Chromatography Data Systems

INTRODUCTION

Chromatography is an analytical technique used in virtually all areas of the pharmaceutical and biotechnology industries to detect or measure compounds during the course of product development and manufacture. It can be used for the measurement of active ingredients, raw materials, impurities and determining the stability of active substances in final preparations. The chromatograms from these analytical methods produced are generated, displayed, integrated and results calculated by a software application called a chromatography data system (CDS).

This chapter presents some approaches to prospectively and retrospectively validating client server networked CDS based on case studies; in addition the business benefits that can be exploited from the implementation of electronic signatures when remediating or updating a legacy chromatography data system are presented.

What is a Chromatography Data System?

This section discusses the operation of a chromatography data system from the perspective of a typical laboratory process or workflow; Figure 1 shows the overall sequence of events that a typical data system should perform. This is a generalised approach to the operation of a "typical" data system; further details on the subject are in the articles by McDowall [1, 2] and the book by Dyson [3].

Method Files

The start of the data acquisition operation of a chromatography data system is to build a method file. This tells the data system how to acquire data and process and interpret the results. A method file should control:

- The data sampling rate of the analogue to digital (A/D) converter [4],
- When to start and stop the integration of the chromatogram,
- Whether peak areas or heights should be used,
- Retention time windows and identification of the analytes and internal standard
- Allocate the method to calculate the analyte amount or concentration.

A name, number or a mixture of both should identify individual method files within the system. In addition, the system should be able to provide facilities for version control of method files to ensure that control is maintained over the method for the lifetime of its use. Part of the control function must be access control to identify the individuals who can create, modify or delete analytical methods. If a method has been modified then copies of the modifications must be stored with the data processed by that method. This is to provide an audit trail for the data and results produced by a version of a method. However, when developing methods, flexibility with method files is essential and a default method should be available to acquire data and then feedback to a normal method.

Naming Conventions

When a laboratory uses a client-server CDS there will be an urgent need to consider naming conventions for method, sequence and all data files within the data system. Any CDS must have sufficient capacity for naming all of files that would be created by the system over a reasonable time period to aid efficient archiving and unambiguous identification of these files. Therefore for efficient management of data files and methods, naming conventions should be introduced. Any naming convention system must aid users, quality assurance, and regulatory inspectors.

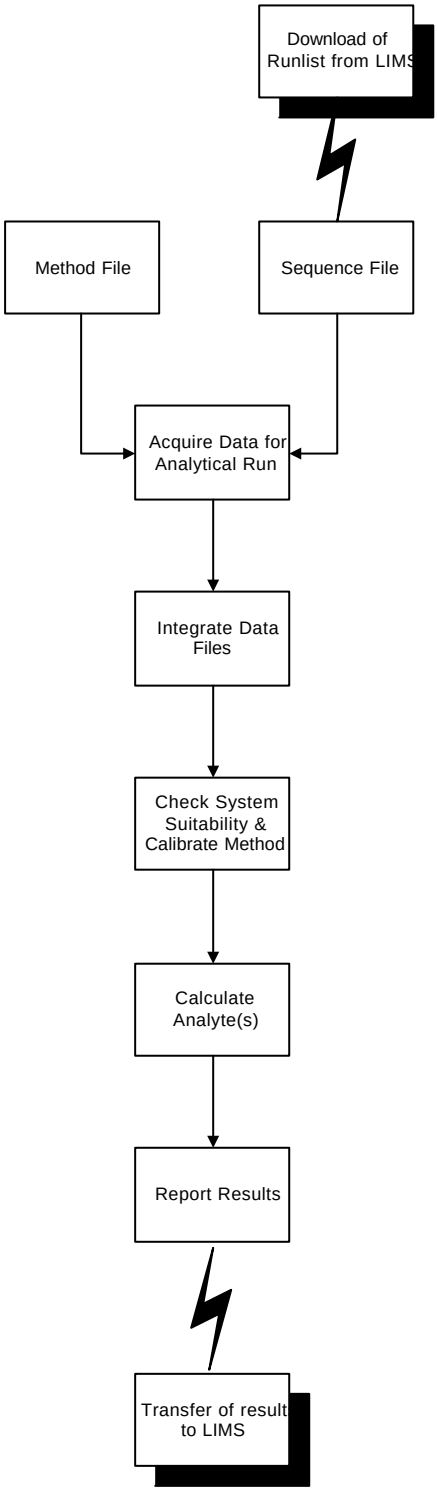

Figure 1: Workflow for a Typical Chromatography Data System

A naming convention should be based on the workflow undertaken by a laboratory. This is to allow efficient archiving of data but just as importantly, the efficient retrieval of data. Some ideas might be:

- Organise the data around drug products or development projects, as this is how the work is structured and how project teams are organized as this will help retrieve data to aid 21 CFR 11 compliance for ready retrieval of electronic records.
- Major subdivisions of each project should be based around the type of work done e.g. method development, method validation, pre-formulation etc.

Sequence File

The sequence file is the run list or order that the samples, standards, quality control samples and blanks will be injected into the chromatograph; this is essential as it puts in context to the individual data files. Each sequence file or each injection must be linked with a method file to process the resulting data. For laboratories with large numbers of samples for a single method, the sequence file will usually be linked with a single method. Smaller laboratories may need the flexibility to link the sequence file with several methods during the course of a single analytical run for maximal use of equipment resources.

Each sample to be analysed should be identified in the sequence file as one of the following types:

- Unknown
- Calibration standard
- Quality control
- Blank

Depending on the data system involved, at least the first two options are available to a user. There may also be a sample number to link the injection to the physical sample used for analysis.

Interpretation of Chromatographic Data

After the method file and the sequence file have been set up the analytical run is started and data are collected. A data file containing the A/D data slices will be obtained for each chromatographic run and sample injected. It is important from scientific and regulatory considerations that the data files must not be capable of alteration.

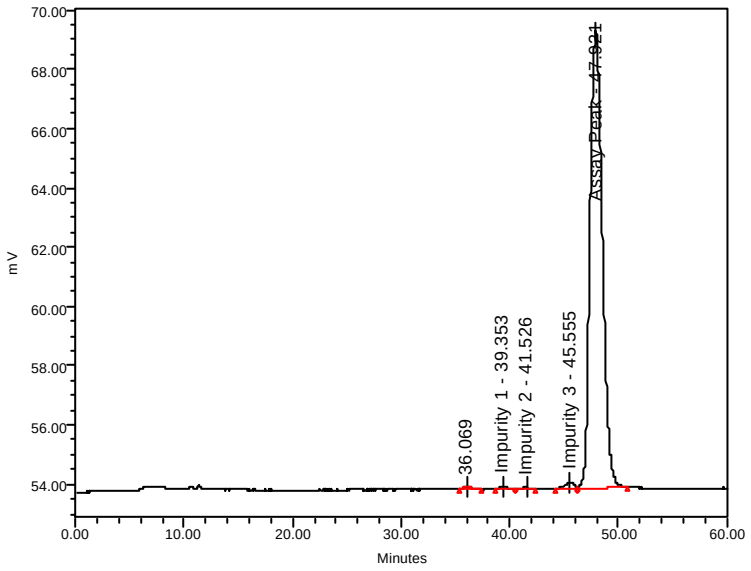

Figure 2: A typical chromatogram of an active substance separation from impurities/ degradation products

Moreover, they must not be overwritten either if the same sample information is assigned to an assay or if the disc becomes full. This is an area for consideration when validating the chromatography data system; as you must know what happens to your data files, especially in a regulated environment.

The data system will interpret each data file, identifying the individual peaks and fitting the peak baselines according to the parameters defined in the CDS method as shown in Figure 2. The data systems should have the ability to identify whether the peak baselines have been automatically or manually interpreted. This is a useful feature for compliance with Part 11 to indicate the number of times a chromatogram has been interpreted.

Most data systems should be able to provide a real-time plot, so that the analyst can review the chromatograms as the analytical run progresses. In addition, the plotting options of a data system should include:

- Fitted baselines;
- Peak start/stop ticks;
- Named components;
- Retention times;
- Timed events e.g. integration start/stop;
- Run time windows and user defined plotting windows;
- Baseline subtract;

Each of these options should be enabled or disabled by a user.

An overlay function should be available to enable you to compare results between samples. This will be used to compare chromatograms from the same run sequence as well as chromatograms from different sources. The maximum number of overlays will vary from data system to data system but a minimum of 6-8 is reasonable and practicable. More overlays may be technically possible, but the amount of useful information obtained may be limited. Overlays that can be offset by an amount determined by the user are useful to highlight certain peak information. Ideally, the overlay screen should have hidden lines removed and be able to be printed.

Calibration

Calibration is a weak area with most data systems, as most chromatographers use many ways to calibrate their methods as evidenced by the multitude of calibration options available. Often these methods are basic and lack statistical rigour, as the understanding of many chromatographers, where calibration is concerned, can be poor.

Within a pharmaceutical analysis laboratory, the number of calibration model options that can be successfully used is usually limited to:

- Bracketed standards at one concentration or amount for bulk drug or finished product assays
- Response function for all analytes
- Average by amount for bulk drug and finished products
- Multi-level or linear regression for related substances and degradation products

Within each calibration type, the data system must be able to cope, sufficiently flexibility, with variations in numbers of standards used in a sequence and types of standard bracketing. The incorporation of a blank standard into the calibration curve should always be an option.

Each plot of an analyte in a multi-level or linear regression calibration model must contain an identifier for that calibration line and the analyte to be determined. The calibration curve should show all calibrating standards run in any particular assay. In assays containing more than one analyte it will be necessary to interpret all the calibration graphs before the calculation of results. Again, this is an area that is poor for data system as many only offer one line fitting method for all analytes in the run resulting in compromises.

User Defined Analytical Run Information

The system should be capable of collating user-defined parameters (e.g. height, area, ratios, concentrations etc) for selected analytes from a sequence of runs. After collation system defined and/or user defined statistical calculations will be carried out on the data generated. The type of calculations required should include mean, standard deviation, analysis of variance and possibly significance testing.

Reports and Collation of Results

Ideally, the report following an individual chromatogram should contain both elements that are user definable and those, which are standard; this should enable the laboratory to customise a report. At the end of the analytical run, a user defined summary report containing information such as sample ID, area or height, baseline and calculated analyte concentration should be created. This report can either be printed out or transferred to a LIMS for further analysis and interpretation.

Instrument Control

The primary interaction of the CDS with analytical instrumentation is with the output from the detector; however, there are other considerations such as instrument control. These can vary from system to system and the following options are available:

- Contact closures for the control of chromatographic valves or associated equipment during analysis is usually available for other supplier's equipment.
- When the same supplier makes the data system and the chromatography equipment; control is more sophisticated and more tightly integrated with the data system functions so control of the instrument and set up of the data system can be achieved from a single workstation.
- Communication with the auto-sampler via binary coded decimal (BCD) or equivalent communication for sample continuity is, in my view essential, but is usually ignored by many and offered as an option by many suppliers;
- Remote monitoring of the chromatography system output including the instrument conditions.
- The ability to list the items of equipment (pump, detector etc) used for a particular analysis is a function to help to automate the administrative records associated with an analysis and help meet GMP compliance.

Architecture of a Networked CDS

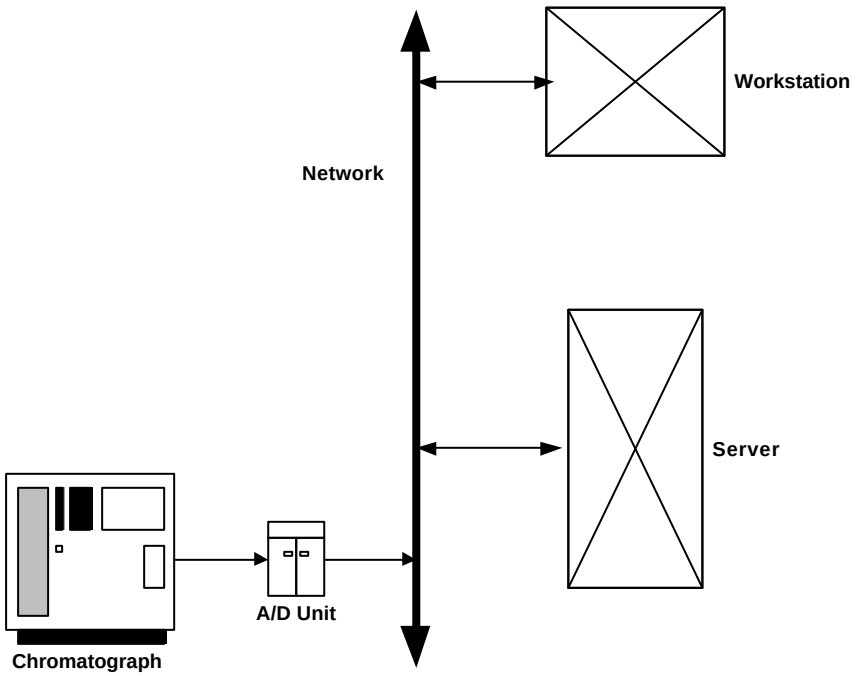

Figure 3: Schematic Diagram of a typical networked Chromatography Data System

A typical networked chromatography data system will consist of several hardware components as shown in Figure 3:

- Chromatograph: this is the instrument that performs the analytical separation and can be a high performance liquid chromatograph (HPLC), gas chromatograph (GC) or a Capillary Electrophoresis (CE) instrument.
- Data acquisition via an analogue to digital (A/D) converter from the instrument detector to the CDS and converts the continuous analogue signal to a number of discrete digital data readings. This is an optional item; if the instrument is controlled by the CDS then data are transferred digitally via the network cables running to the data system. Often the A/D unit can have buffering capability if the network is temporarily unavailable and to prevent data loss.
- Network: transport medium for moving the data from the instrument to a server for secure data storage.
- Workstation (Client): for operating the CDS setting up an instrument, checking that the separation is working correctly and interpreting the resultant chromatograms after the run is finished and reporting the results.

