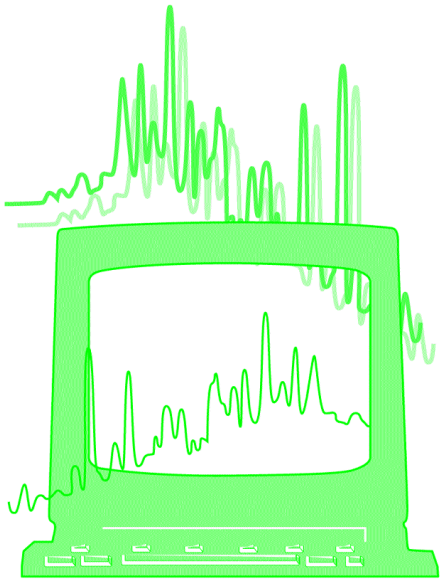

Chromatography Data Systems V: Data Migration and System Retirement

R.D. McDowall, McDowall Consulting, Bromley, Kent, UK.

At the end of the life cycle of a chromatography data system do you turn off the system and forget it? Alternatively, do you need to consider data migration and retirement of the existing system?

There is always something that satisfies an animal urge in chromatographers when an old chromatography data system (CDS) is to be retired. The emotions can vary from a sigh of relief ('In the Name of God, Go' syndrome) to saying goodbye to an old friend.

The longer serving members of the department may remember the work to acquire (1), validate and qualify (2) and operate the system whilst maintaining its validated status (3).

However, you'll not be going back to the 'Jurassic Age of Chromatography' will you? There'll be a replacement system that will have been properly specified, selected, installed and qualified. What you have is a situation in which you are at the end of one life cycle and at the beginning of another (Figure 1).

The issues to face are these:

- What to do about your existing system and the accumulated data it has generated and processed, especially if there are electronic data files?
- How do you cut over to the new system and what is the impact on the work being undertaken by the laboratory?

I'll not be discussing data archive in this article, only looking at migration to a new system and the problems that you could have.

The discussions in this article will be general and not specific, as usually the decisions around retirement and migration are made on a case-by-case basis and not to a generic formula. The reason for this is that each laboratory and its organization tends to have individual requirements in this area, some internal and some external. Moreover, as technological advances occur, the existing platform may need to be replaced when upgrading to the next version of a vendor's system, as has occurred with some CDS applications already.

Options for Retirement and Data Migration

There are several approaches to the retirement of a system that could be considered:

- Turn off and forget: as the name suggests this is turn off the old system and forget it, then start using the new CDS.
- Phased cut-over: complete the existing work on the old system and undertake all new work on the new system.
- Data migration and retirement of the old system components.

Why bother to go through all this fuss over an old system, I hear you say? Let's look at some of the reasons for and against the argument for formal data migration and system retirement.

Consider an overview of your existing system. Most CDS will have been operational in your laboratory for between 2 and 10 years. Either stand-alone or multi-use systems, they will have analysed many thousands, hundreds of thousands or even millions of samples. Some work, especially that assessing the stability of products may have run for a number of years. Cut-over to a new system will inevitably involve studies assayed under the two systems.

Drivers for System Retirement

The drivers for system retirement are usually from two sources: internal and external. We'll look at both to examine the reasons for this.

Internal drivers: Here the replacement of a system may result for several reasons:

- User input: Does the current system(s) do the job required? The business function may either drift or change dramatically over time and the CDS may not meet the current requirements. Alternatively, an increase in functionality is required as the users become more

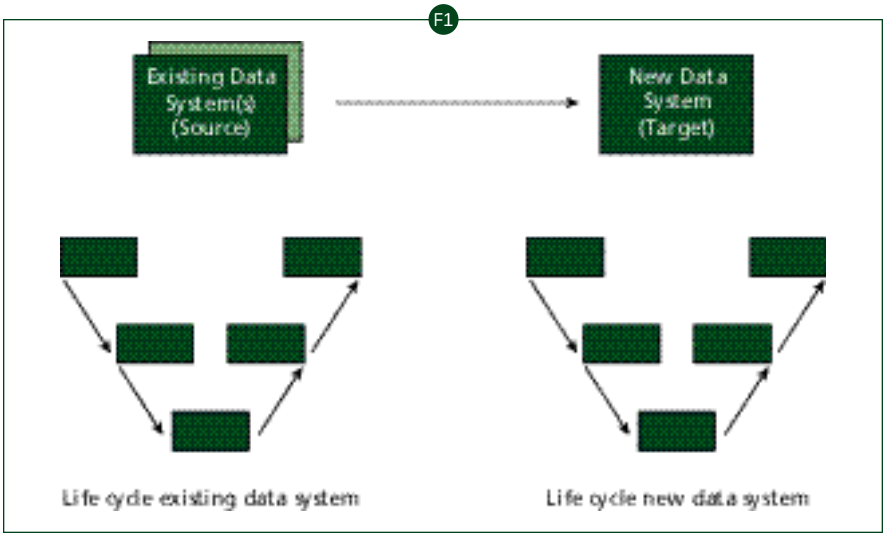

Figure 1: Commission a new system and retire the old one.

