Experiences Migrating Mass Spectrometry Data Between Platforms and Applications and Retiring Obsolete Legacy Systems:

A Case Stud y

By David Browne
Covance Laboratories
Terry Thompson
Covance Laboratories
David Mole
McDowall Consulting
and R.D. McDowall
McDowall Consulting

This article discusses the overall strategy and experiences observed during mass spectrometry migration of electronic records, generated by commercial applications, from a Macintosh platform to one based on Windows NT. Differences between the two environments included:

- Operating systems and hardware platforms
- Incompatible file formats
- Operating parameters are named differently due to variation in design philosophy in Mac and NT applications

A generic methodology for migrating electronic records will be discussed with experiences from this case study. Issues experienced while completing the work included:

- Advantages and disadvantages of the overall migration strategy
- Defining the electronic records to migrate so that the data can be reanalyzed in the new system
- The need to understand the complexity of the migration problems and how this impacts the validation of the process
- The need to map operating parameters between the two environments to obtain equivalent values for validation of the data migration

"Data migration can be the worst part of validating an existing computerized system because the system may have been operational for a number of years. "

- Conversion of Macintosh electronic records to NT format, and recalculation of data such as peak areas, calibration curve parameters, and calculated concentrations
- The authors' experiences in setting acceptance criteria
- Limitations of the data migration tools provided by the vendor

Overview

Data migration and system retirement occur at the end of the lifecycle of any computerized system, however, there is little or no direct regulatory requirements for formal system retirement, nor general advice on how to undertake the task. Retirement in many instances may be a euphemism for simply throwing the system components out, however, this paper intends to present justification for a more formal approach.

Data migration is necessary for a number of reasons such as a change in the:

- Data processing algorithms following a software upgrade of an application
- Use of a different software application
- Computing environment such as an operating system or computing platform
- Data file formats

Data migration may also be required for the duration of the records retention period under the electronic records and electronic signatures Final Rule (21 CFR 11).[1] 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 validating an existing computerized system because the system may have been operational for a number of years. The data may have been shared between several departments, and the original staff of the project may no longer work for the company or in the same area. This challenge can be compounded by reorganizations within a firm, and the system boundaries are different compared with the original installation. Fortunately in this case study, the data were generated within a single department with a single system owner, making the project simpler than other comparable data migration projects.

This paper describes our experiences designing and validating a mass spectrometry data migration between two different platforms. The initial event that caused the company to migrate data was when the vendor of the mass spectrometry equipment and application software moved to a new computing platform and declared the current one obsolete.

Design of the Overall Mass Spectrometry Validation Project

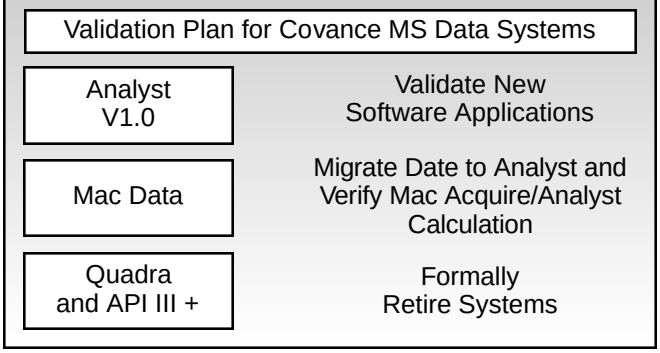
Computer system validation is concerned with producing documented evidence that the system in question produced using quality standards, accurate when qualified, and remains so throughout its operational life. The need for fully validated chromatography data systems used for Good Manufacturing Practice (GMP) and Good Laboratory Practice (GLP) submissions has been well documented over the past few years. As the pharmaceutical industry adopts more computerized systems, and attempts to comply with a reasonable interpretation of FDA21 CFR Part 11, full validation of the electronic record and signature program will be imperative.

This project was conducted using the lifecycle approach to validation of Chromatography Data Systems (CDS) as described by McDowall[2-4], and consisted of three processes under a single validation plan as shown in *Figure 1*. These processes were:

❶ Prospective validation of the new application software (Analyst® version 1.0) and qualification of new instruments associated with them

Figure 1

Overview of the Whole Mass Spectrometry Validation, Data Migration, and System Retirement Project



❷ 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
❸ Formal retirement of obsolete mass spectrometry and Macintosh computer hardware

This article concentrates on processes two and three, as the prospective validation of the analyst has been published separately.

Overview of the Mass Spectrometry Systems Used in the Case Study

The mass spectrometry equipment, current software options, and computing environment used is presented below and summarized in *Figure 2*.