KEY REGULATORY REQUIREMENTS AND ISSUES WITH A CDS

Before discussing how to validate a chromatography data system, it is important to understand the regulatory requirements and their interpretation. The responsibility for the validation rests with the system or business process owner but from experience most do not understand fully the regulations they work under or the risk mitigation strategies that need to be undertaken when validating a CDS.

The regulations and guidelines have a view on what is expected during the implementation and release of a CDS as well as what is expected when the system is operational and when it is retired. In general, the emphasis is concerned with generating the proof to demonstrate that the computerised system is accurate when validated and continues to be so when it is operational and that there is sufficient proof of management awareness and control. To obtain proof of an action usually means that it must be documented, although the format of documentation (paper or electronic) is left open by all schemes.

The definition of Performance Qualification is *documented verification that the computer related system performs it functions in accordance with the computerised system specification while operating in its* ***normal operating environment***. [5].

The major point to make is that the laboratory must test the CDS as they use it and not how the supplier has tested it (i.e. in the laboratory's operating environment, using the laboratory's analytical methods, specifications and capacities and using the laboratory's networks).

Warning Letters and 483 Observations Involving CDS

Some of the key 483 observations and warning letters involving chromatography data systems are discussed in this section. This is not an all-inclusive list of non-compliances and the reader is encouraged to look at the FDA web site to keep abreast of any changes in emphasis of inspections.

Gaines Chemical Company 483 Observations

In December 1999 [6] the FDA inspected the client server CDS operated in the QC Laboratories of the company and found the following observations:

- The CDS had never been validated and there no documentation to assure that the system can operate as intended
- There was no change control
- No security was enabled and anyone could access the system
- No record of system configuration
- The application audit trail had been deliberately turned off by the staff
- No documentation of calculations performed by the system
- The application security could be bypassed by using Windows Explorer; implying that files could be deleted outside of the application and with no record
- Passwords consisted of four characters and never expired

- When the system was operational anyone could access the application; the workstation had to be turned on otherwise data could not be acquired
- There were no SOPs for the operation of the system
- Backup and recovery was not demonstrated and the storage conditions of backup tapes was not verified

These observations reflect the situation in many small to medium sized companies that work in the regulated environment.

Glenwood Warning Letter

In May 1999, Glenwood LLC received an FDA warning letter [7] that contained the following non-compliance relating to their CDS software:

> *"Failure to validate the software programs, _____ and ____, that are used to run the laboratory HPLC equipment, during analysis of raw materials and finished products. The ____ software does not secure data from alterations, losses, or erasures. The software allows for overwriting of original data. There are no written procedures for the use of passwords, levels of access, or data back-up".*

Apart from failure to validate the CDS application,. Therefore protection of all electronic records created by any CDS is vitally important.

Gensia Scicor Warning Letter

A warning letter sent to the company in July 1999 [8] again reiterates the importance of protecting and preserving electronic records:

> *"Failure to maintain laboratory records to include complete data derived from all tests necessary to assure compliance with established specifications and standards [21 CFR 211.194]. Specifically, your firm failed to properly maintain electronic files containing data secured in the course of tests from 20 HPLCS and 3 GLCS. Additionally, no investigation was conducted by your company to determine the cause of missing data and no corrective measures were implemented to prevent the reoccurrence of this event".*

The critical problem was loss of electronic records coupled with a failure to investigate the problem to stop it happening again. Note the use of the predicate rule citation rather than 21 CFR 11.

Noramco 483 Observations

A bulk chemical company was inspected in May 2001 [9], the 483 observations highlight the detail of due diligence that any CDS validation requires in the 21 CFR 11 world. The observations are reproduced below:

> *There was no assurance that data acquired on the _____ chromatographic client-server data system was accurate, reliable and reproducible for analyses of …..*
>
> - *The CDS was not validated to ensure the system produced accurate and precise data.*
> - *There was no documentation to show that the system's ability to handle overload situations in an orderly fashion.*
> - *There was no assurance of the program's behaviour when working at its limit. Functional testing that includes volume and stress testing was not conducted to demonstrate the system's behaviour.*
> - *Confidential and unique user log-ins and passwords were not assigned to each analyst to ensure data authenticity and integrity. Each workstation had a single log in name and password, which was shared with all users.*
> - *There were no automatic computer generated time-stamped audit trails to ensure authenticity and integrity of analytical data that was acquired and processed with the CDS. Analyst's transactions were not documented to show whether the analytical data were modified, copied or deleted.*

- *There was no documented evidence that the CDS was adequately configured and performed as intended.*
- *The firm did not have a system administrator that was responsible for system configuration and control of access to configuration tools that can modify or delete electronic records. System administrator permissions and rights were given to some QC analysts who were also responsible for analysing samples.*
- *There was no control over how analysts interacted with analytical data on the system.*
- *The universal log-in and password system gives users rights and permissions to edit, modify and delete data files. The system was not configured to deny analysts rights to directories and users did not have read / write access to analytical data on the system. Users could not only modify their records but all records on the server. There was no written documentation that established what limits and rights the IT groups assigned QC laboratory users.*
- *There was no documented evidence to show that the firm periodically restored analytical data from its tape backup medium to ensure that data files could be reconstructed and were not corrupted. IT personnel did not know how to reconstruct the graphic data on workstations and referred us to analysts in the laboratory to perform system administrator tasks.*
- *There was no documentation to show that analytical data on the chromatography network could not be altered or modified by authorised users of the corporate network. The networks are connected by a router, which enables data packets to move between networks. The chromatography network did not have capabilities for tracking and controlling the integrity of each sample throughout its retention period. There were no protocols that explained the logical security procedures in place to prevent unlimited and unauthorised access to chromatographic data files.*

Key Inspection Learning Points

Some of the key learning points from these inspections and warning letters that we need to remember for the validation of any CDS are:

- The CDS must be validated and the scope of work includes documenting any customisation or configuration of the system
- Include in the PQ testing, capacity tests for stress and overload conditions to comply with §211.63 ("adequate size"). The nature and extent of these capacity tests will vary depending on the architecture of the individual CDS system and also how an individual laboratory uses it.
- Effective preservation of electronic records is vital to passing any inspection: have a procedure, follow it and have documented evidence that it works. Use redundant hardware such as RAID disks (Redundant Array of Inexpensive Disks) and uninterruptible power supplies (UPS) as a first line of defence against electronic record loss.
- Change control is vital and the process must include the IT department and the network
- Security must be enabled, documented and tested.

EXPLOITING THE BENEFITS OF ELECTRONIC SIGNATURES WITH A CDS

Rationale for Using Electronic Signatures

The Electronic Records; Electronic Signatures (21 CFR 11) final rule is an integrated regulation: subpart B (electronic records) has requirements for signing electronic records whilst subpart C (electronic signatures) has controls that are as important for ensuring the trustworthiness and reliability of electronic records as well as electronic signatures. Therefore, to use legacy systems in a hybrid mode is just a temporary solution before working completely electronically. In this section, the ways that the design of electronic signatures can be implemented into a chromatography data system (CDS) will be discussed.

A prerequisite for this approach to succeed is the need for any software to be technically compliant with the requirements of 21 CFR 11. Therefore, it is important that before implementing electronic signatures that the software used is technically compliant with the requirements of the regulation and the laboratory's interpretation of the regulation.

The key principle is that to implement electronic signatures on an existing paper based process is not just a matter of electronically signing the calculated results. It requires a different philosophy and also requires a good understanding of the regulations that an organisation has to comply with and also the business processes that will use electronic signatures.

It is unlikely that an organisation will benefit implementing electronic signatures on an existing process unless it has been implemented to work electronically [10].

To illustrate this principle, the interim results from a laboratory where electronic signatures have been designed into the process will be presented and discussed. The CDS is installed in a pharmaceutical quality control laboratory where the system used for both raw material and finished product analysis; there are approximately 50 part-time users of the system. The current CDS version was not fully compliant with the technical requirements of 21 CFR 11 and was to be upgraded to a new compliant version of the software from the same vendor. Before the implementation of the new version, the current process was mapped and analysed to see if there were any opportunities for improvement and to make effective use of electronic signatures.

There is also a LIMS that is operational in some of the sections within the Laboratories, however at the moment there is a mixture of both lab notebooks and a LIMS being used.

Mapping and Understanding The Current Process

The first task when considering implementing electronic signatures is to map the current process. This is relatively quick and the current laboratory high level process is shown in Figure 4. We can see that there are parallel electronic and paper activities when chromatographic analysis is undertaken. For example, when a chromatograph is set up, a paper record (Lab Book) needs to be updated and checked. When results are calculated the report and chromatograms printed out and the Lab Book updated and checked again.

It is important to analyse the current process:

- What are the process metrics? For example:
  How many samples are analysed?
  What are the turnaround times?
- Once this information has been obtained, analyse the turnaround times and find out the reasons for fast and slow turnaround?

Answers to these questions will give you the information to start to improve the process and make it more effective and efficient.

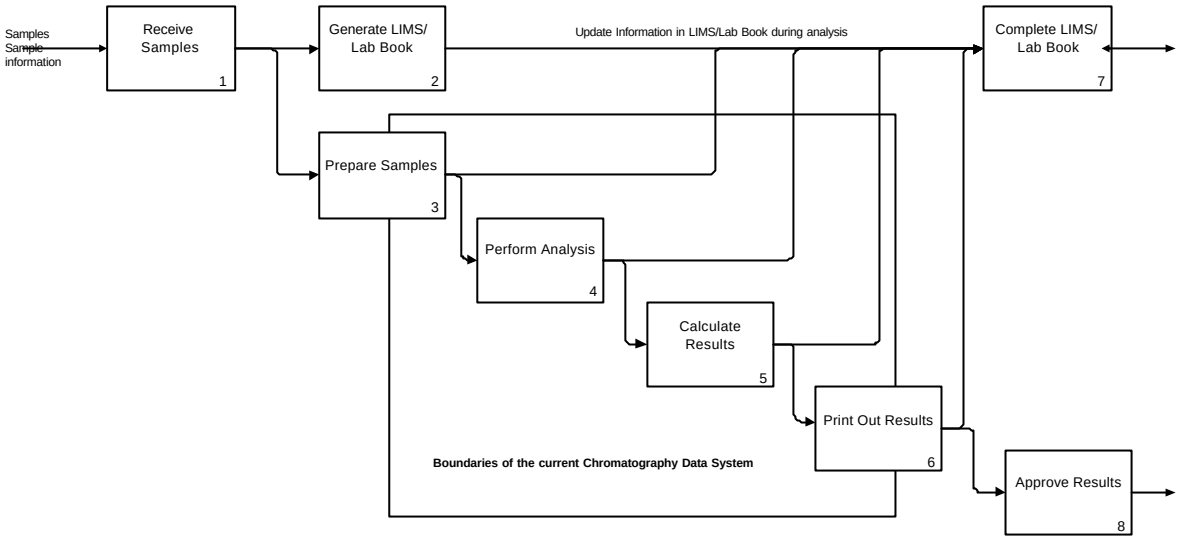

Figure 4: The current process highlighting the boundaries of the current version of the CDS

The boundaries of the current version of the chromatography data system are also shown in Figure 4. In the current system the approval of results occurs outside of the chromatography data system on paper.

Optimising the Workflow to Use Electronic Signatures

Knowing the problems and improvement ideas from the analysis of the current ways of working, a new process can be designed to exploit the use of electronic signatures. It is important at this stage to ensure that the new process is compliant with 21 CFR 11 and any predicate rule requirements and that the new version of the CDS can support the new process as well. For example, where in the process will you use signatures and where will identifications of actions be sufficient?

In the example, the redesigned process is shown in Figure 5; the main differences are:

- Elimination of the need to update the Lab Book for chromatographic analysis. This is a quick win that is estimated to save about 0.3-2.6 FTE (Full Time Equivalents or person years). This is independent of implementing electronic signatures in the CDS
- Expanding the scope of the CDS. In effect the approval of electronic records and calculated results takes place in the CDS and the printout is an option.
- Using the CDS to carry out all calculations rather than use a calculator or spreadsheet, this streamlines the whole process for calculating, reviewing and approving results.

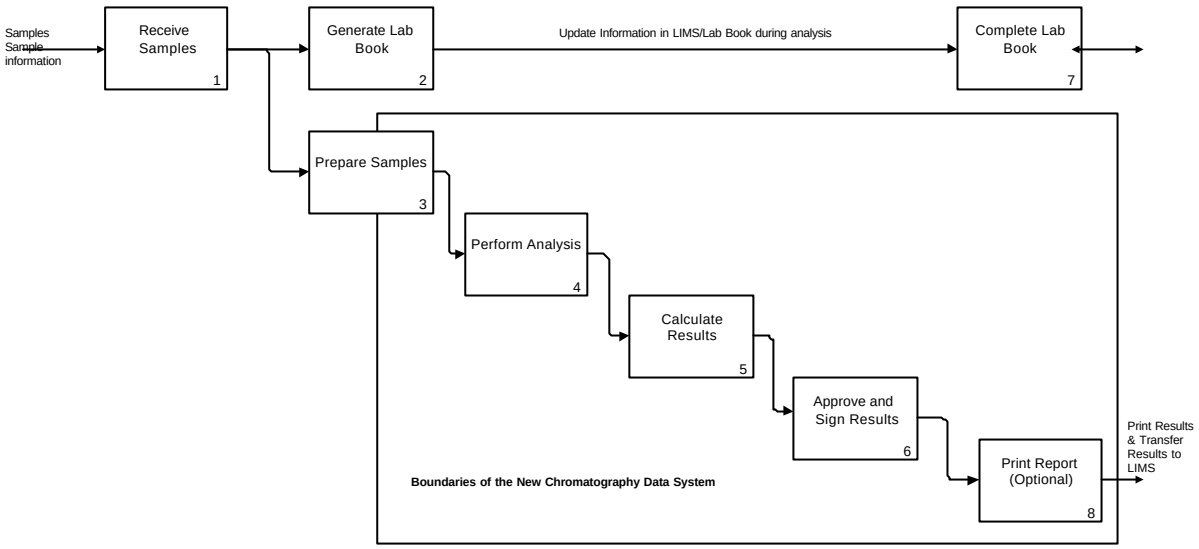

Figure 5: The redesigned process highlighting the extended boundaries of the new version of the CDS

The benefits of the process redesign when the CDS is linked to the LIMS would be in the region of 6-12 FTE. This is a surprising benefit but enables more capacity to be generated with the current laboratory resources. This is against a one off cost of about 2 FTE for the process redesign, linking the system to the LIMS and validation of the CDS and the data link to LIMS.

