Focus on Quality

A Flexible Analytical Data Life Cycle?

Does a one-size-fits-all data life cycle meet laboratory requirements for data integrity? No! The reason is that there are many different types of analytical procedures, which means that a flexible analytical data life cycle is required.

R.D. McDowall

One of the requirements of the data integrity guidances for regulated laboratories is a data life cycle that covers the birth to death of regulatory records. The data life cycle is defined in the recent MHRA data integrity guidance as "All phases in the life of the data from generation and recording through processing (including analysis, transformation or migration), use, data retention, archive/retrieval and destruction" (1).

While this is rather vague and needs a degree of interpretation, it gives an outline for a data life cycle.

A Generic Data Life Cycle, 1

After the publication of the MHRA GMP guidance in 2015 (2), I developed a data life cycle (3) that consists of two phases, active and inactive. The active phase is where most of the laboratory work occurs from acquisition to data use and short-term retention, but this is the shortest part of the life cycle. The inactive phase is where the data and records are stored for the remainder of the record retention period.

The active phase of the data life cycle consists of the following tasks:
- Data Acquisition: Controlling and recording the observation or generating data
- Data Processing: Interpretation or processing of the original data
- Generate Reportable Result: Calculation of the reportable result
- Information and Knowledge Use: Using the result to make a decision

- Short-Term Retention: Secure storage of the data and information in a secure but accessible environment for any further use—for example, audits, trending, and so on. Note that GLP requires that all study data be placed in the GLP archive at the end of the study (4,5).

The inactive phase of the data life cycle consists of the following tasks:
- Long Term Archive: The movement of the records into a secure location
- Data Migration: Migration of data from one system or repository to another over the retention period, only necessary for electronic records and not paper
- Data/Record Destruction: The formal process of destruction at the end of the retention period.

A Generic Data Life Cycle, 2

Another data life cycle interpretation was published in the GAMP Guide for Records and Data Integrity (6). This interpretation, which is consistent with the WHO data life cycle (7), defined a generic life cycle consisting of five phases:
- Creation: Acquisition of data
- Processing: Transforming and interpreting the data and calculation of the reportable result
- Review, Reporting and Use: Consisting of data review, audit trail review, data reporting, data distribution, and use of the information
- Retention and Retrieval: Covering the availability of the records, the security to ensure that unauthorized changes
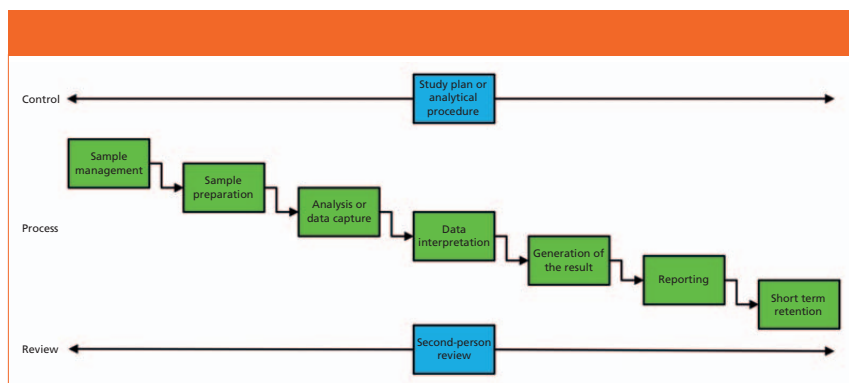
Figure 1: The active phase of an analytical data life cycle. (Adapted with permission from reference 8.)

are not made, backup and recovery, physical storage, and security and archiving, as needed

- Destruction: The formal process of destroying the records at the end of the record retention period.

My main criticism of this model is that review, reporting, and use are crammed into a single activity. Second-person review is an integral part of the creation and processing portion of laboratory work to ensure the quality and integrity of reportable results.

Generic Data Life Cycles Don't Work in the Laboratory

Generic models are good in that they provide a simple basis for understanding a data life cycle, but both models described above (3,6) may be too simplistic to apply to all situations, especially in a GXP regulated laboratory. I believe that an analytical data life cycle must be sufficiently granular for a laboratory environment. Where is sample management and sample preparation? Are these key analytical aspects all bundled into "data acquisition" or "capture"? Furthermore, where do the controlling elements of a life cycle, such as a study plan, validation plan, or analytical procedure, come into the life cycle? All are key requirements for any regulated laboratory.

I suggest that the two life cycles above are not suitable for regulated laboratories, and that an alternative approach is required.

You may think that I am being very critical of the two data life cycle models above, but if you are going to man-

age data integrity in a laboratory, you need to have a good understanding of each process down to a data and record level. These two models are not sufficiently detailed, and therefore we need an analytical data life cycle. The analytical data life cycle described below is essentially an expansion of the first data life cycle (3), adapted to chemical analysis.