Mass Spectrometry Equipment

There were three main models of mass spectrometer currently operating, API® models III+, 365, and 3000. Of these, the API III+ was obsolete because the Macintosh PC used to run the software was no longer in production. Therefore, the three systems using the API III+ mass spectrometer were to be formally retired, and only the API 365 and 3000 models would be used thereafter.

Data Acquisition and Processing Software Applications

The MassChrom mass spectrometer software currently used was 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 was to be retired under the work described in this paper.

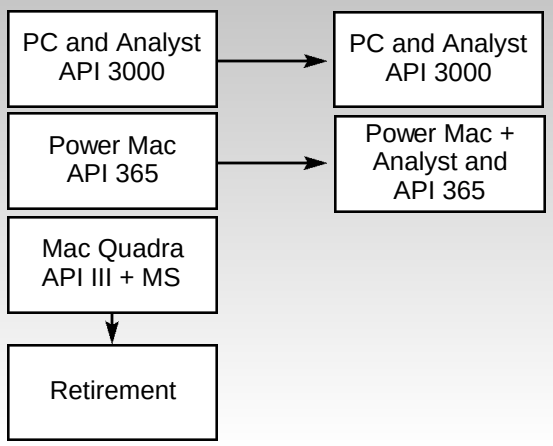
Amixed environment was operated during a transition period where data were acquired by sample control on a Macintosh, but all data processing and quantification ran on the Analyst. After retiring all of the Macintosh computers, there was to be only Analyst running on Windows NT.

Computing Environments

The existing environment was Macintosh with mass spectrometry being downloaded to a server after it had been acquired. Introduction of the Analyst started a migration to an NT operating environment that continued after the completion of the data migration outlined here in *Figure 2*.

Figure 2

Overview of Mass Spectrometry Equipment and Data Systems



Differences Between the Two Systems

It is vitally important to understand the differences between the two environments before progressing further with any data migration. Here were the major differences between the two systems and their impact on the data migration. Generally, the problem was that we had incompatible:

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

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

Computing Platform Differences

The Macintosh and Intel hardware computing platform and operating system software were 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 could run the software and be supported by the vendor.

Raw Data File Format Differences

The file formats for the chromatograms produced by the same instrument in the two environments were completely different. The Macintosh used a different file format compared with the Analyst that used Waveform Interchange File Format (WIFF), and could have had 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 required three files to set up and acquire data: the method, state, and experiment files. The method and experiment files were 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 were just two such files used within the Analyst: The Data Acquisition Method (DAM) and instrument (INS) files. The mapping of the MassChrom and Analyst files was not one to one: parameters in the experiment file were 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 were 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 specifically designed for trained mass spectrometrists.

Over time, the instrument was used more by chromatographers. The Analyst software is an application that is simpler, and uses chromatographic terms more than mass spectrometry terms. This difference in design philosophy complicates data migration because terms had to be mapped between the applications, as described later in this article.

Generic Data Migration and System Retirement Process

A generic seven-step process, shown in *Figure 3*, describes system retirement and migration of data. Each stage will be described as an overview, and is a summary of the work described by McDowall.[4]

❶ 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.

❷ 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.

❸ Write the Retirement Plan

Using the data generated from Step one, 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

❹ Detailed Information Gathering

In this part of the process you need to know the details of the computer hardware, including any specialized devices, software, and 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?

❺ System Decommissioning and Data Migration Plan

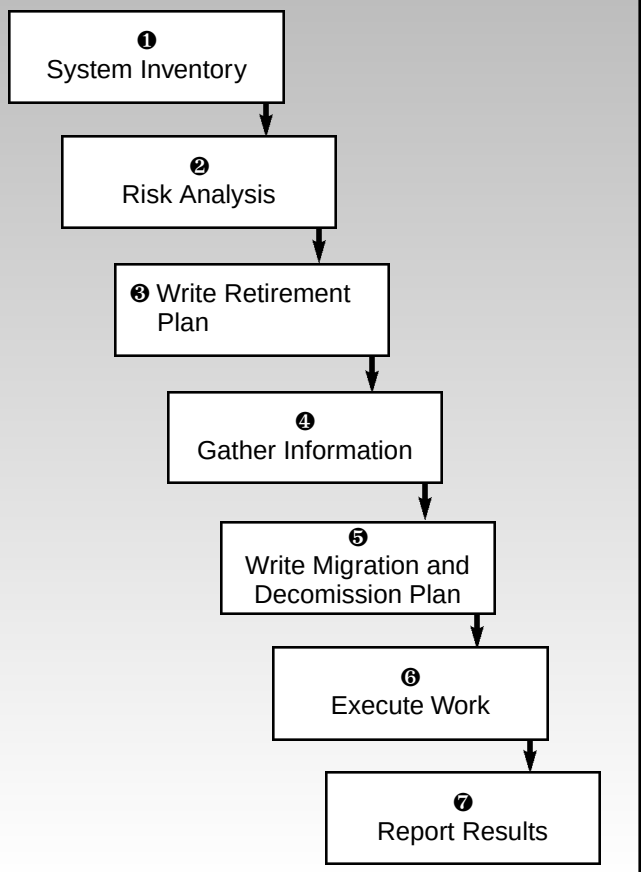
This document is a detailed presentation of the approach you'll take towards the system, and describes the roles and responsibilities of people involved in the project, systems, data to migrate, test scripts needed, and what each test script will contain to document the process.