LIFE CYCLE APPROACH TO VALIDATION

GAMP Software Classification

A Chromatography data system should, in the author's opinion, be classified as GAMP category 4 and where customized macros or calculations are involved GAMP category 5. The rationale for this is that all commercial CDS applications need configuration at least to acquire data from the various chromatographs they are connected to or control these instruments.

Therefore the discussions on the life cycle and the validation will be based around this premise of a GAMP 4-5 software application.

CDS Life Cycle

An International Standards Organisation (ISO) system development life cycle model is shown in Figure 6 and is depicted in the shape of a V; it is different from the GAMP "V model" [11] as the model below more accurately represents user and supplier relationships in regards to COTS products. It is important to realize that there is a division between the user (above the line) and the supplier (below it). The qualification stages are condensed into a single stage under the control of the user below rather than presented as three distinct stages that never occur in practice.

The left hand side of the V represents the design stages of the application, the bottom is the programming and the right hand side is the testing stages of the life cycle.

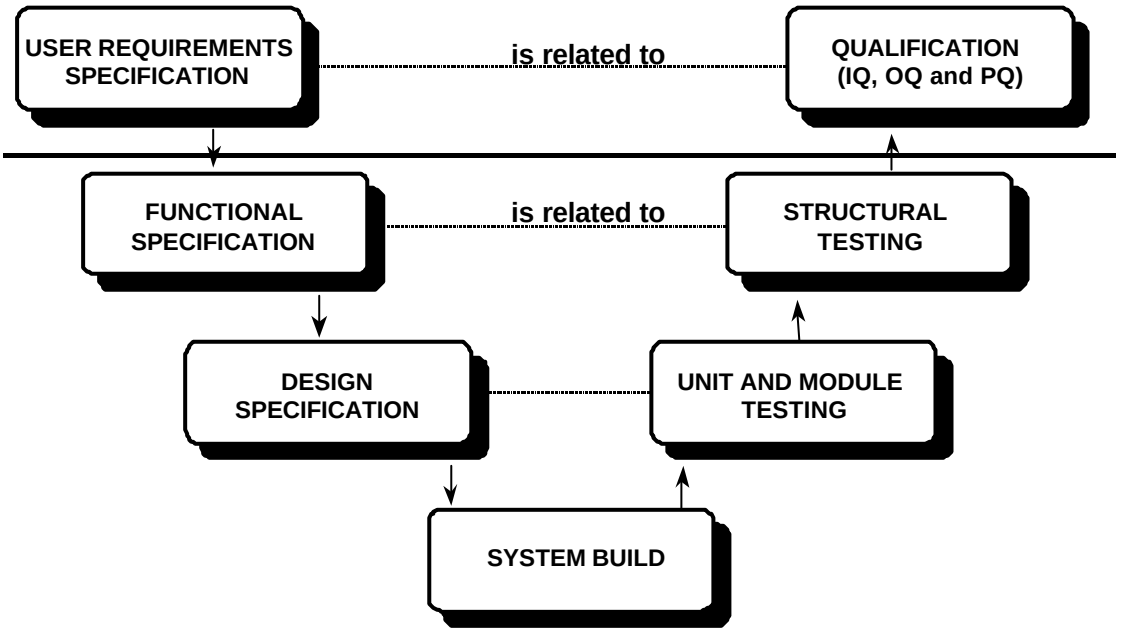

Figure 6: A System Development Life Cycle (SDLC) of a Chromatography Data System

This model can be used to generate the documentation that could be produced during the system development life cycle; the documents that could be produced are presented in Table 1 and the key ones are discussed in more detail in the next section. Taken together all of these documents will provide the validation package to support the contention that the chromatography data system is fit for purpose. Note please, that this is a suggested minimum list; you may write less or more documents than outlined here.

The extent that an individual validation differs to this approach will depend on the amount of regulatory risk that the organisation or laboratory management wishes to carry after the validation.

Document Name	Outline Function in Validation
Validation Plan	• Documents the intent of the validation effort throughout the whole life cycle • Defines documentation for validation package • Defines roles and responsibilities of parties involved
Project Plan	• Outlines all tasks in the project • Allocates responsibilities for tasks to individuals or functional units • Several versions as progress is updated
User Requirements Specification (URS)	• Defines the functions that the CDS will undertake • Defines the scope, boundary and interfaces of the system • Defines the scope of tests for system evaluation and qualification
Risk Analysis and Traceability Matrix	• Prioritising system requirements: mandatory and desirable • Classifying requirements as either critical or non-critical • Tracing testable requirements to specific PQ test scripts
System Selection Report	• Outlines the systems evaluated either on paper or in-house • Summarises experience of evaluation testing • Outlines criteria for selecting chosen system
Supplier Audit Report & Supplier Quality Certificates	• Defines the quality of the software from suppliers perspective (certificates) • Confirms that quality procedures matches practice (audit report) • Confirms overall quality of the system before purchase
Purchase Order	• From supplier quotation selects software and peripherals to be ordered • Delivery note used to confirm actual delivery against purchase order • Defines the initial configuration items of the CDS
Installation Qualification (IQ) Operational Qualification (OQ)	• Installation of the components of the system by the supplier after approval • Testing of individual components • Documentation of the work carried out • Testing of the installed system • Use of an approved suppliers protocol or test scripts • Documentation of the work carried out
Performance Qualification (PQ) Test Plan	• Defines user testing on the system against the URS functions • Highlights features to test and those not to test • Outlines the assumptions, exclusions and limitations of approach
PQ Test Scripts	• Test script written to cover key functions defined in test plan • Scripts used to collect evidence and observations as testing is carried out • Documents any changes to test procedure and if test passed or failed
Written Procedures	• Procedures defined for users and system administrators including definition and validation of custom calculations, account management and definition of logical security • Procedures written for IT related functions • Practice must match the procedure
User Training Material	• Initial material used to train super users and all users available • Refresher or advanced training documented • Training records updated accordingly
Validation Summary Report	• Summarises the whole life cycle of the CDS • Discusses any deviations from validation plan and quality issues found • Management authorisation to use the system

Table 1: Typical Documentation for a CDS Validation

KEY VALIDATION DOCUMENTS FOR A CDS

The main validation documents will be presented in this section, typically in the order in which they are written and used in the system development life cycle, however there are differences that will depend on individual circumstances.

Specifying the CDS Requirements

Defining the Basic CDS Functions

The first document in the validation is usually the URS as this can influence the validation strategy outlined in the validation plan. From Figure 6 that the system requirements are related to the tests carried out in the performance qualification. Therefore, it is important to define the requirements for the basic functions of the CDS, the adequate size, 21 CFR 11 requirements and consistent intended performance in the URS. Remember from the draft FDA validation guidance [12] the URS provides a laboratory with the predefined specifications to validate the CDS; without this document you cannot validate your CDS.

It is important to realize that the URS is a living document and must be updated as the system changes and evolves; for example an URS should be written to select a system, it will then be reviewed and updated to reflect the selected CDS and version that will be validated and the functions specific to the laboratory where it will be installed.

The main elements in an URS should include the following major areas; each requirement must be individually numbered and written so that it can be tested as noted later in this section:

- Overall system requirements such as: number of users, locations where the system will be used and the instruments connected to the system; will terminal emulation be used?
- Compliance requirements from the predicate rule and 21 CFR 11 such as: open or closed system definition, security and access configuration of the software application including user types, requirements for data integrity, time and date stamp requirements, electronic signature requirements.
- Data system functions defined using the workflow outlined in Figure 1 but ensure that capacity requirements are defined such as maximum number of samples to be run, custom calculations and reports for the initial implementation and roll-out etc.
- IT Support requirements such as: backup and recovery, off line archive and restore.
- Interface requirements such as will the CDS be a standalone system or will it interface with a LIMS and if so how?

System Specification Issues to Consider

Therefore, the first stage in the considerations for validating a CDS is to define all functions in a URS; for example some or all of the following requirements will be included in the document:

- Data capture rates across all chromatographic techniques connected to the CDS. For example conventional chromatography with a run time in the order of 20 minutes a data capture rate of 1Hz is usually adequate. However for capillary GC 10-20 Hz may be appropriate and for CE a higher rate may be required depending on the overall migration time and the analyte peak shape.
- Depending on your data system, several chromatographs may be linked into a collection workstation or an A/D unit. Here consider if crosstalk (the interference from one channel to another) could be an issue if the A/D chip is multiplexed across two or more channels and / or total sampling capacity of the data collection and buffering unit.
- Has the maximum number of injections for an analytical run been defined? This is a critical component, if 100 vials are routinely injected in a run, the system can't be tested with a run of only 10 samples as a user has not demonstrated adequate size. The specification must match the use of the system including replicate injections.
- Some data systems will be configured to collect data from Diode Array Detectors (DAD). If this is required, especially to analyse product, then the data collection and analysis will need to be checked as part of the adequate size as some data files can be in the Mb range. The file delete option should not be enabled to protect the electronic records generated.
- Virtually all client server CDS systems will have a buffering capacity within their A/D or data

collection units (if acquiring digital data from chromatographs via network interfaces). Therefore, so part of the adequate size requirements must be the ability to capture and buffer data if the network is unavailable, followed by the succe ssful transfer of data to the server when the network connection is re-established.

- How many users will there be on the system at the same time and will the system still perform its functions reliably? This number may be lower than the number of concurrent users that you have a license for but this is a major requirement to define in the URS and test during the PQ. If the system becomes unreliable or unstable as the number of users increases then the system owner cannot state that the system has adequate size or can perform as intended.

These are some of the considerations for each installation of a CDS, once installed in a laboratory environment and on the organisation's network it becomes unique. The number of users, network components, server components, operating systems, software patches and laboratory configuration make it so, therefore you need to demonstrate that it works under you operating environment.

Documenting the System Requirements for Traceability

Although not mentioned in the regulations specifically, traceability of system requirements to the testing phase is important for any system including a CDS, therefore the way that system requirements are presented and managed is important.

It is all very well the regulations stating that a user must define their requirements in a URS, what does this mean in practice? Table 2 illustrates one way that capacity requirements can be documented; each requirement; note that each requirement is:

- Uniquely numbered
- Written so that it can be tested, if required, in the PQ
- Prioritised as either mandatory (M = essential for system performance) or desirable (D = nice to have and the system could be used without it). This prioritisation can be used in risk analysis of the functions and also for tracing the requirements through the rest of the life cycle as will be discussed in a later section.

Remember as shown in Figure 6 that the URS functions are related to the tests carried out in the qualification phase of the life cycle. Therefore, if you have not specified the requirements in this document; how can you test them?

Req No.	Data System Feature Specification	Priority M/D
3.3.01	The CDS has the capacity to support 10 concurrent users from an expected user base of 40 users.	**M**
3.3.02	The CDS has the capacity to support concurrently 10 data acquisition channels from an expected 25 total number of channels.	**M**
3.3.03	The CDS has the capacity to support concurrently 10 digital acquisition channels from an expected 25 total number of channels.	**D**
3.3.04	The CDS has the capacity to control concurrently 10 instruments from an expected 20 total number of connected instruments.	**M**
3.3.05	The CDS has the capacity to simultaneously support all concurrent users, acquisition and instrument connects whilst performi ng all operations such as reprocessing and reporting without loss of performance (maximum response time is <10 seconds from sending the request).	**M**
3.3.06	The CDS has the capacity to hold 20 GB of data live on the system.	**D**

Table 2: How System Requirements for CDS Capacity can be Documented

Furthermore each requirement must be written so that it can be tested if required; according to IEEE standard 1233 [13] a well-defined requirement must have the following:

- Capability:
- Condition:
- Constraint:

For more detail on how to write system requirements, refer to the paper by McDowall [14].

Review of the URS

Ideally, an independent group of users (persons not involved in writing the document) should evaluate the URS and challenge each requirement and any interfacing requirements for the chromatographs or any other computer applications. If any missing requirements or inconsistencies can be found at this stage they are easy and inexpensive to correct. Therefore, the extra work in ensuring that the system requirement specification is correct are time and resources well spent; problems that can be rectified at this stage are far cheaper to solve than those identified later in the life cycle. When the system requirements specification is complete, the outline selection tests can be generated that can be used to select a potential system and reused later in the life cycle during the PQ testing.

Validation Plan

The name for this document varies so much from laboratory to laboratory: validation plan, master validation plan or validation master plan or even quality plan. Regardless of what it is called in an organisation it should cover what steps will be taken to demonstrate the quality and compliance of the CDS in the laboratory. Ideally it should be written as early in the process as possible to define the overall steps that are required and the documents to be produced from each. Please see Chapter 5 for more details

System Selection

The purchase of a new CDS system should be a formal selection process to see if an application matches the main requirements of the URS. The outline tests can be used to screen and select the system; an in-house test can be an option if there is sufficient time and resources to do this. A selection report would be the outcome of this phase of the work and would form part of the supporting evidence for the CDS validation.

Supplier Audit

The majority of the system development life cycle for a commercial CDS will be undertaken by a third party: the supplier; this is shown in Figure 6 as all of the operations under the horizontal line. As recommended by the GAMP Guide, Appendix M4 [11] a supplier audit should be undertaken to ensure that the software was developed in a quality manner.

A supplier audit is also a requirement of European Union GMP requirements outlined in Annex 11 [15] namely:

- Clause 11.5: *The software is a critical component of a computerised system. The user of such software should take all reasonable steps to ensure that it has been produced in accordance with a system of Quality Assurance.*

The supplier audit should take place once the product has been selected and the purpose is simply to see if the ISO 9000 quality system is operated effectively. The evaluation and audit process is very important part of the life cycle as it ensures the design, build and testing stages (which are under the control of the supplier) have been checked to ensure compliance with the regulations. The audit should be planned and cover items such as the design and programming phases, product testing and release, documentation and support; a report of the audit should be produced after the visit.