An Analytical Data Life Cycle

The analytical data life cycle presented here still consists of two phases, active and inactive. The active phase of the analytical data life cycle, shown in Figure 1, consists of three subphases (8):

- Control: A study plan, an analytical procedure, and an analytical request that defines the work to be performed
- Process: The analytical process from sampling to reportable result
- Review: Second-person review to ensure the quality and integrity of the result.

Each task within each subphase of the life cycle will be described in more detail below.

Controlling the Analytical Data Life Cycle

In all instances, there needs to be control of the active phase of the analytical life cycle, and this control is shown in Figure 1 as either a study plan or an analytical procedure to define how work will be conducted. Even method development under the quality by design (QbD) approach advocated by the draft *USP* chapter <1220> (9) will have control by defining the analytical
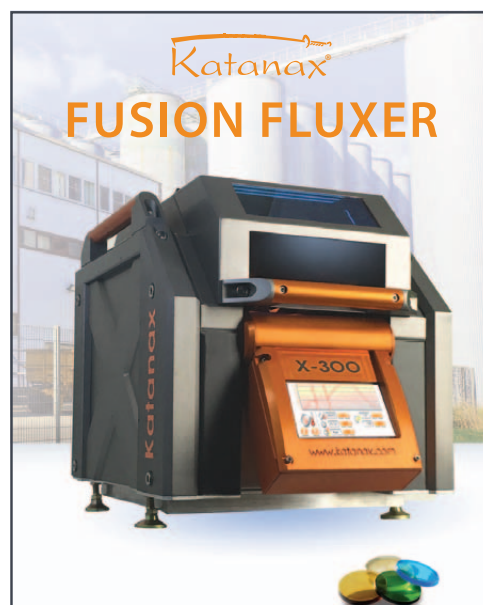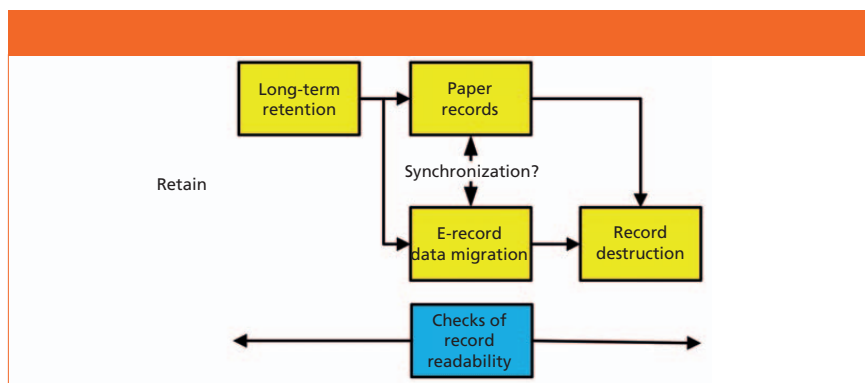
Figure 2: The inactive phase of an analytical data life cycle. (Adapted with permission from reference 8.)
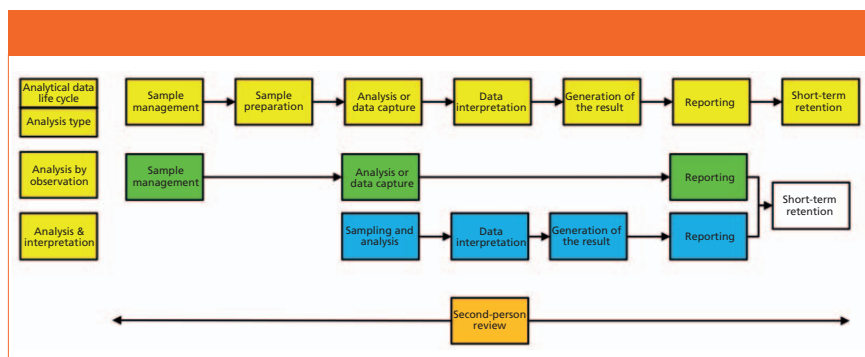
Figure 3: Different analytical procedures require a flexible analytical data life cycle. (Adapted with permission from reference 8.)

target profile (ATP) and controlling the development of a procedure. The plan or procedure must be considered as an essential part of the analytical data life cycle, because it directly determines the data that will be collected and processed. In addition to these controlling documents of the analytical life cycle, there will be standard operating procedures or work instructions for performing component tasks within the overall life cycle. These are not considered in this discussion, but are an essential part of performing and reviewing the work carried out in any analytical laboratory.

Tasks of an Analytical Data Life Cycle
In Figure 1, underneath the study plan or the analytical procedure is the main analytical data life cycle from sampling to generating reportable results. That life cycle consists of six analytical tasks together with short-term data storage.