sophisticated and need more from the system to work effectively.
- Several data systems reduced to one: Over time a laboratory may acquire a number of data systems, either because of changes in purchasing policy or because of personal preference. This means that a single chromatographer would need to be trained on several different data systems to be effective within such an environment; this would be coupled with the different capabilities of each type of data system. Retirement of all but one, or indeed a move to another data system, would mean that a chromatographer could be trained for a single system.
- Corporate policy: This is when an organization has made a decision to use a single CDS for either financial and/or standardization reasons. Alternatively, the introduction of a standard or common office environment (SOE or COE) for the PC desktop could trigger the decision for a change to comply with a COE.
- Business decision: The contraction or closure of a department or even a site may necessitate the retirement of a system and also data migration to another department or site.
- Computerized system validation (CSV) policy: Some organizations include a retirement and data-migration phase in their CSV policies and this is merely the execution of that policy on retirement.

External drivers: The rationale for system retirement can derive from such factors as
- Existing system obsolete: This occurs when a vendor makes a system obsolete or changes the operating or hardware platform. The old system is declared obsolete and will not be supported after a specific date, and the laboratory has to move to a new system if it wishes to continue to receive effective support from its vendor.
- Interpretation of existing regulations or guidelines: This may result in an action at your laboratory from the regulating or certifying authority that means you must improve your data system. Alternatively, this occurs at a laboratory outside of your organization; wishing to avoid similar action triggers the search for a replacement.
- Introduction of new regulations or guidelines: Here, an existing system does not comply with these new guidelines, and improvements in the system functionality are required if it is to comply with them. An example of this is the electronic records and electronic signatures rule 21 CFR 11 for the pharmaceutical industry (4). Here there are other issues concerning electronic records and raw data definitions that will be discussed in forthcoming "Questions of Quality" columns in more detail.

So, for whatever reason, let's assume that you're going to change your current CDS. What do we do with the old system(s) and the data generated by it (them)?

System Retirement and Data Migration Strategy

The major issue is to decide who has the overall responsibility for system retirement and data migration. If you have a single PC acquiring data from a small number of chromatographs for one department, you may be questioning my sanity. You may be right but that's a different discussion. However, if you are looking at a multi-user client/server system servicing several laboratories, this is a very pertinent question that needs to be resolved before you go much further. For a system of this size, the regulatory or quality impacts could be quite large and the impact if you got it wrong would also be large. This tends to get the attention of senior management rather quickly.

At this point, the size and regulatory impact of the system may require the setting up of a specific project team to manage the retirement and any data migration. Alternatively, it falls to the chromatographer who happens to be "available" at the time. Regardless of the approach, the retirement and migration activities must be coordinated with the efforts to acquire the new CDS to ensure the technical feasibility of any data migration. A single owner of the system to be retired must be identified as the responsible individual. This person could also be the project owner of the new data system to ensure overall coordination.

What Does the System Owner Do?

Let's start from the beginning:
- Identify the user groups (stakeholders) involved.
- Identify who owns the data on the system. This can vary from one individual to several departments, depending on the size of the overall system.
- Identify the completed work to be archived.
- Identify the completed work to be migrated.
- Identify the ongoing work to be migrated.

The system owner needs to involve all stakeholders in the system to establish a retirement and data-migration team.

Seven Steps to Heaven?

Described below is a seven-step process for the retirement and migration of data, this is also outlined in Figure 2.

The issue here is that heroic measures are not required; the key to success is to be realistic in your data-migration aims and

