Waters™

Creating Sustainable Data Using Empower LMS

Kate Lenton, Fiona O'Leary, Arjan Timmerman, Tracy Hibbs, Kate Wearden.

BACKGROUND

Data is everywhere in a laboratory. Data is being produced from a multitude of informatics systems, for example Empower™ Lab Management System (LMS), waters_connect™ Software etc., and business systems, such as Enterprise Resource Planning (ERP) and Document Management System (DMS). The data may take the form of an original direct measurement or observation (raw data) or it may have had some additional analysis or decision making involvement (processed data).

While this abundance of data presents vast potential in insight generation and decision making, it can also come with several challenges. One of the biggest challenges can be how to effectively manage the data, so the laboratory can access and benefit from it in the future.

An example of this is when data is backed up and archived but not catalogued fully or correctly, or even not catalogued at all. This makes it difficult for laboratory personnel to be informed about the data in the first place, and then to locate and reuse it.

DATA CHALLENGES

The requirement to maintain data is well documented by many Regulatory Bodies (21 CFR 11[1], 21 CFR 211[2], EU GMP Part 1[3], and Annex 11[4]), yet to truly ensure durable data, it must be sustained throughout its complete lifecycle. The challenge however, comes with the variability of information that is needed at each stage of the data lifecycle.

For example, when raw data is first captured, perhaps the only recorded information required to identify that data is the product code, sample name, and date. However when this raw data is processed the new sample name, date, and/or other attributes, are also required to be captured. Fast-forward five years, if the same product is analyzed without capturing the critical product code in the archiving system, it would result in the inability to automatically locate and create a digital relationship with the older, related sample data captured in the five years prior.

The result of a scenario like this is that the original sample data cannot be reused efficiently, the overall efficiency and productivity of the laboratory is compromised, and valuable time has been wasted. This can be avoided if all relevant data is recorded at the time.

To understand what is relevant, it is recommended that regulated companies must map their Data Flow paths and all interactions that occur with the data within these paths. This will ensure they have understood and documented the attributes that must be captured and associated with the data at every stage of its lifecycle. These activities will strengthen and sustain this data knowledge and data integrity within the company.

By truly sustaining the data at each stage of its lifecycle, the data will work harder, so you do not have to.

DURABLE DATA VS SUSTAINABLE DATA

In the previous white paper: Using Empower LMS to build Intrinsic Compliance and Durable Data[5], we described that intrinsic compliance requires that data integrity, as a foundation to compliance and product safety, is so inherent in the business processes and computerized systems that it just happens without conscious thought or action. As we look more closely at sustainable data, let's review the core differences between data durability and data sustainability:

- Data Durability – refers to data having security, integrity, and availability to support product quality decisions at the time of the decision-making and for its lifetime of use.
- Data Sustainability – refers to systems, behaviours, and activities to maintain, uphold, and defend a particular entity or resource continuously over time.

Durable data leads to sustainable data. Implementing robust controls to ensure data survives the retention period and is still accurate and complete enables it to sustained through the data lifecycle.

An example of this resiliency is Empower Software projects. Given appropriate technical controls are in place, projects are considered durable, as all project data is backed up together at the same time so that it can be restored in the future. This durability leads on to data sustainability as the projects can be restored when needed to the same or a later version of Empower Software - this is the activity to maintain, uphold and defend.

For more information about durable data, see the whitepaper: Durable Data for Non-IT: A Lab Manager's Guide to Ensuring Your Empower Data is Secure and Available from Creation Through the Full Mandated Retention Period

ENSURING SUSTAINABLE DATA THROUGH DATA LIFECYCLE MANAGEMENT (DLM)

Traditionally, the data lifecycle consisted of six key stages: Plan, Acquire, Process, Analyze, Report, and Store. As understanding of the importance of effective DLM has increased, so have the number of stages involved in the data lifecycle (see Figure 1).