Sample Management
Sample management covers a sampling plan, sampling process, defining the sample containers to be used, sample preservation requirements, transport, and storage in the laboratory. Because sampling is the most critical part of the analytical process, it must be performed correctly with the right documentation. The integrity of the data gathered here is essential to support the final results of the analysis and these results will include any environmental monitoring records of storage conditions during transport or storage. The major problem is that much of the sampling process is usually manual, can contain errors, and can be falsified easily. This life cycle task is not discussed in any data integrity guidance.

Sample Preparation
Preparing the samples for analysis can be as simple as transferring a liquid sample to a vessel and presenting it to an instrument, through dissolving and diluting, to complex liquid–liquid or solid-phase extraction. Although some sample preparation techniques can be automated, many of the steps are manual and typically recorded on paper.

The scope of work may include preparation of reference solutions, buffers, and mobile phases using instrumentation such as sonic baths, analytical balances, pipettes, volumetric glassware, homogenizers, and pH meters. Data demonstrating that this work has been performed are essential for demonstrating both the integrity and quality of the work, including appropriate instrument calibration checks and associated instrument log book entries. Like sampling, the sample preparation task is not covered in regulatory or industry guidance documents on data integrity.

Newton and McDowall, in an article series about data integrity in the chromatography laboratory, have discussed both sampling and sample preparation in more detail for *Spectroscopy*'s sister publication, *LCGC North America* (10).

Analysis or Data Capture
The spectrum of analytical techniques applied to a sample can vary from observation (color, appearance, or odor) through wet chemistry, such as loss on drying and water content to instrumental techniques such as spectroscopy or chromatography. This task can vary from a simple observation through to the setup of a spectrometer with an appropriate software configuration to protect records, calibration, and point-of-use checks. This step is followed by the acquisition of data from the sample by following the applicable analytical procedure. The data collected here will include, as appropriate, the instrument setup, any system suitability tests or point of use checks before committing the samples for analysis, data values or data files for interpretation in the next stage of the analytical data life cycle, or just the result. Both sampling and sample preparation can be automated using either an electronic laboratory notebook (ELN) or laboratory execution system (LES).

Newton and McDowall have discussed this task in their article series on data integrity in the chromatography laboratory, but the principles described can be easily adapted for spectroscopic analysis (11).

Data Evaluation or Interpretation
In this step, the data acquired during the analysis are interpreted to obtain

processed data. Where an instrument such as a spectrometer is used, the data need to be interpreted by an analyst to obtain an identity, absorbance at a specific wavelength, or peak area counts. It may also involve the comparison of a sample spectrum with a spectral library to confirm the identity of a sample. This key task of the analytical data life cycle needs to be controlled carefully to ensure the integrity of the data. This task is also the subject of rigorous regulatory scrutiny and is the source of many data integrity citations and warning letters. This is the subject of the third part of Newton and McDowall's discourse on data integrity (12).

Generation of the Results

Following the interpretation of the data, the next task is the generation of the reportable results. This calculation can be made by a variety of means, such as manually using a calculator, using a spreadsheet, or incorporating the data into an instrument data system or other informatics application. Where possi-

ble, calculations should be performed by a validated software application, thus avoiding manual data entry. At this stage, the outliers can be identified, such as out-of-specification (OOS) results or a value on a time-versus-drug concentration curve for further investigation (OOS results, for example) (13,14).

Reporting

Following the calculation of the results, the next task in the life cycle is reporting. There are many forms that the report can take, such as a method validation or transfer report, certificate of analysis (COA), or study report. Newton and McDowall have discussed calculation of the reportable results in the fourth part of their data integrity series (13).

Second-Person Review

Before a report or COA is formally issued, the complete data package needs to be subject to a second person review. This is a critical subphase of the analytical data life cycle. The laboratory

reviewer needs to be suitably trained and the review will include any instruments and computerized systems involved in the analysis. The aim of the second-person review is to ensure that the work has been carried out correctly, procedures have been followed, data have been interpreted correctly, results have been generated accurately, and the report is complete. In addition, the second-person reviewer needs to check that there have not been any data falsification or poor data management practices. The penultimate part of Newton and McDowall's series in *LCGC North America* covers the review of analytical data in more detail (15).

Short-Term Retention

Data are retained in a secure manner regardless if the records are paper or electronic. Often, the complete set of data is hybrid; consequently, there will be paper and electronic records that need to be synchronized if any changes are made after the analysis is completed and reported, such as in

response to complaints or regulatory questions.

When the short-term data retention period has elapsed, the data are retained for the applicable retention period mandated either by the regulations or company policy.

Implicit throughout the whole of the analytical process is the applicable standard operating procedures, work instructions that describe how work should be conducted by the analytical staff. These should include how results are trended in compliance with EU GMP Chapter 6 requirements (16) and the identification of OOS or out-of-trend or out-of-expectation (OOT and OOE) results (14).