Although many CDS suppliers are certified to ISO 9000 of some description and will offer you a certificate that the system conforms to their quality processes. This is fine but please remember that there is no requirement for product quality in any ISO 9000 schedule and if you look at the warranty of any software product there is no guarantee that the CDS is either fit for purpose or error free. The certificates are fine

but if the system is critical to your operation my advice is to consider a supplier audit.

The details of the format and how to conduct a supplier audit can be found in the GAMP guide [15] and in two papers by McDowall [16, 17].

Requirements Traceability and Risk Assessment

The next stage in the process is to carry out a risk assessment of each function determined on if the function is business and / or regulatory risk critical (C) or not (N). The tables from the URS have two additional columns added to the as shown in Table 4.

Req No.	Data System Feature Specification	Priority M/D	Risk N/C	Test
3.3.01	The CDS has the capacity to support 10 concurrent users from an expected user base of 40 users.	**M**	**C**	**TS05**
3.3.02	The CDS has the capacity to support concurrently 10 data acquisition channels from an expected 25 total number of channels.	**M**	**C**	**TS05**
3.3.03	The CDS has the capacity to support concurrently 10 digital acquisition channels from an expected 25 total number of channels.	**D**	**N**	**-**
3.3.04	The CDS has the capacity to control concurrently 10 instruments from an expected 20 total number of connected instruments.	**M**	**C**	**TS05**
3.3.05	The CDS has the capacity to simultaneously support all concurrent users, acquisition and instrument connects whilst performing all operations such as reprocessing and reporting without loss of performance (maximum response time is <10 seconds from sending the request).	**M**	**C**	**TS05**
3.3.06	The CDS has the capacity to hold 20 GB of data live on the system.	**D**	**N**	**-**

Table 3: Part of a Combined Risk and Analysis and Traceability Matrix for a CDS

When the risk analysis column has been completed, the priority and risk assignments can be assessed together. For a GAMP category 4, such as a chromatography data system, this approach to risk analysis is far simpler to understand and perform than the Failure Mode Effect Analysis (FMEA) process outlined in the GAMP Guide [16].

The risk analysis methodology outlined here is to take only those functions that are classified as both Mandatory and Critical and consider them for testing in the qualification phase of the validation. Therefore functions 3.3.03 and 3.3.06 are not considered for testing, as they do not meet the criteria. Of the remaining four requirements these all constitute capacity requirements that can be combined together and tested under a single capacity test script, which in this example is called Test Script 05 (TS05). In this way, requirements are prioritised and classified for risk and the most critical one can be traced to the PQ test script.

Installation Qualification and Operational Qualification

Installation Qualification (IQ)

Put simply this is the installation of the components of the system with a check that each works correctly. The best people to undertake this work will be the supplier, as they know their products best. However, there could be several groups working on the installation qualification depending on the complexity of the configuration of the CDS: supplier, system administrator from the laboratory, IT department.

For networked CDS systems the following activities would also be required, again depending on the configuration of the system:

- Server (for data storage) installation by the IT department, server supplier or manufacturer
- Installation of the A/D units or data collection servers to the corporate LAN
- Processing or data review workstations either the IT department or contractors working on their behalf (typically with an operating system configured to corporate requirements)
- Network connection of the workstations to the corporate LAN
- Installation of the CDS application software for data processing on the workstations
- Connection of the chromatographs to the A/D units or data collection servers

Many chromatographers are not familiar with the detail of the regulations or guidelines that they are operating under and it is essential to ensure that essential documentation to be collected from these activities is planned and collected proactively. Retrospective documentation of any phase of this work is far more costly and time consuming. Therefore, reiterating the advice given earlier; plan the work in the validation plan otherwise you'll end up with little from this phase of work and a large compliance hole; typically this will involve an installation plan.

Establish the Initial Configuration Baseline

The configuration baseline was established by doing an inventory of the whole system. This resulted in a description of all the parts making up the Millennium system including hardware, software and documentation.

Operational Qualification (OQ)

The operational qualification is carried out after the IQ and is intended to demonstrate that the application works the way the supplier says it will. Most suppliers will supply OQ scripts. These of necessity will only cover a subset of functions and will not be a substitute for the user acceptance tests or PQ tests. Many enterprising suppliers will sell part or the whole of their in-house test suites as their OQ packages, what you have to be aware of is that after completion of the OQ the system will be configured and you may have to execute tests in the PQ that reflect the way the system is now configured.

Typically the OQ is carried out immediately after the IQ and the same person will execute both. Ensure before they start that they are trained to do this and you have documented evidence of this such as a training certificate that is current at the time that the work was carried out.

What should be in an OQ? Here this depends on a supplier and the marketing approach to this “value added” package. Here are my views on the subject: the purpose of an OQ is to show that the software and system works the way that the suppliers state it should.

To understand the purpose of an OQ more fully you need to understand how software is produced. As the FDA acknowledge in their guidance for industry called General Principles of Software Validation [18], the critical phase of development is the design, writing and testing of the application. Production of the software is simply the production of CD media and verification that the disk has been burnt correctly. Therefore the main emphasis in software production is the correct design and release of the system, this is where the supplier’s certificate (or equivalent) of conformance / validation / compliance with their internal procedures is important. Most of the work is done at the supplier’s site and the IQ (have the files been installed in their correct locations) and the OQ (does the software work correctly) are confirmation that the software is the same on your system.

Therefore the amount of OQ testing can be relatively small, as the supplier has carried out the bulk of the work at their development site. The OQ is just a confirmation that the out of the box software works as expected: no configuration will be carried out, as this is the laboratory’s responsibility.

In most cases, the OQ does not need to be very extensive to demonstrate this, especially when the software is to be configured before the PQ is carried out e.g. security options, macros, custom calculations etc. The reason is that extensive testing of the baseline package is of little value, as it will bare little relationship to

the final operating software application.

However, before dismissing any supplier's OQ as a total waste of time and effort, you should, as part of a critical review of the approach, map your requirements to the supplier's package and find out what is being done and can it form a substitute for work you would need to do in the PQ. Some examples may be detailed instrument control functions and where your requirements match what is undertaken in the OQ (typically for simpler software applications). Where there is a lot of laboratory customisation of the application e.g. chromatographic spectral library involving your specific compounds, then the supplier's OQ package is of less or little help here.

Assess Supplier IQ and OQ Documentation

Any documentation provided by a supplier must be critically reviewed. Never accept documentation from a supplier without evaluating it and approving it. Why? Go back to the regulations; look at the 21 CFR 211 current Good Manufacturing Practice requirements under Laboratory Controls and read section § 211.160 subtitled "General Requirements [19]".

> *(a) The establishment of any specifications, standards, sampling plans, test procedures, or other laboratory control mechanisms required by this subpart, including any change in such specification, standards, sampling plans, test procedures, or other laboratory control mechanisms, shall be drafted by the appropriate organizational unit and* ***reviewed and approved by the quality control unit****. The requirements in this subpart shall be followed and shall be documented at the time of performance. Any deviation from the written specifications, standards, sampling plans, test procedures, or other laboratory control mechanisms shall be recorded and justified.*

In essence, there needs to be a written plan that is approved by the quality control or quality assurance group within your organisation before it is executed. However, very few supplier IQ and OQ documents give space for the QC or QA group to sign off that they have reviewed the documentation. However, the regulations go much further.

> *(b) Laboratory controls shall include* ***the establishment of scientifically sound and appropriate specifications, standards, sampling plans, and test procedures*** *designed to assure that components, drug product containers, closures, in-process materials, labeling, and drug products conform to appropriate standards of identity, strength, quality and purity.*

Now you see the reason for assessing the supplier IQ and OQ documentation. The regulations require that before execution the protocols have to be approved by the QC/QA unit and also that whatever is written in them needs to be scientifically sound. That is why you must review this documentation.

Also look at the requirements of the draft guidance for industry on 21 CFR 11 validation [12]; in section 5.4.3, entitled "How Test Results Should Be Expressed" there is the following comment:

> *Quantifiable test results should be recorded in quantified rather than qualified (e.g., pass / fail) terms. Quantified results allow for subsequent review and independent evaluation of the test results.*

Therefore this gives you an additional factor for critical review of what you are purchasing. Explicitly stated acceptance criteria must also be available rather than implying if all expected and observed results match, then system passes.

If in doubt here's an example of someone who did not do what is suggested here; in the warning letter sent to Spolana [20], a Czech company, in October 2000, there is the following citation:

> *Furthermore, calibration data and results provided by an outside contractor were not checked, reviewed and approved by a responsible Q.C. or Q.A. official.*

Performance Qualification (PQ)

The PQ stages of the overall qualification of the system can be considered as the acceptance testing (this can also be called end user testing), undertaken by the users and based upon the way that the system is used in a particular laboratory. Therefore, a CDS cannot be considered validated simply because another laboratory has validated the same software: the operations of two laboratories may differ markedly even within the same organisation.

The functions to be tested in the PQ must be based on the requirements defined in the URS and with the numbering of individual requirements can be traced back to the system requirements. The main issue is how users can test software?

PQ Test Plan and Test Scripts

One way to document the is using an overall PQ test plan that outlines the features of the CDS to test and those that will not be tested and a discussion of the assumptions, exclusions and limitations of the testing undertaken. A documentation standard for the PQ test plan can be found in the IEEE standard 829-1998 [21] presented in Table 4.

1. Test plan identifier;
2. Introduction;
3. Test system/item
4. Features to be tested;
5. Features not to be tested;
6. Approach to be adopted
7. Pass/fail acceptance criteria for all features to be tested
8. Suspension criteria and resumption requirements;
9. Test deliverables;
10. Testing tasks;
11. Environmental needs;
12. Responsibilities;
13. Staffing and training needs;
14. Schedule (Test order)
15. Risks and contingencies;
16. Approvals.

Table 4: Outline of a Test Plan from IEEE Standard 829-1998

The key sections of a PQ test plan are the features to test and those that will not be tested and associated with the features to be tested are the written notes of the assumptions, exclusions and limitations to the testing undertaken. The assumptions, exclusions and limitations of the testing effort were recorded in the appropriate section of the qualification test plan to provide contemporaneous notes of why particular approaches were taken. This is very useful if an inspection occurs in the future, as there is a reference back to the rationale for the testing. It is also very important as no user can fully test a CDS or any other software application. For example, the operating system was explicitly excluded from testing as the CDS application software implicitly tested this.

Release notes for the CDS application version being validated will document the known features or errors of the system. PQ tests carried out in any validation effort should not designed to confirm the existence of known errors but to test how the system is used by the users on a day to day basis. If these or other software errors were found during the PQ testing, then the test scripts have space to record the fact and describe the steps that were taken to resolve the problem.

PQ Test Scripts

In the same IEEE standard [21] can be found the basis for the test documentation that is the heart of any PQ effort i.e. the test script; in essence this document will:

- Outline one or more test procedures that are required to test the CDS functions

- Each test procedure will consist of a number of test steps that define how the test will be carried out
- For each test step the expected results must be defined
- There will be space to write the observed results and note if it the test step passes or fails when compared with the expected results
- There is a test log to highlight any deviations from the testing
- Sections will collate any documented evidence produced during the testing; this includes both paper and electronic documented evidence
- Definition of the acceptance criteria for each test procedure and if the test passes or fails
- A test summary log collating the results of all testing
- A sign off of the test script stating if the script has passed or failed

Testing Overview

One key point is that to ensure that the PQ stage progresses quickly, a test script should test as many functions as possible as simply as possible (great coverage and simple design). Software testing has four main features, known as the 4Es [22]:

- Effective: demonstrating that the system tested meets both the defined system requirements and also finds errors
- Exemplary: test more than one function simultaneously, where feasible
- Economical: tests are quick to design and quick to perform
- Evolvable: able to change to cope with new versions of the software and changes in the user interface

Manual or Automated Testing?

By the way, if a laboratory is tempted to use an automated test tool for their PQ execution consider Graham's words on the subject [22]:

- Automated testing tools take longer to use the first time compared to manual testing
- Expectation will exceed the delivery
- To be economical the test suite must be reused many times
- Automated tools are best used for regression testing (to see if operation of the software remains the same after change)
- Automated testing is not a substitute for manual testing

Therefore don't use automated testing tools for the PQ as they will cause more problems than they will solve. If a supplier offers an automated tool for the IQ and or OQ, then this will probably be useful, as it will establish if the system has been installed correctly and the software functions as the supplier intended it to. However, evaluate the tool critically to see that it meets your needs and is compliant with GXP as described earlier in this chapter.

Write the Test Scripts

Dependent on the complexity of the system requirements, the overall architecture of the CDS and whether electronic signatures have been implemented the number of PQ test scripts needed for a chromatography data system typically falls in the range of 15-30 to provide adequate coverage for the important functions documented in the URS.

Typically the test scripts will cover the following areas of system functionality:

Chromatography Data System Functions e.g.

- Data acquisition from the different types of chromatograph interfaced to the system
- Crosstalk of A/D converters [23]
- Calibration methods used within the laboratory: are they mathematically correct
- Analyte calculation
- System suitability test parameters
- Reporting data
- Sample continuity

- Unavailability of the network: buffering of the A/D or data collection devices
- Remote processing over the network
- Data acquisition and data processing using a diode array detector (DAD) and / or dual wavelength detector
- Creation and management of DAD spectral libraries
- Custom calculations implement calculations on data
- Macros used to perform functions automatically
- System capacity tests e.g. analysing the largest expected number of samples in a batch, were incorporated within some test scripts to demonstrate that the system was capable of analysing the actual sample volume that could be expected in the laboratory.
- Interfaces between the CDS and other software applications e.g. LIMS

21 CFR 11 and other Regulatory Requirements e.g.

- Preservation of electronic records e.g. Backup and Recovery; Archive and Retrieve
- Data file integrity
- System security and access control including between departments or remote sites
- Audit trail
- Date and time stamps
- Electronic signatures
- Identifying altered and invalid records

Outline Test Case Design

The considerations for designing stress and capacity tests for a CDS will be discussed here and will be based on the client-server architecture shown in Figure 3. Note that all requirements must be written in the system requirements specification.