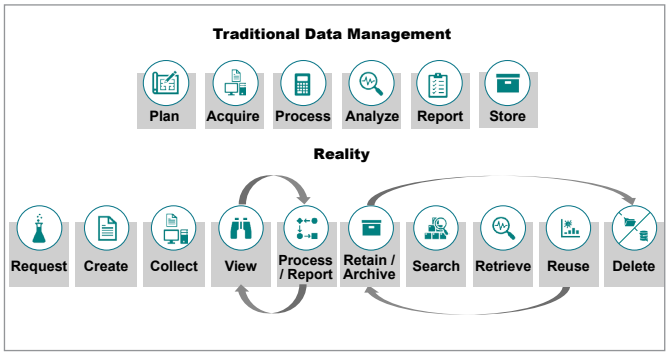

Figure 1. The stages of traditional data lifecycle management vs the stages of data lifecycle management today.

As we now appreciate the increased complexity in the stages of data lifecycle management, we are also able to investigate how each stage individually impacts data sustainability.

Stage 1 & 2*: Request & Create* - During these stages, key record documentation is created, resulting in the need for all relevant data to be captured at both these stages to ensure data sustainability.

Stage 3*: Collect* – At first, the raw data must be connected to the metadata to aid at Stage 4 (View) and later stages. Second, during collection of the raw data it must be clear what is the master data. Third, the storage location of this data must be considered. To ensure data sustainability at this stage, it's important to state the connection points between the data sets from each informatics system.

Stage 4*: View* – At this stage the relevant information for the data needs to be known in order to find the data to be viewed. This relevant information should not rely on a single person, team or department to recall the relevant information, but must be unambiguous and traceable from the start to the end of the data lifecycle.

Stage 5*: Process/Report* - Connecting the data and outcomes/results together at this stage allows for the relevant information to be documented accurately, supporting robust data sustainability.

Stage 6*: Search* – Knowing which terms to use to find the relevant data is made much simpler if the data lifecycle has been traced out so the appropriate data can be captured at the time and tagged onto the dataset.

Stage 7*: Retrieve* – When the data and records can be found back by searching in the informatics system, then these records can be reused again.

Stage 8*: Retain/Archive* – Cataloguing the data at the time of data storage is very important. Ensuring that archive location is secure, so the data can be recalled securely in preparation for stage 9.

Stage 9*: Reuse* – The ability to effectively reuse data to ensure the laboratory, is receiving the highest return on investment from its' data.

Stage 10*: Delete* – A large part of data governance and sustainable data is determining the end of life of data. This demonstrates that the data is not of use or relevant anymore (this can differ for the type of data, and is described in specific guidance).

HOW DOES ALCOA++ LEAD TO SUSTAINABLE DATA?

ALCOA++ is a set of principles for the life science industry to follow for intrinsic compliance (see Figure 2). Originally introduced by the US Food and Drug Administration (FDA), the ALCOA++ principles, when followed, are critical, but not limited to ensuring data integrity in life sciences laboratories. They provide clear guidelines on the elements to consider when pursuing data integrity and are becoming increasingly important to Current Good Manufacturing Practices (cGMP) particularly with the transition many labs are making to digital processes as part of Pharma 4.0.

As the amount of data generated by labs increases and companies begin to adopt data-driven technologies like automation, artificial intelligence/machine learning, and connectivity, the knowledge and data flow documentation around these technologies must also keep pace to assure and maintain data integrity and ALCOA++ principles.

When reviewing the ALOCA++ principles (see Figure 2), it's clear that data that doesn't adhere to these principles cannot be sustainable.

For example if the data is not legible, it is not usable for future decision making.

Implementing software products that have been developed to aid the user in adhering to the ALCOA++ principles, such as Empower LMS, can go a long way to removing the burden from effective data lifecycle management and data sustainability.

