Data Integrity with NIR-Spectroscopy Software

Nicolas Rühl

Data Integrity is currently a hot topic issue that has created much attention and has raised concern within companies working in regulated environments. This White Paper explains some of the key terms used in the context of Data Integrity and outlines how the requirements of Data Integrity can be understood and implemented.

Metrohm International Headquarters; Herisau, Switzerland

Metrohm

Introduction

Recent news about various Data Integrity violations revealed by governing bodies has created much attention and has raised concern within companies operating active in regulated environments. Figure 1 shows the increase in violations of data integrity aspects during cGMP (current good manufacturing practice) inspections. (1)

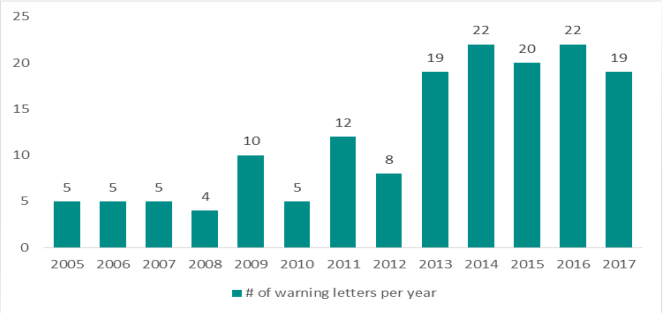

Figure 1. Number of data integrity violations over the last years.

To help understand the topic Data Integrity and its requirements, this White Paper provides an overview of important terms used in this context and presents examples of software functionalities which help to fulfill requirements stipulated in the FDA guidelines for data integrity. (2)

Data Integrity Terms – An explanation

ALCOA/ALCOA+

Data integrity includes several aspects which are summarized under the acronym ALCOA and ALCOA+. ALCOA stands for the requirements attributable, legible, contemporaneous, original, and accurate. This set of requirements has been recently extended by the aspects complete, consistent, enduring, and available (ALCOA +). A detailed explanation of these requirements with examples can be found in the chapter The ALCOA Principle.

Data and Metadata

The FDA distinguishes in their Data Integrity guideline between "data" and "metadata". Data is understood as a collection of facts, which can, for example, be a number, a list of measurements, or a record. Metadata, or contextual data, on the other hand, provides additional information describing data, for example, by whom and when a measurement was done. For Data Integrity and more specifically for the requirements of transparency, consistency, and reproducibility, both data types are always required (see figure 2).

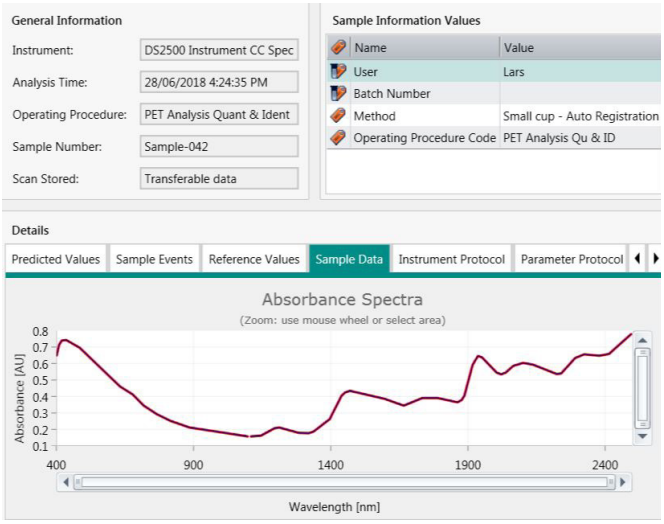

Figure 2. Display of metadata in Vision Air for a single measurement. Time stamps, user names, operating procedure type, to name only a few, are displayed together with the collected spectrum and the predicted results.

Dynamic and static data

Apart from the distinction between data and metadata, data can be classified as being dynamic or static. Static data includes all non-changeable data, such as records on paper or stored as PDF. Electronic records, on the other hand, where user have the possibility to make changes, are defined as dynamic. While the latter offers users convenience, it also involves a risk to lose transparency. This risk needs to be accounted for, e.g., by means of automatic history tracking.

Figure 3 shows a typical example of how dynamic data can be tracked. Signing events (signing on different levels, adding comments, or withdrawing signatures) are recorded automatically in the Vision Air software.

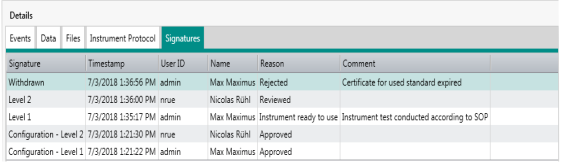

Figure 3. Display of changes for dynamic data. All signing and withdrawing steps are clearly documented, allowing users to track the individual steps performed.