Analytical Run Capacity

First consider an analytical run and the capacity test considerations that will need to be evaluated. You'll know from the URS the maximum number of vials that you'll inject in a single run, this will include standards, samples, quality control and blank reagents that you may run as part of your normal procedures. A test should be designed to run the maximum samples including replicate injections.

Analogue to Digital Unit Capacity

Depending on the type of A/D unit this test can have one or more of the factors that will be discussed below:

- Crosstalk: if two or more channels are multiplexed through a single A/D chip, then a crosstalk test is recommended to see the impact of an overloaded signal on one channel impacts another.
- Data Acquisition Rate: compare the specified rata acquisition rate for a data server to the data rate of chromatographs attached to the unit including any diode array detectors.

In both or either of these instances the validation team may decide that the total data rate is close to the specification of the unit and test this to ensure that the A/D unit is not compromised during normal operation. If the data rate is far below specification then an alternative path you may decide is not necessary to test and to document a rationale for this approach that is scientifically sound. Balancing the regulatory risks is one of the factors in computerised system validation, do you want to do this or test this function?

Unavailability of the Network

There will be times when the network is unavailable and data will be buffered in the A/D unit or data server. You'll want to ensure that this function works during the PQ or you will have failed in your due diligence. The worst-case example for the buffering will be defined in your URS and will be the number of injections with the longest run time. The run should be started, then the network is disconnected and the data accumulated in the A/D unit or acquisition unit until the end of the run when the network is reconnected and the buffered data are transferred to the server. There should be no loss of data integrity in any of the buffered and transferred files if this test is to pass.

System Capacity
The capacity of the system needs to be tested in a way that reflects on the way the system will be used and there are several approaches to take. If you have a 30 user license then one of the simplest ways of assessing the capacity is to run all systems simultaneously, however this will only test the data acquisition and transfer to the sever via the network. As the A/D units, buffer acquired data until transferred to the server this test will also implicitly evaluate the transfer with the network traffic at the time of the test. However, one of the main causes of performance degradation will be integration of data and this must also be included as part of any test of system capacity.

Logical Security and Access Control
Whilst logical security appears at first glance to be a very mundane subject, the inclusion of this topic as a test is very important for regulatory reasons as it is explicitly stated in 21 CFR 11. Also when explored in more depth it provides a good example in the design of a test case.

The test design consist could consist of three basic components:
1. A test sequence where the incorrect account fails to gain access to the system
2. A single test case where the correct account and password gain access to the system
3. A test sequence where the correct account but minor modifications of the password fail to gain access to the software

The important considerations in this test design are:
- Successful test cases are not just those that are designed to pass but also are designed to fail. Good test case design is a key success factor in the quality of validation efforts. Of the test cases above 75% are designed to fail, to demonstrate the effectiveness of the logical security of the system
- The test relies on good practices to ensure that users change or are forced to change their passwords on a regular basis and that these are of reasonable length (minimum 6-8 characters).
  1.

Other test case designs are defined below:
- Boundary test: the entry of valid data within the known range of a field e.g. a pH value would only have acceptable values within 0-14.
- Stress test: entering data outside of designed limits e.g. a pH value of 15.
- Predicted output: knowing the function of the module to be tested, a known input should have a predicted output.
- Consistent operation: important tests of major functions should have repetition built into them to demonstrate that the operation of the system is reproducible.
- Common problems: both on the operational and support aspects of the computer system should be part of any validation plan e.g. backup works, incorrect data inputs can be corrected in a compliant way with corresponding audit trail entries. The predictability of the system under these tests must generate confidence in the CDS operations (Trustworthiness and reliability of electronic records and electronic signatures) and the IT support.

The format of the document and more detail of PQ testing, the articles on the retrospective and prospective validations of CDS systems are recommended [23, 24].

Personnel and Training Records

All involved with the selection, installation, operation and use of a CDS should have training records to demonstrate that they are suitably qualified to carry out their functions and maintain them. It is especially important to have training records and curricula vitae of installers and operators of a system as this is a particularly weak area and a system can generate an observation for non-compliance.

Major suppliers of CDS will usually provide certificates of training for installation of the system and software. However, a major weak spot with many CDS that have the IT Department running the system do not have training records or curricula vitae.

The types of personnel involved that could be involved in a validation are:
- Suppliers staff: who were responsible for the installation and initial testing of the data system software, left copies of their training certificates listing the products they were trained to work on. These were

checked to confirm they were current and covered the relevant products and then included in the validation package.
- System managers: training in the use of the system and administration tasks were provided by the supplier and documented in the validation package.
- Users: were either analytical chemists or technicians whom had their initial training by the supplier staff to use the data system and this was documented in their training records.
- Consultants: any consultants involved in aiding a validation effort must provide a curriculum vitae (resume) and a written summary of skills to include in the validation package for the system.
- IT Staff: training records and job descriptions outlining the combination of education, training and skills that each member has.

Training records for CDS users are usually updated at the launch of a system but can lapse as a system becomes mature. To demonstrate operational control, training records need to be updated regularly especially after software changes to the system. Error fixes do not usually require additional training, however major enhancement or upgrade should trigger the consideration of additional training. The prudent laboratory would document the decision and the reasons not to offer additional training in this event.

To get the best out of the investment in a CDS, periodic retraining, refresher training or even advanced training courses could be very useful for large or complex ones. Again this additional training should be documented.

Service Level Agreement

In the case of outsourcing the support for the hardware platforms and network that run the chromatography data system software to the internal IT Department, a Service Level Agreement (SLA) has to be written. This SLA should cover procedures such as:
- Backup and recovery
- Archive and restore
- Storage and long term archive of data
- Disaster recovery.

This SLA will cover the minimum service levels agreed together with performance metrics so that they can be monitored for effectiveness.

System Documentation

Documentation

The documentation supplied with the CDS application or system (both hardware and software), user notes and user standard operating procedures will not be discussed here as it is too specific and also depends upon the management approach in an individual laboratory. However, the importance of this system specific documentation for validation should not be underestimated. Keeping this documentation current should be considered a vital part of ensuring the operational validation of any computerised system. The users should know where to find the current copies of documentation to enable them to do their job. The old versions of user SOPs, system and user documentation should be archived.

Standard Operating Procedures (SOPs)

Standard Operating Procedures are required for the operation of both the CDS applications software and the system itself; as explained above, we not consider user SOPs in detail. SOPs are the main medium for formalising procedures by describing the exact procedures to be followed to achieve a defined outcome. According to Hambloch [25], SOPs have the advantage that the same task is undertaken consistently, is done correctly and nothing is omitted and a written procedure means that new employees are trained faster. The aim is to ensure a quality operation. Laboratory staff are used to working with SOPs, however if a central computer group supports a large system they may not be used to working with SOPs and even less ready to document their work. However, to provide a service to a regulated laboratory, a computer

department must provide a suitably documented procedure. Indeed this is a requirement under EU GMP Annex 11 [15], where a third party supplier should have a documented operation.

According to Hambloch [26] there is a minimum list of twelve SOPs required for the operation of a computer system in a regulated or accredited laboratory. These are:

- SOP on SOPs: this should describe the approach taken to the writing of SOPs within the functional group, the sections, who can authorise the procedure, description of the procedure and distribution list.
- Description of responsibilities: the roles and responsibilities of staff supporting the computer system are defined.
- System description of hardware and change control procedures: describes how the hardware components will be maintained (equivalent to the hardware configuration log) with the procedure to be adopted when the system configuration is changed.
- Preventative maintenance: described the procedures for preventative maintenance of the hardware components
- Prevention, detection, and correction of errors: the measures and procedures for finding, recording and resolving errors in the system. This can be a complex SOP covering many different aspects of the system and may refer to sections of the technical manuals provided with the system. This SOP includes good housekeeping such as disc defragmentation or monitoring the space available on all discs.
- System boot and shutdown: This is a special SOP that should contain all the specific instructions for starting up and shutting down the system. This SOP may be required in an emergency and therefore should be written well and be easily available for use.
- Control of environmental conditions: For systems that require a controlled environment, an SOP that defines the acceptable ranges of temperature, humidity, and power supply. Other environmental considerations may be what to do in the case of electrostatic discharges, power surges, fire, lightening strikes or the use and maintenance of an uninterruptible power supply (UPS).
- Contingency plans and emergency operation: this a disaster recovery plan and the use of alternative plans until the computer system has been recovered. It is important that any disaster recovery plan is tested and verified that it works before any disaster occurs. This is covered in more detail in section 9.
- Backup and restore of data: Describes the procedures for backup of data and software programs and how to restore data to disc.
- Security: The logical (software) and physical security of the system is covered with the procedures for setting up and maintaining security.
- Installation and update of software: Procedures to be undertaken before, during and after installing software. This should start with the complete backup of all discs and then installation of the software and any testing and validation that may be required.
- Development and update of system software procedures: Software can be written to control the system or help execute functions, this SOP outlines the procedures for the creation, documentation and modification of these procedures

The reader is referred to the article by Hambloch [25] For more details on these SOPs. However, it is important to realise that the list above refers to a relatively large computer system that is run by a centralised IT group. Therefore, for smaller items of laboratory computer equipment the list should be reviewed for applicability and suitability. Where a system does not have the facility to store raw data e.g. disc drive, then no SOP is required for backup and restore. The same logic should be applied to the whole list. The converse is also true, this is a generalised list of SOPs, and if there is a specialised application there may be the need for more SOPs than appears above.

Write Validation Summary Report

The validation summary report brings together all of the documentation collected throughout the whole of the life cycle and presents a recommendation for management approval when the system is validated. The emphasis is on using a summary report as a rapid and efficient means of presenting results as the detail is contained in the other documentation in the validation package; see Chapter 11 for more details.

RETROSPECTIVE VALIDATION

Guidance for retrospective validation in the literature tends to be presented in overview only, for example Huber [26] outlines the approaches for retrospective evaluation of computerised analytical instruments. Whilst some of the approaches outlined by Huber and those used in this paper are similar in some cases (e.g. document collection, describe and define the system, qualify the system, update documentation) there are some differences as well.

In the paper by Wikenstedt el al [23], a retrospective validation of a CDS is described in detail. The key difference between a prospective and a retrospective CDS validation is the gap and plan phase.

Gap and Plan for Retrospective Validation

The Gap and Plan phase is an essential stage in the retrospective validation of any computerised system.

Collect Existing CDS Documentation

First all of the existing documentation on the system must be collected; this could include items such as:

- Validation plan,
- URS,
- Documentation from the selection process,
- Purchase order, packing lists,
- Qualification tests and documentation,
- PQ tests,
- Training materials,
- Operating manuals both in-house and from the supplier,
- Standard Operating Procedures.

In this example, the system was relatively new and most of the available documentation was retrieved, as documentation was easily available. Furthermore, the personnel operating the system have been involved with the project from the start. This is in contrast with a system that may be much older where documentation may be non-existent and personnel may have left the company or indeed the company has reorganised or merged.

When all the documentation has been collected, a list is made. This can be compared against the current regulatory regulations, industry guidelines and the corporate validation policy. This generates a list of missing documents and defines the gap to be filled.

Review Existing Documents

Next, the existing documentation must be reviewed to see that each item is of suitable quality, coverage and fitness for purpose. The mere existence of a document does not mean that its quality and coverage is good. Poor documents must be completed, or otherwise discarded and a new one written that meet the current compliance requirements. For instance, if there is a current system requirements specification (URS) is it specific enough to allow qualification tests to be constructed? If a URS consists of one or two pages of general statements for a data system, such as

- The data system performance must be fast
- User friendly operation

This means that there is no firm requirement to allow a meaningful test to be constructed. The assessment of documents may result in more documents being added to the gap list.

Planning to Bridge the Gap

Once the gap has been defined, there must be a decision made to either write the key documents and fill the gap or for management to take the business risk not to write them, if they are not available. Times and resources must be included in this plan. This list of documents to be written, authorised by management, is the output of the Gap and Plan phase.

The Gap and Plan identified that there were several key documents that were required in the example [23]; these were:

- Validation plan
- Workflow analysis
- User requirements specification
- Test plan for the qualification of the system
- User test scripts (Performance Qualification)
- Change control and configuration management SOP
- System description

The process for the retrospective validation is to write these documents and execute any PQ testing as necessary.

MAINTAINING THE VALIDATION STATUS DURING OPERATIONAL LIFE

After operational release comes the most difficult part of computerised system validation; maintaining the validation status of the system throughout its whole operational life. Look at the challenges that will be faced when dealing with maintaining the validation of a CDS or indeed any system; some of the types of changes that will impact an operational CDS are:

- Software bugs will be found and associated fixes installed
- Application software, operating system, plus any software tools or middleware used by the CDS will be upgraded
- Network improvements: changes in hardware, cabling, routers and switches to cope with increased traffic and volume
- Hardware changes: PCs and server upgraded or increases in memory, disk storage etc
- Interface to new applications e.g. spreadsheets or laboratory information management systems (LIMS)
- Expansion or contraction of the system due to work or organisation reasons
- Environmental changes: moving or renovating laboratories

All of these changes need to be controlled to maintain the validation status of the CDS.

In addition there are other factors that impact the system as well from a validation perspective, such as:

- Problem reporting and resolution
- Software errors and maintenance
- Backup and recovery of data
- Archive and restore of data
- Maintenance of hardware
- Disaster recovery (business continuity planning)
- Written procedures for all of the above

In this section, the number of measures will be discussed that need to be in place to maintain the validation status of a chromatography data system.

Change Control and Configuration Management

Changes will occur throughout the lifetime of the system from a variety of sources such as:

- Upgrades of the CDS software
- Upgrades of network and operating system software
- Changes to the hardware: additional memory, processor upgrade, disc increases etc.

- Extension of the system for new users

This is the key item from the installation of the system to its retirement. Changes must be controlled. From a regulatory perspective there are specific references to the control of change in both the OECD consensus document [27] and EU GMP regulations [15].

Change control was implemented through an SOP that defined the procedure for change control. A change form was the means of requesting and assessing change:

- The change requested was described first by the submitter.
- The impact was assessed by the system managers and then approved or rejected by management.
- Changes that were approved were implemented, tested and qualified before operational release.

