The Ideal Chromatography Data System for a Regulated Laboratory, Part II: System Architecture Requirements

Here in the second part of this series, the key system architecture requirements for a chromatography data system (CDS) in a regulated environment are discussed.

In the first article in this series (1), we looked at the role of the laboratory and discussed the concept of the analytical factory together with the controllable and uncontrolled factors influencing the analytical process. In addition, we looked at the requirements for ensuring data integrity throughout the analytical process. In this second installment, we start by defining in more detail the requirements for the future of chromatography data systems (CDS) in a regulated laboratory. Specifically, we discuss the overall system architecture for a compliant CDS in a regulated laboratory. There are a number of requirements that, in our view, a system needs to meet before a CDS can be considered to be capable for operation in a regulated laboratory.

Where Are We Now?

Current chromatography data systems come in a variety of shapes and sizes and the choice available to a laboratory will depend on the overall system size and available budget. There are three possible CDS configuration options (2):

- A standalone workstation controlling two or more chromatographs
- A standalone workstation controlling a single chromatograph including liquid chromatography–mass spectrom-

etry (LC–MS) or gas chromatography (GC)–MS instruments
- Networked CDS system controlling multiple instruments in one or more laboratories

What Do We Need?

In our view, five main requirements are essential for a CDS operating in a regulated environment:

- Networked CDS
- Data management via a database
- Independent IT support
- Ability to interface to other instruments and systems
- Nonproprietary data file formats including the metadata

We will discuss each one of these in the following sections.

Networked CDS Architecture

Let us be very clear that, in our opinion, for regulated analysis standalone workstations are not fit for purpose and should not be used. Only a networked CDS architecture solution should be considered. Let's look at the rationale for our position, in general:

- Standalone workstations have a problem with resource contention—for example, access to the system by different users at the same time. When data are being processed, the same

R.D. McDowall and Chris Burgess
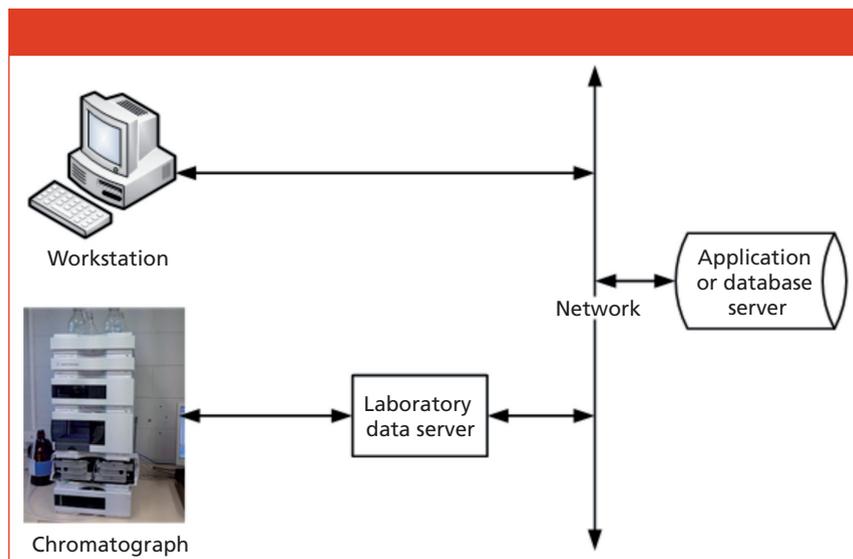
Figure 1: Overall CDS architecture diagram.

workstation cannot typically be used to set up another analysis, which can reduce the throughput of the system.

- Furthermore, if during an overnight run the workstation is left unattended, it might be possible for someone to make changes that are attributed to the user who initiated the run and who may not be in the laboratory when the changes were made.
- Data can be subject to manipulation as evidenced by the number of warning letters (3).
- There is a single point of failure with the workstation hard drive, coupled with the potential loss of regulated data.

So, from the perspectives of regulatory compliance and practical use of the system, a networked CDS solution is the only option that should be considered for regulated laboratories. This statement applies even if only a single chromatograph is used. With a networked system, data can be acquired on one instrument, but processed on a different workstation in an office because the data are available via a central server. In addition, a networked CDS has one or more data servers located in the laboratory to buffer data, add resilience, and avoid data loss.