The ALCOA Principle

The ALCOA and ALCOA+ requirements are the basis to achieve data integrity. All these requirements have a significant overlap with requirements stated in the FDA 21 CFR Part 11 regulation. Therefore, when trying to achieve Data Integrity, it is advisable to use software which has been developed in accordance with 21 CFR Part 11 requirements to simplify the effort. However, it is important to highlight that similar to achieving compliance with the 21 CFR Part 11 guideline, software functionalities do only cover parts of these requirements, more precisely technical ones. That's the reason why compliant-ready software can only act as a solid base to achieve Data Integrity. To achieve Data Integrity process, aspects must also be defined within the company, which is typically done by implementing standard operating procedures.
In the following chapters, each aspect is first explained briefly and then a practical example based on Metrohm's Vision Air software is presented.

Attributable – When did who do what?

To conform with Data Integrity requirements, workflows need to be traceable. As mentioned in the previous chapter, a combination of both data and metadata is required. The requirement of attributability focuses on the availability of metadata, specifically the user ID and the time stamp for each measurement.

Figure 4 shows the typical sample registration information field created for each measurement in Vision Air. Each measurement follows a predefined operating procedure, which provides guidance to the operator. The time stamp and user information by whom and when a measurement was performed is tracked. The user is identified by his unique user ID and a user name. For additional information, user-defined fields can be added. In this example, an input field for a batch number was created.

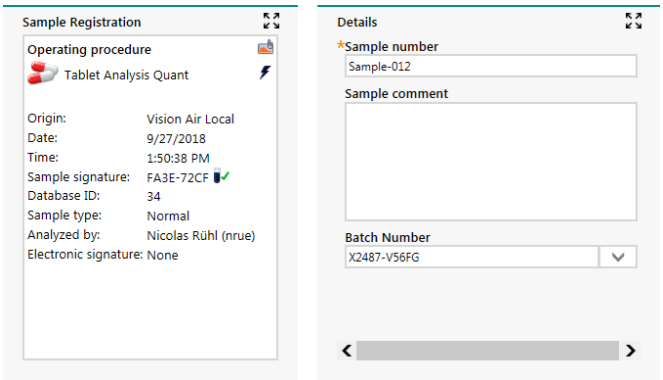

Figure 4. Sample registration window during routine analysis in Vision Air. Both automatically created and user-entered metadata are available for each measurement.

Legible – Do data remain readable through the complete lifecycle?

The requirement of legibility states that data and metadata be readable and permanent. While this may sound trivial, the implementation can be challenging, especially with regard to long-term readability. A common approach for this requirement is to offer the possibility to print or to export data into frequently used formats.

Vision Air provides export options for reports (PDF, XML, and CSV). Furthermore, the Vision Air automatic backup scheduler offers users a convenient way to back up the entire database, which can be reimported at any time (see Figure 5). Both features guarantee long-term readability.

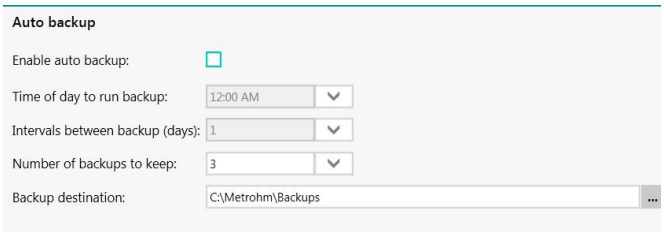

Figure 5. Automatic backup screen scheduler in Vision Air. Vision Air allows users to define an automatic backup procedure to ensure long-term readability.

Contemporaneous – Are records created at the time of the activity?

The paper-based documentation approach, which was common in the past, enabled back-dated entering of data. This approach always involves a significant risk of unprecise and false documentation, since information can be forgotten or misremembered. The nowadays more established form of electronic documentation eliminates this risk by creating records ideally at the time an activity takes place.

In Vision Air, each measurement is instantly stored in the SQL database. Successful storage is clearly displayed with icons (see figure 6). In comparison to a file-based system, the storage location and the data structure is set automatically, which guarantees full functionality of the system.

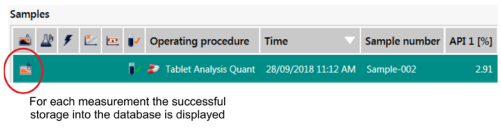

Figure 6. Visual display of successful storage of data. In the sample history view in Vision Air, an icon displays the successful storage of data in the SQL database.

Original – Is this the first recorded observation or a verified, true copy?