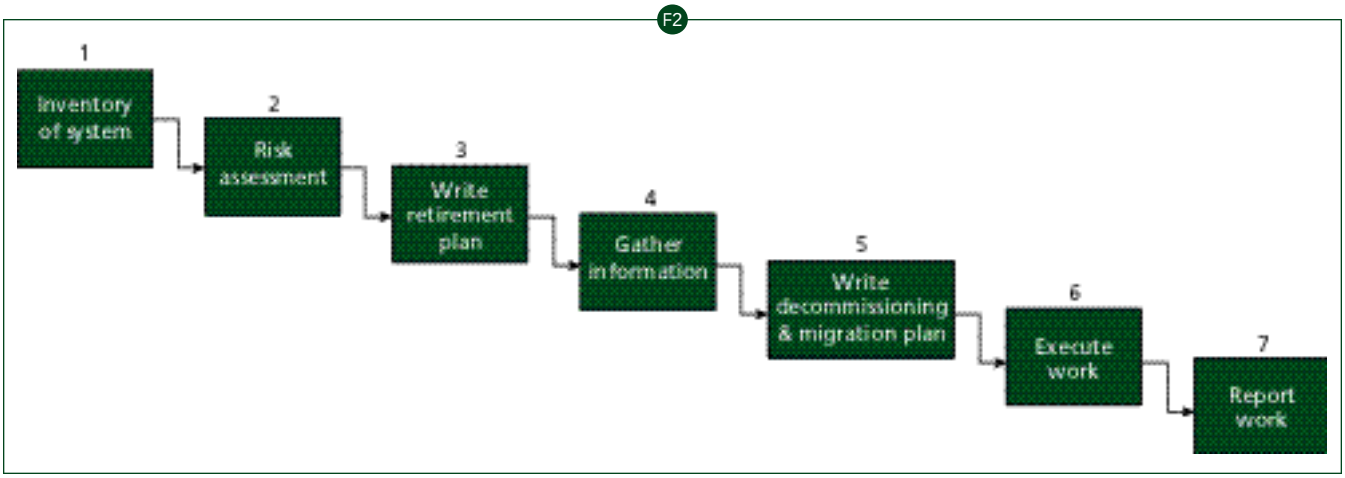

Figure 2: Process for data migration and system retirement.

succeed, rather than be too demanding and fail. This is subject, of course, to meeting your scientific and, where appropriate, regulatory or other quality requirements. So get it right and do it right.

You may think that this process is involved and it is, but the issues concerned are complex. You must plan this carefully. Remember without a plan you can end up anywhere.

Step 1 — inventory of the system(s):

- Identify the scope and boundaries of the system and the departments who use it. Part of this could 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, because data spread over different networks will need to be collated to find out the data volumes and projects/ studies involved.
- Of the work performed on the system, are the data used to support product or sample releases or input to regulatory submissions? Are there data-retention requirements to comply with on system retirement?
- Identify the interfaces to and from the CDS. Are you linked with a laboratory information management system (LIMS) or do you transfer data into spreadsheets?
- Ongoing work needs to be identified: number of studies and progress of each.
- Identify the data volumes.
- Identify the data types: CDS files but also consider whether you need to migrate the method files, run sequence files and the audit trail if the system has one.
- Identify the raw data format used by the existing CDS system.
- Data archive: identify approximate data volumes and the format of the archive.

Step 2 — perform a risk assessment: There are many different ways to perform a risk assessment, and you may have a methodology already in place for this. However, consider some of the following questions:

- Is the CDS system mission critical?
- Are the data generated under regulatory control?
- Are there multiple users?
- Will you need to re-analyse any data?

If you answer 'yes' to one or more of these questions, then in my opinion you have a high-risk system and the formality of system retirement and data migration is mandatory. If the data system is a minor player in a laboratory, then you may want to take a less formal approach. It all comes down to how the system owner and the organization handle risk.

A word of caution at this stage; you'll need to understand the business drivers and be careful of how you interpret any applicable regulations and guidelines. You will also need to understand the skills of the chromatographers and other staff involved in the project.

Step 3 — write the retirement plan: Using the data generated from Step 1, the plan will cover, as a minimum, the

- scope and boundaries of the chromatography data system(s)
- roles and responsibilities
- outline project plan
- process of system retirement
- process of data migration.

One of the issues you'll be facing is whether you have to validate the system on retirement. If the CDS is not validated, what is the risk if you are unable to demonstrate that the system was fit for purpose? Think carefully as this could impact all the data produced throughout the whole of the system's life.

Step 4 — detailed information gathering: In this part of the process you'll need to know the details of the computer hardware including any A/D devices, the software and the documentation associated with the system. If you are migrating to a system from the same vendor, you may be reusing some or all of these components in the new system. Alternatively, if you are closing a facility and moving to another site that uses the same data system, some components may be reused and some retired.

The data must be identified in detail; for example, how many tapes are involved (assuming that your long-term storage is on tape) and what data relating to which samples are on a specific tape. This enables a decision to be taken on the data to be migrated to the new system and that to be archived. See later in this article for more information about data migration and whether it's feasible or not.