❻ Execute Work and Document Activities

Following the tasks described in the decommissioning plan, the data retirement will start first, followed by the system retirement. You will need to

Figure 3

Generic Process for Data Migration and System Requirement



write any scripts to check and document the correctness of the data transfer. This is a critical stage in generating confidence in the process. Once the data has been successfully migrated and or archived, then hardware can be turned off, and the decision of reusing it or removing it from the site can be addressed. Again, this will be documented as the process continues.

❼ Write Retirement and Migration Report

This is simply a summary of the work that was done 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 on the report.

Rationale for the Data Migration

The driver for the migration in this case was the fact that the vendor was rendering the current platform obsolete. No further development occurred, and the users were being encouraged to migrate to a new platform based on Windows NT. As the system has generated large amounts of regulatory data, the case study company decided to ensure that data could be migrated and reprocessed in the Analyst environment if required, either by sponsor companies or regulatory authorities.

Data Migration Strategy

The options for data migration were to assess if it was technically feasible to migrate data. The vendor of the mass spectrometry software systems provided conversion programs that enabled a user to migrate electronic records from the Macintosh to the Analyst system. Conversion was necessary because the file formats were completely different between the Macintosh and NT environments.

Vendor Supplied Data Conversion Utilities

Three file converter programs were supplied for the conversion of the Macintosh format data and meta data files. These included:

- File Translator: Data file conversion program that took Macintosh formatted data files and converted them to single or multiple Analyst format files (WIFF).
- InstFileGenerator: Instrument file conversion program combined Macintosh state and calibration files and generated an Analyst instrument file (INS file).
- ExptFile Converter: Experiment file conversion program combined a Macintosh state file and a Macintosh experiment file, and generated an Analyst Data Acquisition Method (DAM) file.

It was then technically feasible to convert the data and migrate them into the NT environment. The question then became, "Are all data converted or are files converted on an "as needed" basis? The data volume involved was in the range of 100-200 GB of data.

Limitation of the Data Conversion Utilities

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

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

Data Migration Options

There were 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 arose from this approach:

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

Evolution of the Data Migration Design

Adata migration project requires a full understanding of the challenges involved. Therefore, this section of the paper intends to describe the evolution of the project and the extent of the issues as they arose.

Initially, a single test script under the Analyst validation was envisioned. However, as the complexity of the MassChrom software versions was realized, 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.

Acomplicating 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 performed, as this was not considered a 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. This approach increased the number of test scripts to 16.

Detailed design of the test scripts produced a better method of testing, which reduced the number of test scripts down to 12. Three of the remaining test scripts were designated for retirement of the obsolete mass spectrometry systems.

Design of the Overall Data Migration and System Retirement

Since 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 determine if there were any problems with the proposed approach.

Standardized Study Design

The Analyst version 1.0 had been comprehensively validated to include some 21 CFR Part 11 requirements,[5] so it was 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 were analyzed, and then migrated into the Analyst using the vendor's utilities and reprocessed. Therefore, we had a situation where the same sample solutions were acquired and analyzed by the various MassChrom software versions migrated into the Analyst, 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 were 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 that represents the time the chromatographic column retains an analyte. In setting the acceptance criteria, the discussions centered on the conversion of time, and we determined that the retention times should be within one percent 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 three-minute chromatographic run time, taking in consideration that there were 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 observed 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 did not compare the same measuring parameters, and the MassChrom values must be converted to digital minutes to make the comparison valid.

Therefore, all MassChrom retention time values were collated, converted to seconds, then divided by 60, before comparing them to the corresponding Analyst values. After this conversion, the converted retention times were similar to the original results after disregarding the 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, there were design differences between the two software applications, and these was manifested in the instrument control parameters in both that may or may not have a major impact on the data migration. This area required a thorough knowledge of the two applications. Failure to do this caused the migration to be flawed.