Result values collected during analyses are typically dynamic data sets, thus allowing post- or reprocessing activities. In NIR spectroscopy, post-processing is more common than reprocessing, which is a typical activity performed, for example, for titration or ion chromatography data sets. Typical post-processing operation of data in NIR spectroscopy are slope/bias correction or general post-calculation. For the sake of transparency, each of these calculations and the individual results must be clearly displayed.

In Vision Air, users have the possibility to display both the unprocessed (raw) result and the processed result. If configured, post-processing is done automatically during result calculation, thus avoiding human errors or manipulations. Figure 7 shows a typical detailed result display in Vision Air, where the raw, unprocessed result and the post-processed result from a slope/bias adaptation is clearly displayed for the parameter moisture.

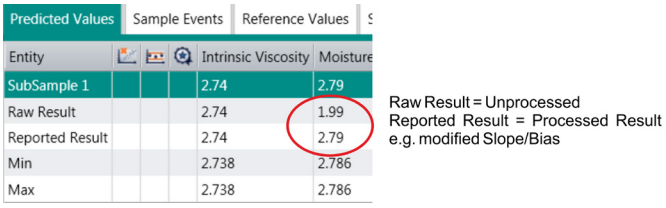

Figure 7. Detail result display. In the detail view in Vision Air, the unprocessed and the post-processed result (e.g., a slope and bias correction) are displayed for each measurement.

Accurate – Is data stored correctly and are modifications clearly documented?

Alteration of data, if allowed, e.g., of dynamic data, has to be documented transparently. For configuration changes of the instrument, extensive review functionalities, typically a two-level signing function as recommended in the 21 CFR Part 11 regulation, should be in place to ensure and help with accurate documentation.

Vision Air Pharma, which was developed to fulfil the requirements of the FDA 21 CFR Part 11 regulation, includes a two-level signing system for every configuration change (see figure 8). Successful application of a configuration change is displayed using unambiguous icons, and each configuration change is listed in a table for later review. Comments can be entered for each modification made to increase further transparency.

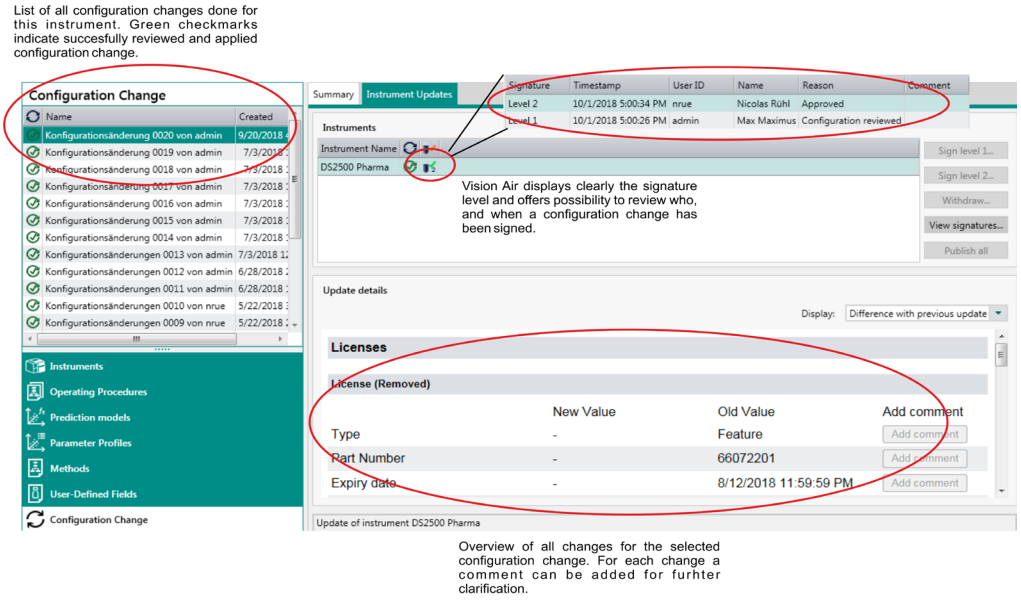

Figure 8. Configuration change viewer in Vision Air. The configuration change viewer in Vision Air allows users to track all changes of the system in a transparent manner.

The ALCOA + extension

ALCOA+ extends the requirements to include aspects such as completeness, consistency, and availability of data:

Complete – Is all relevant data properly stored?

With the change to electronic documentation, assessing whether data has been properly stored becomes more difficult. For both kinds of storage (file-based as well as a data-based), the software support users by showing clearly that data storage was successful.

The SQL database which is used in Vision Air for data storage has the advantage to provide a higher level of data security as opposed to a file-based concept. The risk of corrupted data sets or unauthorized data manipulations can be significantly reduced by using a database system rather than a file-based system. In the Vision Air software, each measurement is linked with a unique signature (see figure 9) indicating that data is not only correctly stored within the database but has also not been modified.