The degree of re-validation work to be done was determined during the impact analysis. Changes that impacted the configuration (hardware, software and documentation) were recorded in a configuration log maintained within Excel.

Operational Logbooks

To document the basic operations of the computer system a number of logbooks are required. The term logbook is used flexibly in this context; the actual physical form that the information takes is not the issue, rather the information that is required to demonstrate that the procedure actually occurred. The physical form of the log can be a bound notebook, a pro-forma sheet, a database or anything else that records the information needed, as long as security and integrity of the records (paper or electronic) are maintained.

Backup Log

The aim of a backup log is to provide a written record of data backup and location of duplicate copies of the system (operating system and application software programs) and the data held on the computer. The backup schedule for the discs can vary. In a larger system, the operating system and applications software will be separated from the data that are stored on separate discs. The data change on a fast timescale that reflects the progress of the samples through the laboratory and must be backed up more frequently. In contrast, the operating system and application programs change at a slower pace and are therefore more static; the backup schedule can therefore reflect this.

For smaller systems, such as personal computers, the data and programs may be located on the same disc and partitioned by the directory structure. If the backup software is capable of performing selective backups then the comments in the paragraph above apply. However, if there is little sophistication the whole disc may have to be backed up routinely. Again, for PC systems this may be an area to evaluate closely before buying. An alternative is a PC network, where the programs and data are held on a central server and can be backed up more efficiently and effectively than stand alone systems.

- Some of the key questions to ask when determining the backup of your chromatography data system are:
- How long should the time between backups be? This can be answered by considering how much data the laboratory can afford to use. If it is up to a week (most unlikely), then the backups can be weekly but typically it is daily. If, you cannot afford to loose any data, then shadowing or duplicate discs are the start of the answer that may lead you to consider RAID (Redundant Array of Inexpensive Discs) technology.
- Who is authorised to perform backups and who signs off the log? The laboratory manager in conjunction with the person responsible for the system should decide this. The authorisation and any counter signature required should be defined in an SOP
- When should duplicate copies be made for security of the data? This question is related to the security of your data and programs. Duplicate copies should be part of the backup procedure at predetermined intervals. The duplicate copies should be stored in a separate location in case of a hazard to the computer and the original backups located nearby. Duplicate backups are also necessary to overcome problems reading the primary backup copies.

Problem Recording and Recovery

During the operation of a computer system, boot up, backup or other system functions, it will be inevitable that errors may occur. It is essential that these errors are recorded and the solution to resolve it also written down. Over time, this can provide a useful historical record to the operation of the computer system and the location of any problem areas in the basic operation.

Areas where this may be the case may be in peripherals where a print queue has stalled. This is relatively minor, however there may be cases where the application fails due to a previously undetected error. In the latter case, there is a need to for link the error resolution to the change control system.

Software Error Logging and Resolution

As it is impossible to completely test all of the pathways through CDS software or any software [27], it is inevitable that errors will occur during the operation of the system. These must be recorded and tracked until there is a resolution. The key elements of this process are to record the error, notify the support group (in-house or supplier), classify the problem and identify a way to resolve it.

Not all reported problems of a CDS will be resolved, they might be minor and have no fundamental effect on the operation of the system and may not even be fixed. Alternatively a work around may be required which should be documented, sometimes even retraining may be necessary. Other errors may be fatal or major, that mean the system cannot be used until fixed. In these cases, the revalidation policy will be triggered and the fix tested and validated before the CDS can be operational again.

Maintenance Records

All quality systems need to demonstrate that the equipment used is properly maintained and in some instances calibrated. Computers are no exception to this. Therefore records of the maintenance of the CDS need to be set up and updated in line with the work carried out on it. The main emphasis of the maintenance records is towards the physical components of a system: hardware, networking and peripherals; the software maintenance is covered under the error logging system described above.

If the hardware has a preventative maintenance contract, the service records after each call should be placed in a file to create a historical record. Also any additional problems that occur that requires maintenance will be recorded in the system log and there will need to be cross-references to the appropriate record there.

Many smaller computer systems have few if any preventative maintenance requirements but this does not absolve the laboratory from keeping records of the maintenance of the system. If a fault occurs that requires a service engineer to visit, then this must be recorded as well.

On sites where maintenance of personal computers is maintained centrally for reasons of cost or convenience, maintenance records may be held centrally. The remit of the central maintenance group may cover all areas of a site or organisation including regulated or accredited as well as non-accredited groups. It is important for the central maintenance group to maintain records sufficient to demonstrate to an inspector of the work they undertake. As defined in EU GMP Annex 11 [15], the third party undertaking this work should have a service agreement and also have the curriculum vitae of its service personnel available and up to date.

Disaster Recovery

Good computing practices require that a documented AND tested disaster recovery plan must be available for all major computerised systems. It rarely is. Failure to have a disaster recovery plan places the data and information stored by major systems at risk, the ultimate losers being the workers in the laboratory and the organisation.

Disaster recovery is usually forgotten, or not considered, as "it will never happen to me". The recovery plan should have several shades of disaster documented. From the loss of a disc drive: how will data be

restored from tape or backup store and then updated with data not on backup, through to the complete loss of the computer room or building through fire or natural disaster.

Once the plans have been formulated, they should be tested and documented to see if they work. Failure to test the recovery plan will give a false sense of security and compound any disaster.

Revalidation Criteria

Any change to a CDS should trigger consideration if revalidation of the system is required. Note the use of the word "consider". There is usually a knee-jerk reaction that any change means that the whole system should be revalidated. One should take a more objective evaluation of the change and its impact before deciding if full revalidation is necessary.

Firstly, if revalidation is necessary, to what extent is it required to test: a software unit, module or the whole system? Thus revalidation is defined by Chapman as *"repetition of the validation process or a specific portion of it"* [28]. There may even be instances where no revalidation would be necessary after a change. However the decision must be documented together with the rationale for it.

Therefore a procedure is required to evaluate the impact of any change to a system and act accordingly. One way to evaluate a change is to review the impact that it would make to data accuracy, security and integrity outlined by Lepore [29]. This will give an indication of the impact of the change on the system and the areas of the application affected. This allows you to target the revalidation effort that is appropriate to the change you are going to make.

CDS DATA MIGRATION AND SYSTEM RETIREMENT

Rationale for Data Migration

Data migration and system retirement occur at the end of the life cycle of any computerised system, however there is little or no directly stated regulatory requirements for formal system retirement nor general advice on how to undertake the task until recently with the 21 CFR 11 guidance on preservation of electronic records [30]. Retirement in many instances may be a euphemism for simply throwing the system components out of the company, however we will highlight reasons in this paper to justify that a more formal approach should be taken.

Data migration will be necessary for a number of reasons e.g.:

- Change in data processing algorithms following a software upgrade of an application
- Change to use of a different software application
- Change in computing environment such as operating system or computing platform
- Change in data file formats

Data migration is required for the duration of the records retention period under the electronic records and electronic signatures final rule (21 CFR 11) [30,31]. The problem is how should this be achieved to allow ready replay of data? What will be the impact on calculated results when, date file formats, calculation algorithms and computing platform change?

Data migration can be the worst part of computerised system validation as a system generating the data will have been operational for a number of years, the data may be shared between several departments and the original staff involved with the project no longer work in the organisation. This can be compounded where there have been reorganisations within a firm and the system boundaries are different compared with the original installation. Fortunately in this case, the data were generated within a single department with a single system owner, making the project simpler than other comparable data migration projects.

The example [32] describes the experiences designing and validating a mass spectrometry data migration between two different platforms. The triggering event was the decision by the supplier of the mass spectrometry equipment and application software to move to a new computing platform and declare the current one obsolete.

Study 2: Chromatography Data Systems

This project was conducted using the life cycle approach to validation of chromatography data systems (CDS) as described by McDowall [2, 33, 34] and consisted of three strands of work under a single validation plan as shown in Figure 7. These strands of work were:

1. Prospective validation of the new application software (Analyst version 1.0) and qualification of new instruments associated with them
2. Validation of the migration of electronic records generated using MassChrom software on the Macintosh systems to the new Analyst NT environment as well as data acquisition on some Macintosh platforms with interpretation using Analyst software
3. Formal retirement of obsolete mass spectrometry and Macintosh computer hardware

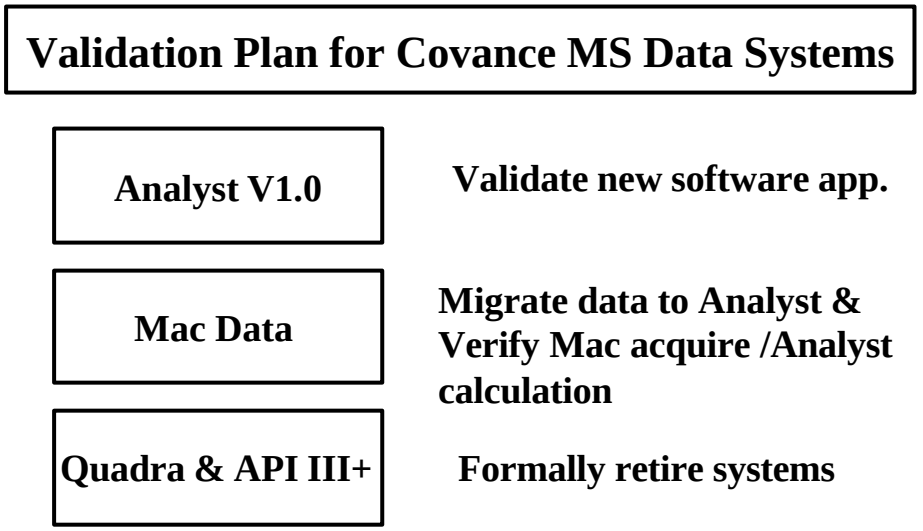

Figure 7: Overview of the whole mass spectrometry validation, data migration and system retirement project

Overview of the Bioanalytical Mass Spectrometry Systems

The mass spectrometry equipment, current software options and computing environment within Bioanalytical Services is presented below and summarised in Table 5 and Figure 8.

Mass Spectrometry Equipment

There are three main models of mass spectrometer currently operating in the Bioanalytical Services Department: API models III+, 365 and 3000. Of these, the API III+ is obsolete as the Macintosh PC used to run the software is no longer in production. Therefore, the three systems using the API III+ mass spectrometer will be formally retired and only the API 365 and 3000 models will be used thereafter.

Data Acquisition and Processing Software Applications

The MassChrom mass spectrometer software currently used in the in the Department is a combination of data acquisition software (three versions of RAD and sample control) and data processing software (two versions) that operates on the Macintosh plus the Analyst software designed for the Windows NT environment. The RAD and MacQuan software running on the Macintosh Quadra will be retired under the work described in this paper.

A mixed environment will be operated for a transition period where data are acquired by sample control running on a Macintosh but all data processing and quantification run on the Analyst. In the future, after retirement of all Macintosh computers, there will be an environment that is only Analyst running on Windows NT.

Computing Environments
The current environment was Macintosh with mass spectrometry being downloaded to a server after it had been acquired. Introduction of the Analyst has started a migration to an NT operating environment that will continue after the completion of the data migration outlined here.

Mass Spectrometry Instrumentation	Computing Hardware	Operating System	Data Acquisition Software	MS Quantification Software
API III+	Mac Quadra	Mac OS	RAD 2.6	MacQuan 1.4
API III+	Mac Quadra	Mac OS	RAD 2.6	TurboQuan 1.0
API 365	Power Mac	Mac OS	Sample Control 1.3	MacQuan 1.4
API 365	Power Mac	Mac OS	Sample Control 1.4	MacQuan 1.4
API 365	Power Mac	Mac OS	Sample Control 1.4	TurboQuan 1.0
API3000	Dell PC	Windows NT	Analyst v1.0	Analyst v1.0

Table 5: Data Processing Options available in Bioanalytical Services Department.

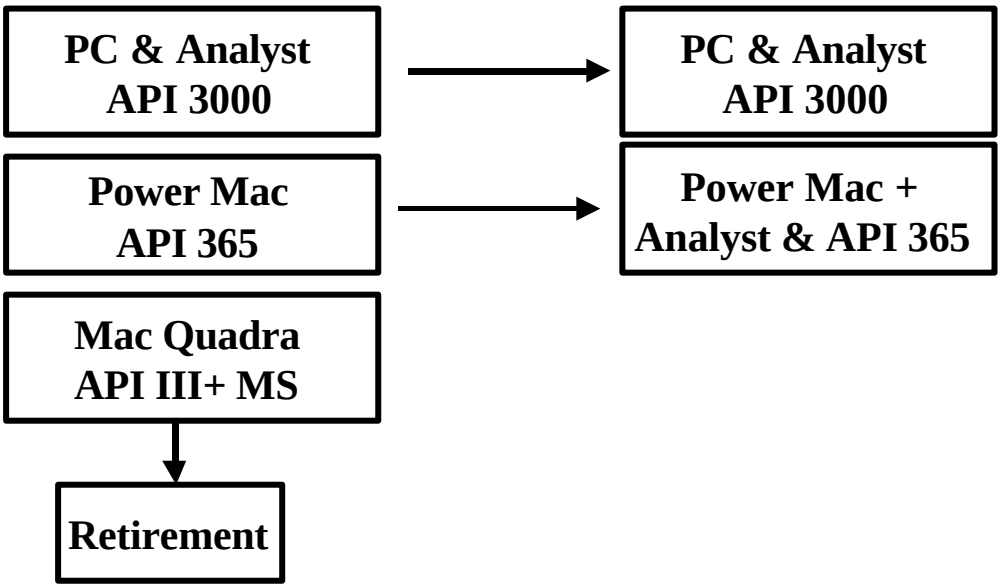

Figure 8: Overview of Mass Spectrometry Equipment and Data Systems

Differences between the Two CDS Systems

It is vitally important to understand the differences between the two environments before progressing further with any data migration. Covered here are the major differences between the two systems and their impact on the data migration. Essentially the problem is that we have incompatible:

- Hardware
- Operating system
- Application software
- Data file formats and
- Application design philosophies

These differences will be discussed below, however the bottom line is that data file conversion is essential for the data migration to succeed.