Inactive Phase

The inactive phase consists of long-term retention, possible electronic record data migration, and the destruction of the records, as shown in Figure 2. In contrast to the earlier model (3), in this model, this phase of the data life cycle (8) separates paper and electronic records. Typically, paper records are stored in an archive or record storage and are not migrated or moved during the retention period. Electronic records may have to undergo data migration during the retention period. Throughout the inactive phase, check that the records are available and can be accessed. This step applies especially to the paper and electronic records from hybrid systems, and must include the synchronization between the two media, and is the most difficult task in this phase of the life cycle.

One Analytical Size Fits All? No!

Although the analytical data life cycle in Figure 1 looks good, there is one small problem: It does not fit all analyses. The problem is that a "one-size-fits-all" approach does not work because there are many ways to analyze a sample. Therefore, this life cycle needs one more attribute to make it applicable in any regulated laboratory: flexibility. The analytical data life cycle must be able to expand and contract to fit an individual analytical process. For example:

• If a sample does not need any preparation stage before testing, will a data life cycle accommodate this situation?

• If the data acquisition is the result, why do we need the interpretation or calculation phases?

To understand this process better, look at Figure 3. This figure shows six analytical phases of the analytical data life cycle model across the top, and there are two analytical procedures modeled below it. The first is analysis by observation, where a sample is taken and the analysis is by observation of color, odor, or appearance. The observation is the reportable result which may be compared with a specification for release. The life cycle is minimal.

The second analysis shown in Figure 3 is instrumental analysis, followed by data interpretation, which is typified by near-infrared identity analysis. The sample management and sample preparation tasks of the analytical data life cycle are minimal (the number of containers to test), because the analysis is usually performed in situ in a warehouse, and the sampling and analysis phases are united in one step as the spectrum of the sample is compared to a composite spectrum in a spectral library.

Summary

As can be seen in Figure 3, an analytical data life cycle must be flexible and must be adapted to meet any analytical procedure and the data generated by it. This approach is far preferable to forcing all processes to fit a one-size data life cycle.

Acknowledgments

I would like to thank Mark Newton and Kevin Roberson for their review comments during preparation of this column.

References
(1) MHRA GXP Data Integrity Guidance and Definitions, Medicines and Healthcare Products Regulatory Agency (London, England, 2018).

(2) MHRA GMP Data Integrity Definitions and Guidance for Industry, Medicines and Healthcare Products Regulatory Agency (London, England, 2nd Ed., 2015).

(3) R.D. McDowall, *Validation of Chromatography Data Systems: Ensuring Data Integrity, Meeting Business and Regulatory Requirements (*Royal Society of Chemistry, Cambridge, UK, 2nd Ed., 2017).

(4) 21 CFR 58 Good Laboratory Practice for Non-Clinical Laboratory Studies. 1978, Food and Drug Administration: Washington, DC.

(5) OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1, OECD Principles on Good Laboratory Practice. 1998, Organization for Economic Co-operation and Development, Paris, France.

(6) GAMP Guide Records and Data integrity. 2017, Tampa, FL: International Society for Pharmaceutical Engineering.

(7) WHO Technical Report Series No.996 Annex 5 Guidance on Good Data and Records Management Practices. 2016, World Health Organization: Geneva.

(8) R.D. McDowall, *Data Integrity and Data Governance: Practical Implementation in Regulated Laboratories* (Royal Society of Chemistry, Cambridge, UK, 2018).

(9) G.P. Martin et al., *Pharmacopoeial Forum,* 2017. **43**(1), (2017).

(10) M.E. Newton and R.D. McDowall, *LCGC North Am.* **36**(1), 46–51 (2018).

(11) M.E. Newton and R.D. McDowall, *LCGC North Am.* **36**(4), 270-274 (2018).

(12) M.E. Newton and R.D. McDowall, *LCGC North Am.* **36**(5), 330–335 (2018).

(13) M.E. Newton and R.D. McDowall, *LCGC North Am.* **36**(7), 458–462 (2018).

(14) FDA Guidance for Industry Out of Specification Results. 2006, Food and Drug Administration: Rockville, MD.

(15) M.E. Newton and R.D. McDowall, *LCGC North Am.* **36**(8), 527–531 (2018).

(16) EudraLex - Volume 4 Good Manufacturing Practice (GMP) Guidelines, Chapter 6 Quality Control. 2014, European Commission: Brussels, Belgium.

R.D. McDowall is the Director of RD McDowall Limited and the editor of the "Questions of Quality" column for *LCGC Europe, Spectroscopy's* sister magazine. Direct correspondence to: SpectroscopyEdit@ubm.com

For more information on this topic, please visit our homepage at: www.spectroscopyonline.com