For example, some parameters were the same in both applications and presented no problem in the data migration project. An example of this was the scan type, such as Multiple Reaction Monitoring (MRM) that was present in both applications, therefore the migration is relatively straightforward, and the set acceptance criteria is an exact match.

However, a parameter can have different terms in the two applications, but still refer to the same measurement, and this started to complicate the migration, as the parameters were analyzed to determine which application reported which measurement. A typical example is the Q0 voltage (MassChrom) that was 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 based upon the fact that we did not know how numbers were held in each system, and by applying a less finite acceptance criteria, rounding errors would be reduced during migration.

Adding further complexity to the migration was when a parameter in Analyst had to be derived from two parameters in MassChrom. Thus, the collision cell exit potential value in the Analyst was 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 included:

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

Since the integration algorithms were different between the two applications, an early decision was made to avoid using the peak area calculations as a comparison factor between the two systems, as noted by both McDowall[4] and Huber and Winter.[5] As noted by McDowall,[4] "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 *Figure 4*.

Figure 4

Comparison Peak Areas

From MassChrom® with the Same Data Converted and Calculated by Analyst®

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

Note that the data at first glance were 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 were auto-processed (baselines were automatically placed using pre-set criteria) using manually input data from the original MassChrom methods, then similar analyte results were obtained.

The major issue was, when quantifying data, we were unable to comply with the full requirements of 21 CFR Part 11. However, there was no need to redevelop any method, as similar results were obtained and being consistent with the comments of McDowall[4] and Huber and Winter.[5]

Calibration curve parameters for original and converted data are shown in *Figure 5*. The values are equivalent. However, the criteria chosen for acceptance of the data migration were based on the calculated results. As the analysis was based upon a comparative method of analysis (chromatography), the results were deemed the best method of evaluating that the conversion was successful. The key question was, 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 coefficient close to 1.0 if the results were the same by both methods. These data are shown in *Figure 6*.

Figure 5

Calibration Curve Parameters

Calculated by MassChrom® and Analyst®

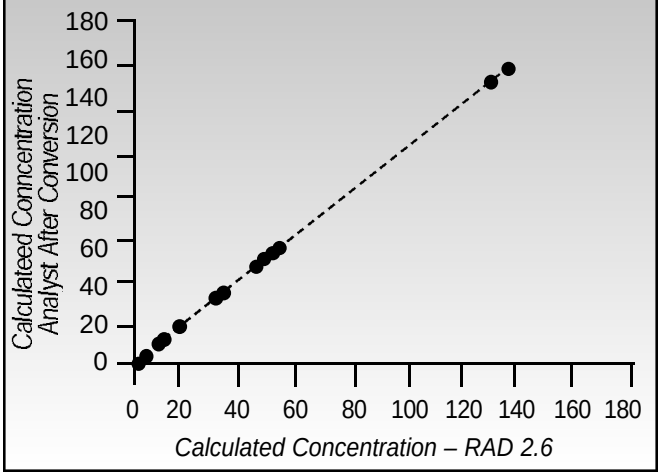
Calibration Parameter	MassChrom	Analyst
Slope	0.00365	0.00362
Intercept	0.00127	-0.00036
Regression Coefficient	0.99726	0.9960

History Logs

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

Figure 6

DM TS 02 – Step 48, 1/x*x Weighted



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 the 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 vendor. For most cases, the tools were successful, however, the inability to migrate the previously fitted baselines is a major flaw that prevented the ready replay of data. However, if data were auto processed, then equivalent results were obtained. A key for success was the technical understanding of both environments, so that parameters can be mapped between the two.