Computing Platform Differences
The Macintosh and Intel hardware computing platform and operating system software are essentially incompatible. An emulator is needed to run Windows software on a Macintosh, but there is no corresponding emulator for the Macintosh in a Windows environment that will run the software and be supported by the supplier.

Raw Data File Format Differences
The file formats for the chromatograms produced by the same instrument in the two environments are completely different. The Macintosh uses a different file format compared with the Analyst that uses WIFF (Waveform Interchange File Format) file format and can have either single or multiple WIFF files. For the work described here, only the use of multiple WIFF files was evaluated.

Meta Data File Format Differences
The MassChrom software requires three files to set up and acquire data: the Method, State and Experiment files. The method and experiment files are used to set up and acquire mass spectrometer data and the experiment and state files used to monitor the performance of the mass spectrometer itself.
In contrast, there are just two such files used within the Analyst: data acquisition method (DAM) and instrument (INS) files. The mapping of the MassChrom and Analyst files is not one to one: parameters in the experiment file are split between the INS and DAM files on the Analyst application.

Design Philosophy of the Macintosh and NT Software Applications
Although the software running on the two platforms can control the same mass spectrometry instruments, their designs are very different. The MassChrom software was designed in the early 1990s for operators with mass spectrometry training; the terminology and instrument set up within the applications are specialist for trained mass spectrometrists.

Over time the instrument has been used more widely by chromatographers and the Analyst software is a response to this as the operation of the application is simpler and uses chromatographic terms more than mass spectrometry ones. This difference in design philosophy is a complicating factor for the data migration, as terms have to be mapped between the applications, as we will describe later in this paper.

Generic Data Migration and System Retirement Process

A generic seven-step process, shown in Figure 9, describes system retirement and migration of data. Each stage will be described in overview and is a summary of the work described by McDowall [34].

Step 1: Inventory of the System
Identify the scope and boundaries of the system and the departments who use the system. Part of this may be the fact that the system may be spread across buildings and even networks. The latter is an issue, as it can complicate the initial work as data spread over different networks will have to be collated to find out the data volumes and projects/studies involved.

Step 2: Carry out a Risk Assessment
How critical is the system? This determines the level of regulatory risk and data criticality and is used to determine the detail required in the remainder of the process.

Step 3: Write the Retirement Plan
Using the data generated from step 1, the plan covers:
- Scope and boundaries of the chromatography data system(s)
- Roles and responsibilities
- Outline project plan
- Process of system retirement
- Process of data migration

Step 4: Detailed Information Gathering
In this part of the process you will need to know the details of the computer hardware including any specialised devices, the software and the documentation associated with the system as well as the data. The

data need to be identified in detail, for example: how many tapes are involved (if your long-term storage is on tape), what data relating to which samples are on a specific tape.

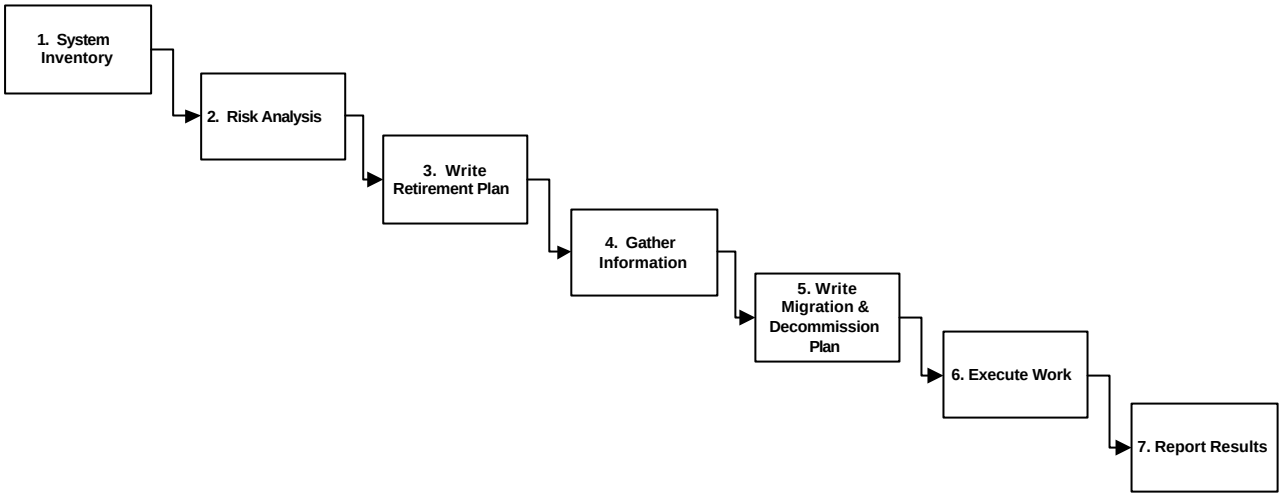

Figure 9: Generic Process for Data Migration and System Retirement

Step 5: System Decommissioning and Data Migration Plan
This document is a detailed presentation of the approach you'll be undertaking on the system and describes the roles and responsibilities of people involved in the work, the systems, the data to migrate, the test scripts needed and what each test script will contain to document the process.

Step 6: Execute Work and Document Activities
Following the tasks described in the decommissioning plan, the data retirement will start first to be followed by the system retirement. You will need to write any scripts to check and document the correctness of the data transfer; this is a critical stage in generating the confidence in the process. Once the data have been successfully migrated and or archived, then you will turn your attention to turning off the hardware and reusing it or removing it from site. Again, this will be documented as the process continues.

Step 7: Write Retirement and Migration Report
This is simply a summary of the work that was undertaken with a description of any deviations from the plan and a discussion of their impact. The data migration together with any validation tests applied will be described and management will sign off the report.

Data Migration Strategy

The options for data migration are to assess if it is technically feasible to migrate data. The supplier of the mass spectrometry software systems (Applied Biosystems/MDS Sciex) provides conversion programs to allow a user to migrate electronic records from the Macintosh to the Analyst system. Conversion is necessary, as the file formats are completely different between the Macintosh and NT environments.

Supplier Supplied Data Conversion Utilities

Three API File Converter programs were supplied for the conversion of the Macintosh format data and meta data files by Applied Biosystems, the software supplier, these are:

- File Translator: Data file conversion program that takes Macintosh formatted data files and converts them to single or multiple Analyst format files (WIFF).
- InstFileGenerator: Instrument file conversion program combines Macintosh state and calibration files and generates an Analyst instrument file (INS file).
- ExptFile Converter: Experiment file conversion program combines a Macintosh state file and a Macintosh experiment file and generates an Analyst data acquisition method file (DAM).

Therefore, it is technically feasible to convert the data and migrate them into the NT environment, the question now becomes 'are all data converted or are files converted on an as needed basis'? The data volume involved is in the range of 100-200 GB of data.

Limitation of the Data Conversion Utilities

These utilities have a number of limitations that were not apparent during the early stages of this work:

- They only work on a PowerMac, therefore the objective of retiring all Macintosh computers cannot be realised as at least one is required to run the data conversion utilities
- The utilities cannot convert RAD version 2.6 files. Only the chromatograms can be converted but the experiment, method and state files cannot and the data contained therein must be manually input into the Analyst. Therefore, in the case of data collected under RAD version 2.6, the requirements of 21 CFR 11 for ready replay of data cannot be met.
- A further limitation of the utilities became apparent during the data migration in that the original baselines were not transferred and new baselines redrawn with the new system.

Data Migration Options

There are essentially two options for the migration of the data from the MassChrom environment:

- Convert all data into the new data format now
- Convert selected data on an "as needed" basis

The second option was chosen for a number of reasons including the time and cost of conversion. However, two main issues arise from this approach:

- The laboratory is totally reliant on the supplier's conversion utilities and their continued maintenance of them over time
- The conversion utilities must be tested to confirm that they continue to operate as expected after every software upgrade

Evolution of the Data Migration Design

It is important to understand that a data migration project requires a full understanding of the problem. Therefore, this section of this paper is intended to provide a measure of the evolution of the data migration project as the understanding of the extent of the issues involved increases.

Initially, a single test script under the Analyst validation was envisioned. However, as the complexity of the MassChrom software versions was understood, a data migration and system retirement test plan was required to explain the overall strategy with five test scripts. Further information gathering revealed more complexity and the number of test scripts rose to 10.

A complicating factor was that each combination of MassChrom software had been validated on its own, comparison of data across all combinations of the software had not been undertaken as this is not normally considered as part of a normal validation study. Therefore, to ensure a comprehensive approach to the data migration, an evaluation of data acquired by all MassChrom software versions was required to ensure that no regulatory questions remained with the data migration. This approach increased the number of test scripts to 16.

Detailed design of the test scripts enabled a better way of testing to be developed and this reduced the number of test scripts down to 12, of which three were for retirement of the obsolete mass spectrometry systems.

Design of the Overall Data Migration and System Retirement

As there was no systematic study of results from all MassChrom software combinations, it was decided to evaluate results from all MassChrom software combinations versus Analyst. In addition, all future data acquisition and analysis configurations were also evaluated to give a comprehensive approach to the data migration and find out if there were any problems with the proposed approach.

Standardised Study Design: as the Analyst version 1.0 had been comprehensively validated to include some 21 CFR 11 requirements [31], we decided that this was the standard to which all data migrations would be measured. A series of 32 sample vials were prepared containing standard and blank solutions that represented a standard curve and a series of unknown samples. This standard set of samples was injected into a mass spectrometer controlled by Analyst software and this set of acquired data was considered the gold standard against which all data migration results were measured.

The standard sample set was then injected into different mass spectrometers controlled by the different software versions, the data analysed and then migrated into the Analyst using the supplier's utilities and reprocessed. Therefore we have a situation where the same samples solutions have been acquired and analysed by the various MassChrom software versions and then migrated into the Analyst and reprocessed and compared against the results of the same samples acquired and processed directly by the Analyst.

In addition, historic study data acquired under MassChrom and archived on tape would be restored to the server, all electronic records then migrated to Analyst, and the results compared.

All test scripts were written, technically reviewed and then approved by the Quality Assurance Unit before execution.

Data Migration: Key Results

In this section, we present a selective review of the key results obtained from the data migration to illustrate the issues in a data migration project. Four areas will be discussed in light of the migration issues we found, the acceptance criteria that we set, and the results that were obtained after the migration.

Retention Time

Retention time is a fundamental chromatographic parameter and is the time that the chromatographic column retains an analyte. In setting the acceptance criteria, the discussions centred on the conversion of time and we determined that the retention times should be within 1% of the original value, especially as the applications were both from the same software supplier. The acceptance criterion of ±1% was determined on the basis of a 3-minute chromatographic run time and that there likely to be differences in the peak integration algorithm that may impact the peak apex in the migrated data.

Reviewing the migrated data, it was seen that there was a large discrepancy between original and migrated results:

- 1.07 (MassChrom)
- 1.12 (Analyst)

Thus, the migration of this parameter appeared to fail against the acceptance criteria. Examining the data more closely, the data formats between the two are different: minutes and seconds (MassChrom) and digital minutes (Analyst). Therefore, we are not comparing like with like and the MassChrom values must be converted to digital minutes to make the comparison valid.

Therefore, all MassChrom retention time values must be collated, converted to seconds then divided by 60 and before comparison with the corresponding Analyst values. After this conversion, the converted

retention times were similar to the original results within rounding errors in the second decimal place. In retrospect, the acceptance criteria could have been set within ±0.5%.

Instrument Control Parameters

As mentioned earlier in this paper, there are design differences between the two software applications and these are manifested in the instrument control parameters in both that can have no or a major impact on the data migration. This area requires a thorough knowledge of the two applications, failure to do this means that the migration will be flawed due to lack of knowledge.

For example, some parameters are the same in both applications and present no problem in the data migration project. An example of this parameter is the scan type such as MRM (multiple reaction monitoring) that is present in both applications, therefore the migration is relatively straightforward and the acceptance criteria that are set is an exact match.

However, a parameter can have different terms in the two applications but still refer to the same measurement, and this starts to complicate the migration, as the parameters must be mapped. A typical example is the Q0 voltage (MassChrom) that is equivalent to the Entrance Potential (Analyst) and illustrates the design differences between the two applications. The acceptance criteria in this instance were set to the nearest volt ignoring differences in the decimal values (e.g. 3.0 versus 3.00), the rationale was that we did not know how numbers were held in either system and that there might be rounding errors involved in the migration.

Adding further complexity to the migration is where a parameter in Analyst has to be derived from two parameters in MassChrom. Thus, the collision cell exit potential value in the Analyst can only be calculated by subtracting the potential for the Rod Offset Potential Q2 from the Inter Quad Lens 3 potential. The acceptance criteria for this were the same as the last example (the nearest volt ignoring differences in decimal values).

Again, this reiterates the need to fully understand the two applications before beginning a data migration. The acceptance criteria for all the instrument parameters monitored in the migration were documented in the appropriate test scripts that were reviewed and approved before the migration.

Integration Algorithms and Calculated Results

When migrating data from one application to another there are a number of results that can be compared. In the example of mass spectrometry these include:

- Analyte peak heights or areas
- Drug: internal standard ratios
- Calibration curve parameters
- Calculated results from unknown samples
- Back calculated standards

As the integration algorithms were different between the two applications, an early decision in the migration was to avoid using the peak area calculations as a comparator between the two systems as noted by McDowall [34]:

> "*What we need to consider here is, when the data files are in the new data system are similar results . . . obtained? Expect to see some differences between the two systems. The main issue is whether it matters from a scientific perspective. . . For instance, if the final calculated result means that a sample that was previously acceptable is now out of specification, the impact of this needs to be assessed . . .*".

This situation was confirmed from the first set of converted data shown in Table 6.

