated industries are ahead in this field, so applying common best-industry-processes is still a dream. However, there are signals that this industry is recognising the need. For example, suppliers of balances and titrators are increasing the value of their instruments by implementing approved and pre-validated methods and industry best-practice workflows in their firmware. Almost all major balance suppliers allow methods to be implemented directly in their balances and other wet chemistry instruments. This may seem a small step, but it can significantly reduce the effort involved in validation of integrity, since these methods reduce the number of failure points during operation and thus the need for customisation.

Another example is the proven integration of LIMS processes with enterprise (ERP) workflows. A significant reduction in the frequency of data integrity issues can be achieved by creating consistent workflow integration between both systems. Process harmonisation will initially increase the validation burden but, in the long run, the effort will pay off. The continued elimination of manual control activities will reduce the number of potential data integrity failure points and boost efficiency of laboratory staff and management.

A final example of how the industry is recognising the need to apply best practices is the approach of mapping the entire laboratory workflow[4] and related operations, from sample receipt to release of results, the benefit being consolidation in operational workflow. The overall net effect will be a reduction in validation effort and decreased data integrity risks.

ADOPT AND USE INDUSTRY STANDARDS AND PROCESSES

The American Association of Pharmaceutical Scientists (AAPS) produces guidance on Analytical Instrument Qualification (AIQ) in the form of a white paper, which has been incorporated as General Chapter <1058> within the *United States Pharmacopoeia* (USP). A proposed update has recently been issued for public comment. The changes that are proposed will have a positive impact on the AIQ process of analytical instruments and laboratory computerised systems[5]. A practical reality is that a computerised system cannot be validated without qualifying the analytical instrument, and vice versa.

A serious concern is the lack of data standards in the scientific community. Without standards, data integrity will remain challenging and auditing and verifying is an expensive exercise. The Allotrope Foundation is an international not-for-profit association of biotech and pharmaceutical companies building a common laboratory information framework for an interoperable means of generating, storing, retrieving, transmitting, analysing, and archiving laboratory data and higher-level business objects such as study reports and regulatory submission files. The deliverables of the foundation – sponsored by industry leaders such as Pfizer, Abbott, Amgen, Baxter, BI, BMS, Merck, GSK, Genentech, Roche and others – are an extensible framework that defines a common standard for data representation to facilitate data processing, data exchange, and verification. Allotrope's vision of the laboratory of the future aligns closely with FDA's regulatory objectives[6].

BETTER COMMUNICATION

Communication is a common dominator and often underestimated as a significant point of failure in multidisciplinary laboratory automation projects; but it is also a critical element to reduce data integrity challenges within organisations. Workflow simplification and the adoption of industry best practice pre-defined workflows will reduce complexity. Modern tools such as LIMS, ELN, LES and mobile devices such tablets are married to each other. However, to connect a balance still needs an IT professor. The challenge to the industry is to make pairing a balance with a computer or tablet using a laboratory software application as simple as connecting a phone in your car. Lowering the barrier to integrate instruments will contribute to lowering data integrity challenges in laboratories significantly.

Peter Boogaard is an independent laboratory informatics consultant and founder of Industrial Lab Automation, which provides services to address harmonisation, integration and consolidation of business processes in development and manufacturing. Industrial Lab Automation organises the Paperless Lab Academy, for which *Scientific Computing World* is media sponsor. Taking place in Barcelona, the 2016 event will focus on finding the speed to innovate. The Paperless Lab Academy is the ideal learning platform, for those considering consolidating, integrating and simplifying laboratory data management systems. There is no conference fee for industry delegates.

References

[1] FDA - 21 CFR 11.3(b)(6) guideline

[2] FDA Pharmaceutical Quality/ Manufacturing Standards (CGMP) Guidanceshttp://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm064971.htm

[3] ISPE International Society for Pharmaceutical Engineering - www.ispe.org

[4] Introducing a paperless Lab – http://multimomentanalysis.com/

[5] Life Cycle Risk Assessment of HPLC Instruments - Paul Smith and R.D. McDowall - LCGC Europe, Volume 28, Issue 2, pg 110-117 www.chromatographyonline.com/life-cycle-risk-assessment-hplc-instruments

[6] Allotrope Foundation - IPQ Monthly Update (Jan/Feb 2015 pp 11-14).– www.allotrope.org