Mass Spectrometry System Retirement

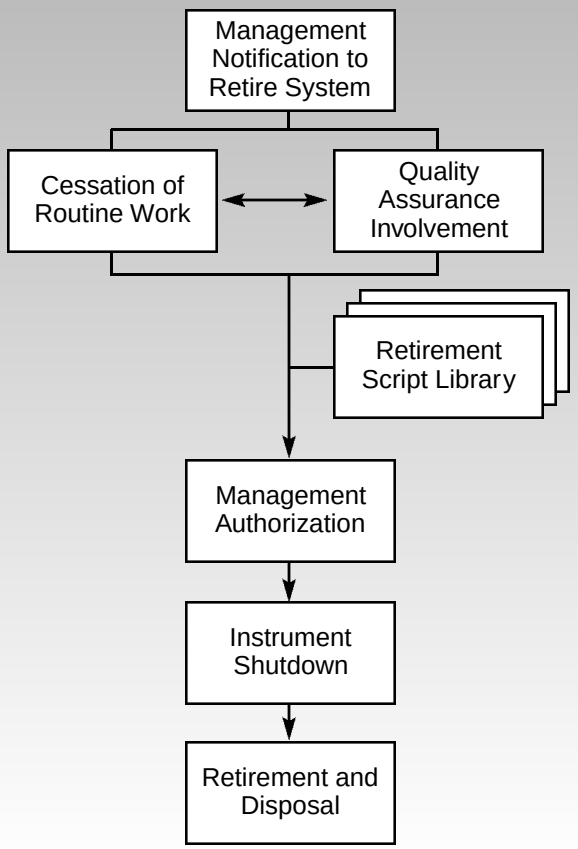
Under the data migration and system retirement test plan, three test scripts were written for the formal retirement of the obsolete mass spectrometry systems. As these systems were essentially the same configuration, the test scripts were identical, and only varied with the name and identification of an individual system. The process flow is shown in *Figure 7*. The involvement of management support in the process is critical.

The essence of each retirement test script was a proforma 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 the following:

- Component inventory: all components of the system including the computer, network connections, software, and MS instruments were listed in the test script (this is supplied from the system inventory and information gathering stages of the process outlined in *Figure 3*).
- Data: It was confirmed that all data were backed-up and then copied across to a server and have not been corrupted. This was followed by deletion of the data on the hard drive.
- Computer: disconnection of the computer from the network and informing the ITdepartment that the socket (IPaddress) could be reallocated if required. The hard drive of the Macintosh was reformatted before the computer was removed from the site to ensure that no confidential data remained.

Figure 7

Process Flow for System Retirement



- 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 contained the expected results and documented evidence that it conformed to acceptance criteria. The retirement test script was subject to management review, and the overall retirement was approved.

Summary

When considering data migration and system retirement, 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 while it will slow the overall project initially, it 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 on the first attempt, therefore adopt an evolutionary approach to the issues. This is illustrated in this paper where the number of scripts rose from one to a final 12.
- Do not rush into actions. First, draw up a data migration plan and review the plan critically and refine the approach. Then the question must be asked, "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 on how to capture and handle these data. These data will be both in paper and electronic files. Manage both well, and establish standard file-naming practices.
- Education of the software supplier; If this is a commercial system, it may need to be factored into the migration. ❑

About the Authors

Dave Browne is currently the Manager of the Mass Spectrometry Laboratory in the Department of Bioanalytical Services at Covance Laboratories, Harrogate, UK. Dave has worked for Covance for over 12 years, and has extensive experience in LC-MS analysis of samples to support pharmacokinetic studies.

Dave Mole is currently a Senior Consultant for McDowall Consulting, Bromley, UK. He has 30 years experience working in the pharmaceutical industry with Wellcome Research covering a wide variety of roles, including automation project, computer validation, data migration, and system retirement.

Bob McDowall is the Principal of McDowall Consulting with nine years experience in this position. Prior to that, he held positions in the pharmaceutical industry for 15 years. Bob has published and presented widely on a number of topics, including computerized system validation. He can be reached by phone at 44-20-8313-0934.

Terry Thompson is the 21 CFR Part 11 Site Coordinator and the Systems Validation Coordinator for Covance Laboratories, Harrogate UK. Terry has a background of clinical chemistry, working for the Medical Research Council before joining Covance.

References

1. FDA. 21 CFR 11, Electronic Records, Electronic Signatures Final Rule. *Federal Register*. 62. 1997. pp. 13430-13466.
2. McDowall, R.D., "Chromatography Data Systems III: Prospective Validation of a CDS," LC-CG Europe Vol. 12 No. 9. 1999. pp. 568-576.
3. McDowall, R.D., "Chromatography Data Systems IV: Managing Change in a Changing World," LC-CG Europe Vol. 12 No. 12. 1999. pp. 774-781.
4. McDowall, R.D., "Chromatography Data Systems V: Data Migration and System Retirement," LC-CG Europe Vol. 13 No. 1. 2000. pp. 35-38.
5. Huber, L. and Winter, W., "Implementing 21 CFR Part 11 – Electronic Signatures and Records in Analytical Laboratories: Part 4, Data Migration and Long Term Archiving for Ready Retrieval." *BioPharm.* June 2000.

Article Acronym Listing

CDS:	Chromatography Data Systems
DAM:	Data Acquisition Method
GLP:	Good Laboratory
GMP:	Good Manufacturing Practice
INS:	Instrument Files
MRM:	Multiple Reaction Monitoring
WIFF:	Waveform Interchange File Format