Analyte Standard Concentration	MassChrom Peak Area	Analyst Peak Area
10 ng/ml	4366	4544
20 ng/ml	7851	8383
50 ng/ml	22867	23160
100 ng/ml	45204	47667
500 ng/ml	205054	205822
1000 ng/ml	399296	401330

Table 6: Comparison peak areas from MassChrom with the same data converted and calculated by Analyst

Note that the data at first glance are very comparable, however on closer inspection the Analyst data were consistently higher. Upon further investigation into the issue, it was discovered that the electronic records were migrated without the original baselines set in the Macintosh environment. However, if the migrated data are auto-processed (baselines were automatically placed using pre-set criteria) using manually input data from the original MassChrom methods, then similar analyte results are obtained.

The major issue from a therefore when quantifying data we are unable to comply with the full requirements of 21 CFR 11. However, there was no need to re-develop any method as similar results were obtained and being consistent with the comments of McDowall [34].

Calibration Parameter	MassChrom	Analyst
Slope	0.00365	0.00362
Intercept	0.00127	-0.00036
Regression Co-efficient	0.99726	0.9960

Table 7: Calibration curve parameters calculated by MassChrom and Analyst

Calibration curve parameters for original and converted data are shown in Table 7; the values are equivalent. However, the criteria chosen for acceptance of the data migration were based on the calculated results. As the analysis is based upon a comparative method of analysis (chromatography), the results were deemed the best way of evaluating if the conversion was successful. The key question is would the same decision be taken on the data? Therefore, a regression line of the MassChrom versus the Analyst across all concentrations should have a correlation co-efficient close to 1.0 if the results were the same by both methods. These data are shown in Figure 10.

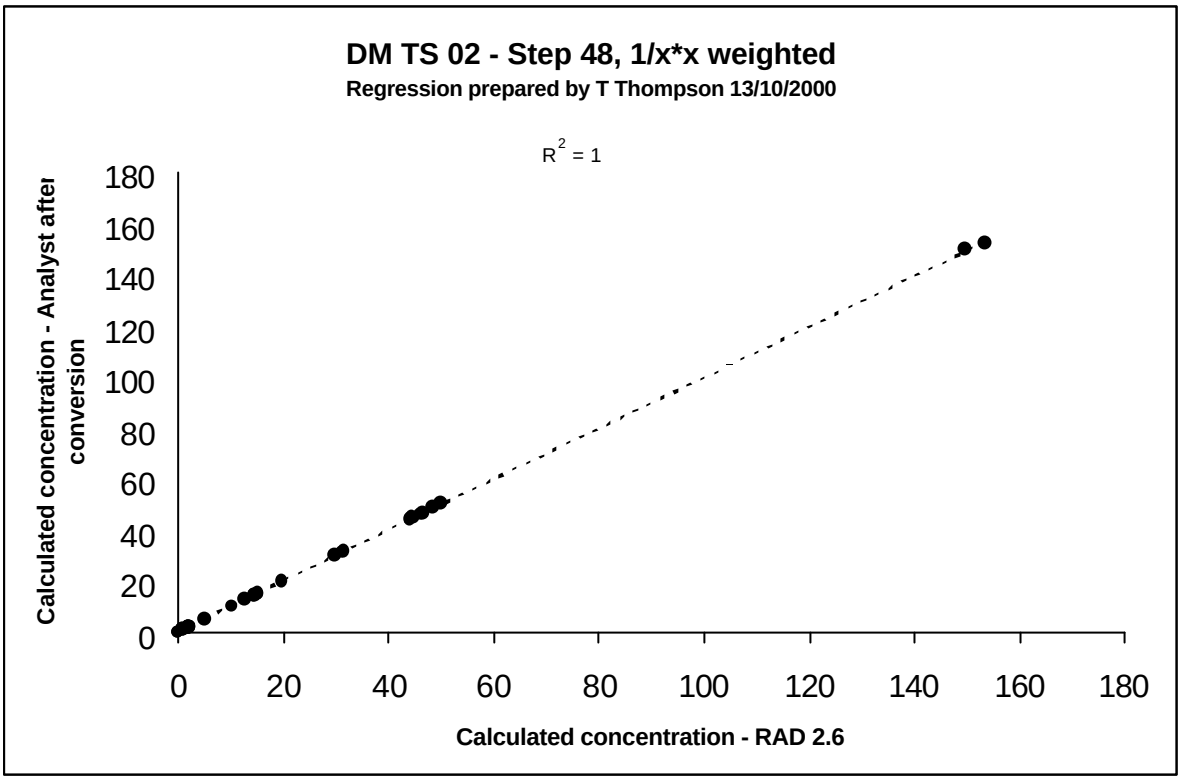

Figure 10: Regression analysis of Macintosh and Analyst data showing equivalent results obtained from the data migration

History Logs

MassChrom does not have an audit trail associated with the data but it does have a history log associated with each data file that notes data and time of creation and changes made to the data. The entries created in the Macintosh environment were migrated to the Analyst environment exactly and were updated following change of a baseline or similar events.

Data Migration from Archive

The final segment of the data migration was to take an archived study, restore the data into the Macintosh environment, reprocess them and then migrate them into Analyst environment for further processing. The two sets of calculated results were compared as above and the results were equivalent.

Data Migration Summary

Data migration from one platform and environment to another was accomplished using the utilities supplied by the supplier. For most cases the tools were successful, however the inability to migrate the previously fitted baselines is a major flaw that prevents the ready replay of data. However, if data are auto processed, then equivalent results are obtained. A key for success is the technical understanding of both environments so that parameters can be mapped between the two.

CDS System Retirement

Under the data migration and system retirement test plan outlined above, three test scripts were written for the formal retirement of the obsolete mass spectrometry systems. As these systems were essentially the same configuration, the test scripts were identical and just varied with the name and identification of an individual system. The process flow is shown in Figure 11, the involvement of management support in the process is key.

The essence of each retirement test script was a pro-forma checklist for the systematic collection and confirmation of activities involved in retirement of an instrument. Sections within each test script for the retirement of a system included:

- Component inventory: all components of the system including the computer, network connections, software and MS instruments are listed in the test script (this is supplied from the system inventory and information gathering stages of the process outlined in Figure 9)
- Data: It was confirmed that all data have been backed up and then copied across to a server and have not been corrupted. This is followed by deletion of the data on the hard drive.
- Computer: disconnection of the computer from the network and informing the IT department that the socket (IP address) can be reallocated if required. The hard drive of the Macintosh was reformatted before the computer was removed from site to ensure that no confidential data remained.
- Mass Spectrometer: There were several stages to this where it was confirmed that the instrument was biologically and radiologically decontaminated before allowing it to be removed from the site.
- Finance: the fixed asset numbers and identities of the components retired were passed to the Finance Department to update the asset register and show the item as decommissioned.

Each section in the retirement test script has the expected results and documented evidence expected as well as acceptance criteria; the script was completed by management review of the overall retirement.

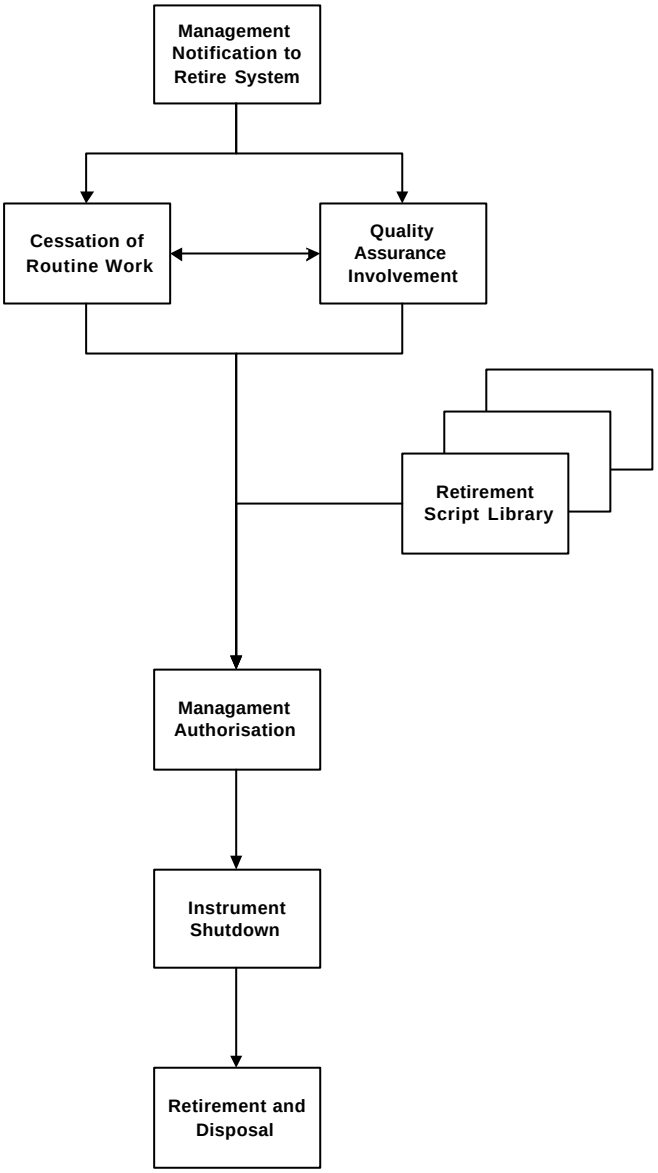

Figure 11: Process flow for system retirement

Data Migration and Retirement Summary

When considering a data migration and system retirement project the following approaches are suggested:

- Think first and understand the complexity of the whole system and technical problems associated with it. This is important and whilst it will slow the overall project initially, will enable the actual work to proceed more smoothly than would be the case if this step were omitted.
- You will be unlikely to solve the problem at the first attempt, therefore adopt an evolutionary approach to the issues. This is illustrated in this paper where the number of scripts rose from 1 to a final 12.
- Do no rush into actions, therefore draw up a data migration plan and then do nothing for at least a week to enable you to review the plan critically and refine the approach: is it feasible and what is the regulatory risk?
- Be practical and flexible, as you will find unexpected issues when least expecting them. The better prepared you are the less likely these issues will be major and affect the data migration adversely.

- Large volumes of data will be produced when validating the data migration process, plan well in advance how to capture and handle these data. These data will be both paper and electronic files, manage both well and have file-naming conventions.
- Education of the software supplier, if this is a commercial system, may need to be factored into the migration.

REFERENCES

1.	R.D.McDowall, Laboratory Data Systems, Chapter 13 in Analytical Chemistry in a GMP Environment, Editors J.M.Miller and J.B.Crowther, Wiley Interscience, New York, 2000
2.	R.D.McDowall, LC-GC Europe 12 (1999) 568-578
3.	N.Dyson, Chromatographic Integration Methods, RSC Chromatography Monographs Series Editor R.M.Smith, Second Edition, Royal Society of Chemistry, Cambridge, 1998.
4.	C.Burgess, D.G.Jones and R.D.McDowall, LC-GC International, 10 (1997)
5.	Validation of Computer-Related Systems, Parental Drug Association, Technical Report Number 18, Journal of the PDA, 49, No 1 January/February 1995, Supplement S1-S17
6.	Gaines Chemical Company, FDA Inspection 483 Observations, December 1999
7.	Glenwood LLC, FDA Warning Letter, May 1999 (m2633n)
8.	Genesia Scicor, FDA Warning Letter, July 1999 (m2819n)
9.	Noramco Inc., FDA Inspection 483 Observations, May 2001
10.	C.Kornbo and R.D.McDowall, Scientific Computing and Instrumentation, January 2002
11.	Good Automated Manufacturing Practice Guidelines Version 4, International Society for Pharmaceutical Engineering, Tampa, Florida, December 2001
12.	FDA Draft Guidance for Industry, Electronic Records; Electronic Signatures, Validation, September 2001
13.	IEEE Guide for Developing Software Requirements Specifications, Standard 1233-1998
14.	R.D.McDowall, Scientific Data Management, March 1998, pp 7-12
15.	Good Manufacturing Practice for Medicinal Products in the European Community, Annex 11, Commission of the European Communities, Brussels, 1997
16.	R.D.McDowall, Scientific Data Management, June 1998 pp 8-14
17.	R.D.McDowall, Scientific Data Management, September 1998 pp 7-11
18.	FDA Guidance for Industry, General Principles of Software Validation, CDRH, January 2002
19.	Current Good Manufacturing Practice Regulations for Finished Pharmaceuticals (21 CFR 211) revisions as of 1st April 2002
20.	Spolana, FDA Warning Letter, October 2000
21.	IEEE Standard 829-1983, Software Test Documentation, Institute of Electronic and Electrical Engineers (1983).
22.	M.Fewster and D.Graham, Software Test Automation, Effective Use of Test Execution Tools, Addison Wesley, London, 1999
23.	B.Wikensted, P.Johansson and R.D.McDowall, LC-GC International 11 (1999) 88-96
24.	D.Browne, T.Thompson, D.Mole and R.D.McDowall, LC-GC International, 14 (2001) 687-694
25.	H.Hambloch, Good Computer Validation Practices; Common Sense Implementation, T.Stokes, R.C.Branning, K.G.Chapman, H.Hambloch and A.J.Trill, Interpharm Press, Buffalo Grove, IL, pp 113 - 140, 1994
26.	B.Boehm, Some Information Processing Implications of Air Force Missions 1970-1980, The Rand Corporation, Santa Monica, CA, 1970.
27.	K.G.Chapman, Good Computer Validation Practices; Common Sense Implementation, T.Stokes, R.C.Branning, K.G.Chapman, H.Hambloch and A.J.Trill, Interpharm Press, Buffalo Grove, IL, pp 47-74 1994
28.	Organisation for Economic Co-operation and Development, Consensus Document on Principles of Good Laboratory Practice applied to Computerised Systems, Paris 1995
29.	P.D.Lepore, Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management, 17 (1992) 283 - 286
30.	FDA Draft Guidance for Industry, Electronic Records; Electronic Signatures, Maintenance of Electronic Records, September 2002
31.	21 CFR 11, Electronic Records, Electronic Signatures Final Rule, Federal Register 62 (1997) 13430 - 13466
32.	D.Browne, T.Thompson, D.Mole and R.D.McDowall, Journal of Validation Technology 8 (2002) 250-259
33.	R.D.McDowall, LC-CG Europe 12 (1999) 774-781
34.	R.D.McDowall, LC-CG Europe 13 (2000) 35-38