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 following:

- roles and responsibilities
- hardware items to decommission or reuse in the new system
- identification of the data to archive and not migrate
- identification of the data to migrate
- plans to migrate any ongoing work
- detailed plans to deal with the migration (e.g., any piloting work to be done before the actual migration, which can be performed 'all at once' or phased over a period of time) (under this section you'll outline the level of validation or checking you'll be doing to confirm that the data have been migrated correctly)
- business continuity planning should any problems arise with the retirement and migration. This stage is critical because you don't want to be left without any data system, but it's not always considered in retirement projects.

This document must be approved by management before any work starts.

Step 6 — execute work and document activities: Following the tasks described in the decommissioning plan, the data retirement will start first and be followed by the system retirement. You'll need to write any scripts to check the data transfers, and check and document the correctness of the data transfer. This is a critical stage in generating confidence in the process.

The process does not always go smoothly or to schedule, so be prepared to be flexible in your approach and document any differences between plans and reality. This is not just generating paper, but if you run into problems then you'll need that documentation to help work out of the hole you are in.

Once the data have been successfully migrated and/or archived, then you'll 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 — retirement report: This is simply a summary of the work 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 Considerations and Issues

We've covered the overall process in the seven steps described above but that is not the full story. We also need to look in more detail at the data-migration issues and problems that can arise. Do not undertake data migration lightly. You'll need to consider the process carefully and evaluate the technical feasibility as well as the scientific impact. After this evaluation as part of the process above, you'll be able to decide whether to proceed or not.

Let's look at some of the data-migration considerations and issues that you'll investigate as part of Step 4:

Technical feasibility (can I migrate my data?): There are several scenarios here and we'll look at just a few. What you

need to consider here is, can I technically migrate my data files from one CDS to another?

- You could have one vendor's data system and want to move to a system from another vendor: is it possible to do this? The problem may be compounded by the fact you are migrating from several systems to a single one. Can you migrate the data from all systems or just a few? Although many CDS vendors adhere to the netCDF format (5), there are small differences that may prevent a full match between systems and the results they produce.
- You may be migrating your data from one vendor's data system running on one platform to the same system running on a different platform. This may entail a change of file format across the two platforms and migration will involve file conversion using a utility provided by the vendor. Happy days are here again…
- You may be thinking if you stay with the same vendor that there will be no problems, but you'd be wrong. A new version of the software could have a new file format (think of the joys that Word 95/97 brought the world) or could involve a modification of the integration algorithms.

Scientific impact (do I get similar results?): What we need to consider here is, when the data files are in the new data system are similar results (notice the avoidance of the words 'same' and 'identical') obtained? This is easily said but not so easy to do in some circumstances.

If you can technically migrate the data files into the new CDS (include also the method information and run sequences if possible as it saves manual retyping) and reintegrate them, what are the results like? Here, we have to consider some of the potential areas for difference in integration between the two systems.

- How is a peak start identified? (will it be in the same place?)
- How is a peak apex defined? (e.g., highest value or an averaged value?)
- How is a peak end defined? (will both systems see the same end point?)
- Are the integration algorithms the same? (probably not, but are peak areas similar?)
- If you see equivalence with large peaks does this apply as well for small peaks?
- Impact of threshold values.
- How is noise calculated?
- How do the two systems handle over-range peaks?

Expect to see some differences between the two systems. The main issue is whether it matters from a scientific perspective. If the new data system reports a peak area of 24 680 μV^{-1} compared with the old value of 24 900 μV^{-1} does this matter? It may not, but similar differences should apply across all samples in the run. I would suggest that you don't look into too much of the detail but at the larger picture (i.e., what is the impact on the calculated results? do the new system's results impact the science or do they change the conclusions reached?) 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 in Step 4 of the process. It may stop you wanting to transfer data to the new system. This whole area needs to be considered carefully and a rational decision reached.

Summary

Data migration and system retirement are areas that are not normally considered in many laboratories. With increasing regulatory and quality guidelines this is changing, as records must be available over a stated retention period. If you change your CDS, what's the impact of this on the retention of data? A seven-step process is outlined and the technical feasibility and scientific impacts of the migration are discussed.

References

(1) R.D. McDowall, LC•GC Europe, 12(7), 422–431 (1999).
(2) R.D. McDowall, LC•GC Europe, 12(9), 568–576 (1999).
(3) R.D. McDowall, LC•GC Europe, 12(12), 774–781 (1999).
(4) Electronic Records; Electronic Signatures Final Rule, Federal Register 27 March 1997.
(5) J.Liscouski, Laboratory and Scientific Computing, A Strategic View, (John Wiley & Sons, New York, USA, 1995) 177–207.

Bob McDowall is Visiting Research Fellow in the department of Chemistry at the University of Surrey, and Principal of McDowall Consulting, Bromley, Kent, UK. He is also a member of the Editorial Advisory Board of LC•GC Europe.