Revisiting Data Integrity and Chromatography Data System (CDS) Software

by Timothy Cross, EMEA Regional Marketing Manager – HPLC, Thermo Fisher Scientific, Hemel Hempstead, UK and Darren Barrington-Light, Software Marketing Specialist, Thermo Fisher Scientific, Altrincham, UK

Data integrity is a key concern for drug, pharmaceutical ingredient and medical device manufacturers and key to Good Manufacturing Practice (GMP) to ensure that final products are of high quality and safe to the patient. The shift from traditional paper and ink to computerised data systems in the pharmaceutical industry has also meant Good Automated Manufacturing Practice (GAMP) has become more and more important. One of the core principles of GAMP is that quality cannot be tested into a batch of product but must be built into each stage of the manufacturing process.

Data integrity is nothing new and has always been fundamental to GMP and GAMP. Data integrity covers all aspects of production, not just the computerised systems - from the company policies, to organisational procedures such as Standard Operating Procedures, through to the training of staff. However the topic has gained renewed momentum in the past couple of years as regulatory bodies have increased their focus on data integrity during inspections due to numerous recent breaches by manufacturers, for example Italian API producer Trifarma S.p.A. in July 2014 and Indian API facility Wockhardt Ltd in July 2013. Breaching data integrity guidelines often leads to serious consequences including product recalls, expensive audits and penalties, as well as damage to the company's reputation.

With this increased focus come extra demands on companies to ensure the integrity of their data to meet current guidelines. In this article we will review the current guidelines issued by the UK's Medicines and Healthcare Products Regulatory Agency (MHRA) on data integrity, the expectations of these guidelines for computerised systems and what requirements this might place on your Chromatography Data System (CDS) software.

Data Integrity Regulatory Framework

Globally, data integrity is regulated by a number of different agencies; in the US this is by the Food and Drug Administration the European Council within Europe and the MHRA in the UK. At the start of the year the MHRA offered new guidance on data integrity, entitled 'MHRA GMP Data Integrity Definitions and Guidance for Industry March 2015' [1] to complement the existing European Eudralex Vol 4 standards [2]. This MHRA guidance identifies opportunities to strengthen both paper and computerised elements of the data lifecycle and sets out the following five expectations that all data must be:

- A – Attributable to the person generating the data
- L – Legible and permanent throughout the data lifecycle
- C – Contemporaneous (i.e. recorded at the time of the event)
- O – An Original record
- A – Accurate

So how are these five expectations applied to the various areas described by the MHRA for data integrity requirements, how do they relate to your CDS software and, most importantly, is your CDS up to the task?

Raw Data, Metadata, Audit Trails, and User Access

MHRA guidelines describe the expectation that raw data must be contemporaneously and accurately recorded, be legible and accessible throughout the data lifecycle, and permit full reconstruction of the activities resulting in the generation of the original data. In order to do this metadata must be recorded. Metadata describes the attributes of other data providing context and meaning to the original data. It also permits data to be attributable to an individual making metadata an integral part of the original record.

The most important form of metadata is audit trails. They record critical information which, in turn, permits the reconstruction of the original process or activity, making them fundamental to data integrity. Audit trails should always be contemporaneously recorded tracking any changes to the data showing who did what, when, and why, making the data attributable to the person making those changes. In order to log who is doing what, user access controls must be in place. These ensure users cannot, for example, amend or disable audit trails or delete data, and that they only have access to functionality that is appropriate for their job role.

CDS software requirements

To meet these expectations your CDS software needs to store all metadata from every operation, such as instrument control, run time events, data object changes, through to user management actions and link it to the original data (Figure 1). This is normally achieved using the protected data storage provided by a relational database such as Microsoft® SQL® Server or Oracle. Some CDS's will prescribe a specific database type, but others offer more flexibility in the database selection that may fit better with your existing IT infrastructure.