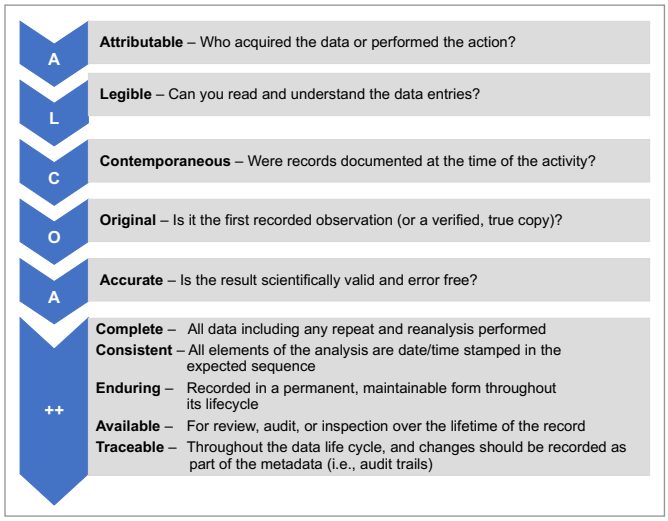

Figure 2. The ALCOA++ principles of data integrity.

REUSE TO REDUCE

As part of successful data lifecycle management and the goal to achieve data sustainability, data reuse serves as a critical element with significant implications. Beyond the initial collection and analysis of data, the ability to reuse data can offer enduring value. Effective data reuse is reliant on the accurate cataloguing, organization, and annotation of datasets to allow for future retrieval and application (think of ALCOA++ principles).

By reusing data, labs streamline their future work, reducing redundancy and optimizing resource utilization, ultimately increasing data sustainability.

Moreover, as datasets evolve and gain relevance over time, their reuse becomes integral to the iterative nature of scientific discovery.

Data lifecycle management, therefore, should embrace reuse as a strategic component, ensuring valuable information continues to contribute to new insights, fostering a sustainable and dynamic approach to data (and knowledge) generation.

Software tools can aid in ensuring effective data lifecycle management, for example Empower LMS has been built with ALCOA++ principles in mind. As a comprehensive laboratory workflow and documentation solution, it provides the ability to capture the relevant data at each stage (Figure 1) of the product lifecycle.

WHY DELETING DATA SHOWS CONFIDENCE IN YOUR DATA MANAGEMENT

While, on the one hand, reusing data is key to data sustainability and data lifecycle management, so is deleting data that is no longer relevant. Deleting data is part of a responsible and strategic data governance. It involves the removal of information that has reached the end of its relevant or legal retention period.

There are several key benefits to data deletion including:

- Ensuring compliance with privacy regulations
- Safeguarding against unauthorized access
- Streamlining storage resources

For example, if a laboratory is burdened with significant amounts of obsolete data, searching through that data will waste precious time and resources. Obsolete data, the retention of which is often not in line with standard operating procedures (SOPs), obscures current data which can cause significant issues, particularly at crucial times such as during an internal investigation or external quality audit.

Implementing a well-defined data deletion strategy within the broader data lifecycle management framework is crucial for maintaining a lean and efficient data environment while upholding data integrity and data sustainability.

Empower LMS has the ability to define the retention times of the data and securely delete the data when it has reached the end of that retention period. In addition, Empower LMS will clearly show the status of the data so the System Administration may always know where that record is at for the data lifecycle.

CONCLUSION

The effective management of laboratory data is critical in generating meaningful insights and maintaining data sustainability. The challenges faced by labs as they continue to generate increasing volumes of data highlight the importance of robust data lifecycle management. Recognizing the intricacies of each stage of the data lifecycle, from initial creation to ultimate deletion, is fundamental to ensuring data sustainability.

By aligning with principles like ALCOA++, embracing data reuse and deletion, and implementing informatics software, such as Empower LMS, which support robust data lifecycle management, laboratories can navigate the evolving data-driven technology landscapes to ensure the ultimate goal of truly sustainable data.

For more information, visit waters.com/EmpowerLMS.

References

1. 21CFR11: ELECTRONIC RECORDS; ELECTRONIC SIGNATURES.
2. 21CFR211.180: CURRENT GOOD MANUFACTURING PRACTICE FOR FINISHED PHARMACEUTICALS.
3. EU GMP Part 1, Ch4: Documentation.
4. Annex 11: Computerised Systems.
5. Using Empower LMS to build Intrinsic Compliance and Durable Data.

Waters™

Waters Corporation
34 Maple Street
Milford, MA 01757 U.S.A.
T: 1 508 478 2000
F: 1 508 872 1990
www.waters.com