Therefore, for resilience, result processing, and review independence a networked architecture is preferred to a single workstation.

Even for a small laboratory working in a regulated environment, the CDS must be networked. Data should be acquired directly to a secure network server that is regularly backed up by the IT department. Using the currently available technology, a virtual network server, rather than a physical server, could be used to store CDS data on a network even for a single instrument. There needs to be adequate redundancy and resilience in the physical hardware platform on which the virtual server runs to reduce data loss. Today, this redundancy is achieved in CDS with the incorporation of a data server in the laboratory to buffer data in case the network is unavailable; this practice should be continued in the future.

Data Management via a Database

To ensure integrity, all data generated during any analysis must be stored safely and securely to prevent deletion, either deliberately or accidently as well as track all changes made to the data by authorized personnel. Therefore, the second architectural requirement for a CDS operating in a regulated laboratory is the need for all data to be managed via a database. Data files stored in directories in the operating system are not fit for purpose in a regulated environment. The reason for this lack of fitness has been shown by numerous warning letters regarding noncompliance and falsification through deletion of unwanted files via the operating system (3). In fact, one way inspectors will demonstrate this is to ask for a file to be created by a chromatograph and then ask a user to attempt to

delete the file via the operating system.

The main reasons for incorporating a database in the system are to

- Manage all chromatographic data and associated contextual metadata
- Provide secure and encrypted storage of chromatographic data (4)
- Provide a secure and encrypted audit trail that is independent of the data files
- Have the ability to monitor, trend, and manage chromatographic data effectively across analytical runs (5)

A database is not simply an add-on to an existing CDS, but needs to be integrated and designed from first principles. You might argue that this is a draconian approach, but there are, sadly, numerous examples of falsification using operating system files. Prevention is always better and cheaper than the cure. In addition, so much mitigation is required to secure flat files that the database is a simpler solution once adequate control of data is considered.

However, some CDS systems on the market use operating system directory structures to store data, so if you insist on using a flat file structure the following issues need to be managed:

- The relationship between records must be embedded, so if files are separated the links can be established between records.
- Files must be protected from modification, copying, or deleting by unauthorized personnel immediately as they are written onto a storage medium.
- Temporary (scratch) files generated by the system must be segregated from data and metadata of interest to users, auditors, and inspectors. These scratch files are intermediate products, used by the system to create user results and metadata. When stored in the same folder as user files, temporary files demand read, write, and delete rights to operate, which opens user files to unapproved changes and deletion. This major design flaw exists in many standalone systems on the market today.
- Audit trail entries are typically embedded in the individual data files; if a file is deleted the corresponding audit trail is deleted as well.
- Access to the operating system and

the system clock must be restricted to authorized administration personnel. No normal users should have the capability to access these portions of the system.

All-in-all, a database is a much better way to go for the future CDS.

Independent IT Support

Independent IT support is essential to separate administration of the system from the normal chromatographic analysis functions of the software. This support ensures that analytical staff do not have access to change items such as turning the audit trail on or off or modify the date and time of the system. Therefore, the following functions need to be included under this section:

- Setup and management of the software application settings: The IT department should set up the configuration software settings that have been defined and documented by the laboratory and maintain them under a formal change control process. This management by the IT department ensures that laboratory staff cannot make changes to the configuration of the software directly.
- User account management: The definition of user types and the associated user privileges will be performed by the laboratory staff, but implemented and maintained by IT.
- Time and date settings: Networking the data system has the benefit of taking the date and time stamp setting out of laboratory control. Time and date settings are potential sources of tampering to affect the results. IT staff should be the only people with access to the network clock, which is synchronized with a trusted time source such as a Network Time Protocol (NTP) server or a government agency such as the US Naval Observatory or Greenwich Mean Time (GMT or UTC).
- Data backup and recovery: If data backup is left to the laboratory the possibility arises that the work is not actioned or not done correctly. Backup problems were found at Ohm Laboratories (6) where the backup was not performed or staff may lose data as the Food and Drug Administration (FDA) found at Cambrex

Profarmaco (7). For IT departments, one of the key tasks is backup and recovery of data and this process can and should be automated and carried out by IT, independently of the originating analytical laboratory.