Without a relational database it can be difficult to preserve the integrity of metadata or to maintain audit trails to track any changes. Your CDS software should automatically record all activities within

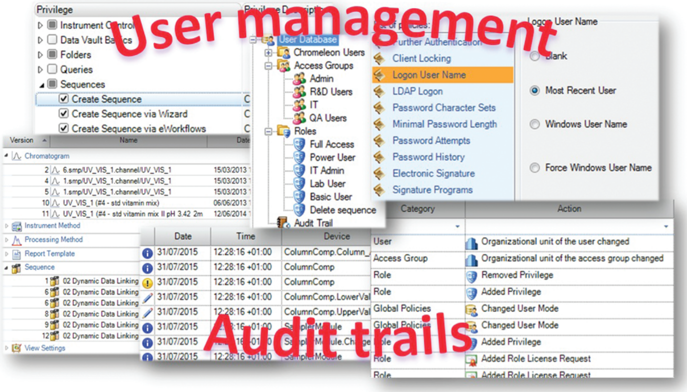


Figure 1: Example of audit trails and user management tools in a commonly used CDS software

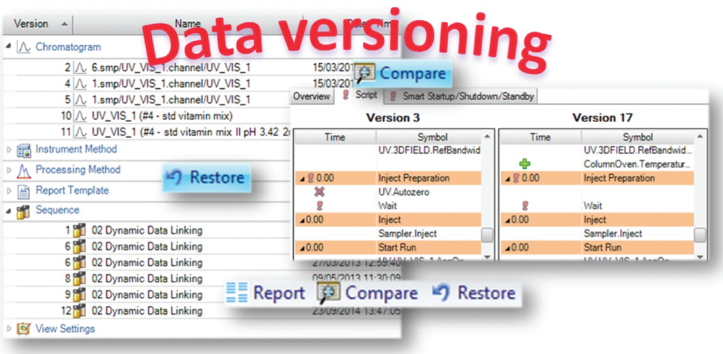


Figure 2: CDS software version comparison tool features

the CDS and provide the ability to view and print these logs. Nowadays it is also common that discrete versions of data objects are automatically created during the software workflow to allow full reconstruction of the activities that generated the data, as discussed next.

Computer System Transactions

A computer system transaction is termed as a single operation or sequence of operations performed as a single 'unit of work'. The individual operation(s) that make up a transaction do not need to be individually saved until the user commits the transaction through a deliberate act, such as pressing the 'save' button, however the computer system should ensure that the execution of critical operations are recorded contemporaneously by the user and not combined into a single computer system transaction with other operations.

CDS software requirements

This is commonly achieved by creating 'versions' of data objects and storing these alongside the original data. There are two possible ways to do this – transactions or versions are automatically created after every individual action, or a version is created when the user actually commits to changes in the current session.

The former option ensures that every single action is immediately recorded as a new version which, while providing a complete history of an object, also creates a huge number of versions. The latter reduces the number of versions created by grouping all the actions together for each new version (Figure 2).

For example, let's say you open a chromatogram and re-integrate by changing a detection parameter such as the integration range or moving a peak baseline, and then change it back to the original setting or discard the change without saving. Should there now be two new versions of the data or just the original file? Is it necessary to record each and every action? It can be argued that the second option fits best with the MHRA requirement for computer system transactions.

File Structure

When it comes to file structure there are two types; flat file and relational database. The flat file structure stores data as individual records which often don't contain all the relevant metadata thereby presenting a greater data integrity risk since data could be manipulated or even deleted without tracking. Conversely a relational database file structure is much more secure as it stores the data and metadata in different places but maintains the relationship between them. This makes it inherently more difficult to selectively delete, amend or recreate the original data and the metadata trail of actions.

CDS software requirements

As previously discussed, to provide maximum data integrity, your CDS software should utilise a relational database rather than a flat file structure. In order to retrieve data from the database most CDS's will provide a database search or query tool. However, the flexibility to quickly and easily find data, including searching on metadata, collate it and then view and utilise that data, for example, to create instrument utilisation statistics or investigate laboratory-wide out of spec results, can be very desirable.

Data Integrity, Lifecycle, Retention and Archiving

Data integrity arrangements must be in place to ensure that the accuracy, completeness, content and meaning of data are retained throughout the data lifecycle. The data lifecycle includes all phases in the life of the data, from its initial creation through processing, use, data retention, archival and retrieval, and eventual destruction.