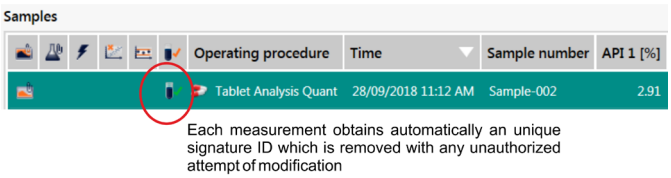

Figure 9. Visual feedback of correct storage of data in the database. The icon highlighted with a red circle indicates a correct hash signature.

Consistent – Can the operation workflow be reconstructed?

The above mentioned requirements all seek to make sure workflows are transparent. However, not only from the point of view of transparency, but also from the point of view of efficiency, workflows should be consistent and easily reconstructible.

Workflows in Vision Air can be easily defined by using so-called operating procedures guiding operators during measurement. Each operating procedure consists of a set of prediction models and an element called method, which defines exactly how measurements are to be conducted (see figure 10).

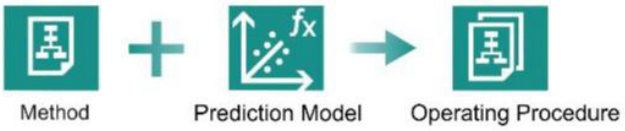

Figure 10. Structure of operating procedures in Vision Air. An operating procedure defining how a measurement is done consists of prediction models and a method which specifies different measurement parameters.

A method can, for example, specify the sample temperature for a measurement (see figure 11), which input fields are accessible, and which input fields are mandatory to be filled.

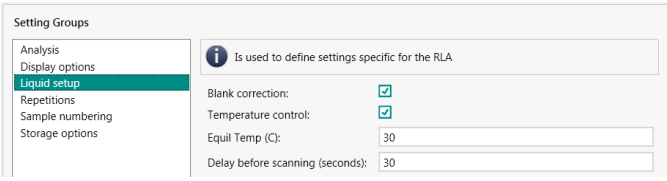

Figure 11. Method setting options for liquid analysis. In the settings for a method, different measurement parameters can be set. For example, the temperature of the sample vessel for liquid analysis.

Enduring – Does stored data remain available longtime? Available – Is data accessible / readable / printable?

The final two aspects mentioned in the FDA Data Integrity guideline are enduring and available. Enduring and available further highlights the importance of longtime availability of stored data and the possibility to access and read data in common ways.

Both aspects are covered by the previously mentioned SQL database storage approach and the availability of audit trails (see figure 12), backup functionalities, and automatic print functions. In the Vision Air audit trail, all relevant activities can be easily reviewed using the integrated filter features.

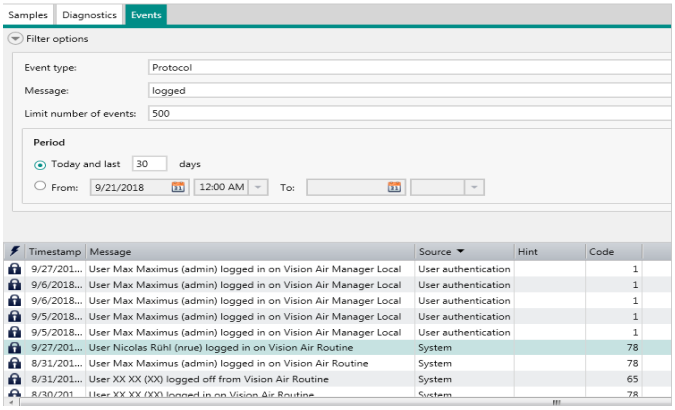

Figure 12. Audit trail view for login events. In Vision Air, different filter settings can be set to find events easily, for example, all events which refer to a login or logout.

Summary of important software functionalities for Data Integrity

The previous chapter discussed many different software functionalities. In general, it can be said that one of the most important software functions to achieve and maintain operation transparency are audit trails. Audit trails allow for the reconstruction of operations performed by clearly displaying by whom, when, and why actions were taken. Besides automatically generated information, such as user ID and time stamps, functions such as comment fields in combination with preselected reasons of action should be available. Besides transparency, data safety is an important aspect of data integrity. This includes mainly data storage and manipulation control. Increased data security can be achieved by following a database approach rather than file-based approaches, where files, for example, can be more easily deleted or modified. Databases furthermore allow for easier tracking of data manipulation using unique signatures for each collected data.

Literature:

(1)[http://www.nsf.org/newsroom_pdf/pb_data_integrity_closer_look.pdf]
(2)[https://www.fda.gov/downloads/drugs/guidances/ucm495891.pdf]

WP-042EN, published November, 2018