Interfaces to Instruments and Systems

As we mentioned in part I of this series (1), a CDS should not exist in isolation. A CDS needs the capability to be interfaced with some analytical instruments as well as other informatics applications for business reasons. Essentially the whole purpose of interfacing is to eliminate manual data entry as much as possible, or reduce it substantially and replace it by seamless data transfers from where the data were originally acquired. For example, the CDS should be able to electronically accept sample identities, electronically match CDS results to them, and forward sample and results to a system, such as a laboratory information management system (LIMS), for batch evaluation.

As an example, the main CDS workflow can be made electronic, but there is still a large amount of manual data that must be input into the application such as sample weights, purities, dilution factors, and so on. To avoid the need to record weights from the balance, manually enter them into the CDS, and check them (these are critical data under clause 6 of EU GMP Annex 11 [4]), analytical balances should be interfaced to an informatics application. This informatics application can either be the CDS itself or another system, such as an electronic laboratory notebook (ELN) or LIMS, from which the weights can be transferred to the correct sequence file using a validated routine.

Following the analysis the electronically signed CDS results need to be transferred for comparison with the specification either to a LIMS or an ELN, thus avoiding the need for transcription checking. In all of these interfaces, audit trail coverage of the transfer is essential to record the acquisition of data from one system in the audit trail.

Another consideration is buffering of data, to prevent loss if the CDS is temporarily down while an assay is running. Interfacing permits the use of buffers to prevent data loss.

Open Data File Formats

In the 1990s there were attempts at data file standardization for chromatography data systems, so the network common data format (NetCDF) file format was adopted for them. However, this approach is inadequate because it only covers the data file itself and not the metadata that surround it, such as sequence, instrument control, data acquisition, and processing files that put the data file in context. Because the regulators are demanding longer retention periods, such as for the time that a marketing authorization is in force (8), then a move to a file format that permits long term access to the data is imperative.

The American Society for Testing and Material's (ASTM) Analytical Markup Language (AnIML) is the main approach for a solution to the archiving issue that has been developed by the ASTM subcommittee E13.15 (9). The solution is text based rather than a binary file format that includes all contextual metadata.

Summary

In this part of our discussion on the future requirements for a CDS for regulated environments, we have discussed how any system must be networked with a database to ensure that any data generated have the integrity from acquisition to reporting. Furthermore, key support functions such as software configuration, user account management, and backup must be controlled by an independent IT group.

Interfaces to other instruments and systems are essential to ensure electronic acquisition and transfer of data while eliminating manual entry and manual transcription checks.

Finally, we need nonproprietary data file formats typically based on the new ASTM AniML standard to provide a mechanism for longer term archiving.

References

(1) R.D. McDowall and C. Burgess, *LCGC North Am.* **33**(8), 554–557 (2015).

(2) R.D. McDowall, *Validation of Chromatography Data Systems: Meeting Business and Regulatory Requirements (*Royal Society of Chemistry, Cambridge, 2005).

(3) R.D. McDowall, *LCGC Europe* **27**(9), 486–492 (2014).

(4) European Commission Health and Consumers Directorate-General, *Eudralex: The Rules Governing Medicinal Products in the European Union, Volume 4: Good Manufacturing Practice, Medicinal Products for Human and Veterinary Use*, *Annex 11, Computerised Systems* (Brussels, Belgium, 2011).

(5) European Commission Health and Consumers Directorate-General, *Eudralex: The Rules Governing Medicinal Products in the European Union, Volume 4: Good Manufacturing Practice, Medicinal Products for Human and Veterinary Use*; Chapter 6 Quality Control (Brussels, Belgium, 2014).

(6) Ohm Laboratories, FDA Warning Letter, December 2009.

(7) Cambrex Profarmaco, FDA Warning Letter, August 2009.

(8) European Commission Health and Consumers Directorate-General, *Eudralex: The Rules Governing Medicinal Products in the European Union, Volume 4: Good Manufacturing Practice, Medicinal Products for Human and Veterinary Use*; Chatpter 4, Documentation (Brussels, Belgium, 2011).

(9) American Society for Testing and Materials (ASTM), *E15.15* Analytical Markup Language (ANiML).]

R.D. McDowall *is the director of R.D. McDowall Ltd., and* **Chris Burgess** *is the managing director of Burgess Analytical Consultancy Ltd. Direct correspondence to: rdmcdowall@btconnect.com* ∎