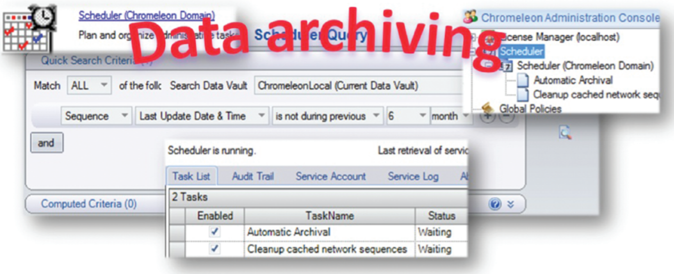
Data retention arrangements (classified as either archive or backup) must be designed to protect records from deliberate or accidental changes or deletion thus ensuring the data integrity of the record throughout the retention period. Data archiving is defined as the long term, permanent retention of completed data and relevant metadata in its final form, whereas a backup is a copy of current data, metadata and system configuration settings for the purpose of disaster recovery. Archived records may need to be stored for many years and must be permanently locked such that no changes can be made without detection or audit trail. In addition to this, at least two years of data must be easily retrieved for regulatory inspections.

CDS software requirements:

Most CDS software provides tools to assist with data retention. The ability to lock injections, sequences, folders, or even data stores can ensure data integrity as well as the use of electronic signatures as legally binding equivalents of an individual's handwritten signature.

Archiving may be as simple as automatically moving data from online to offline storage, for example, after a period of time where the data has been unused. This ensures the data is easily and immediately accessible for regulatory inspection. The flexibility of your CDS query tool to automatically and accurately identify the records to be archived and a tool to schedule and perform the transfer can greatly simplify this task (Figure 3).

Extensive backup tools are not common in CDS software since a system-wide backup


Data archiving
Scheduler (Chromeleon Domain)
Plan and organize administrative tasks
Scheduler Query
Quick Search Criteria
Chromeleon Administration Console
License Manager (localhost)
Scheduler
Scheduler (Chromeleon Domain)
Automatic Archival
Cleanup cached network seq
Global Policies
and
Task List
Audit Trail
Service Account
Service Log
Computed Criteria (0)

Match ALL of the folk Search Data Vault ChromeleonLocal (Current Data Vault)

Sequence | Last Update Date & Time | is not during previous | 6 | month

Scheduler is running. Last retrieval of servic

2 Tasks

Enabled	TaskName	Status
✓	Automatic Archival	Waiting
✓	Cleanup cached network sequences	Waiting

Figure 3: An example CDS software query and scheduler tool for data archiving

for disaster recovery normally requires significant design, implementation, and validation input from the local IT experts. Each company will typically have their own validated processes for disaster recovery although CDS vendors will usually provide guidance during the CDS installation and configuration. Most CDS software will provide backup facilities for smaller portions of data to enable data sharing and transfer. Generally these tools are not designed for disaster recovery.

Summary

Data integrity has always been critical for GMP compliance and a good CDS will provide tools and controls to assist you in ensuring the integrity and quality of your data and to meet the five expectations that all data must be attributable, legible, contemporaneous, original and accurate. However it is important to remember that the data lifecycle may begin and finish outside of the CDS. There may be metadata that needs to be linked to or used with the data inside the CDS, such as weights or dilutions, or the CDS results themselves may be metadata for the next step in the analysis. Whatever the data workflow, you must ensure the integrity of all your data and have relevant procedures in place to meet regulatory standards.

Implementing and validating CDS software across a facility can be a time consuming and costly undertaking and your choice of CDS could have a significant impact on your ability to comply with data integrity requirements and as such it is vital to ensure you choose CDS software that is able to fully meet these requirements.

More information on the Thermo Scientific^{TM} Dionex^{TM} Chromeleon^{TM} CDS can be found at thermoscientific.com/Chromeleon.

References

1. MHRA GMP Data Integrity Definitions and Guidance for Industry March 2015. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/412735/Data_integrity_definitions_and_guidance_v2.pdf

2. EudraLex - Volume 4 Good manufacturing practice (GMP) Guidelines. http://ec.europa.eu/health/documents/eudralex/vol-4/index_en.htm