How Raw Are Your Data — 2000?

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

What constitutes electronic records for chromatography data systems now? Why do we need to change our approach to managing electronic records compared with paper ones?

In December 1996, I wrote a "Questions of Quality" column on the options for defining raw data from chromatography data systems (1). The aim of the column was to review the recent regulatory moves regarding raw data and discuss the options for defining raw data for chromatography data systems (CDS).

Starting from the various regulations, I defined raw data as original observations and noted that exact copies could be used and that these copies could be on magnetic media. I also attempted to relate the regulations back into real life as I saw it within the laboratory and interpret the regulations and guidelines in a practical way.

With a sense of timing that could only be equalled by a sloth pushing a three-wheeled shopping trolley, the column was published four months before the issue of the *Electronic Records and Electronic Signatures Final Rule* or 21 CFR 11 (2). Given the current debate about what constitutes electronic records, I believe that it is time to revisit these definitions of raw data and update them, at least for CDS.

1996 and All That

So what did I conclude at the end of my 1996 article?

- Electronic data definitions were preferable to paper.
- Many laboratory practices involving electronic raw data were unacceptable with informally documented procedures for the use and maintenance of a CDS.

The discussion drew on the updated definition by Furman et al. (3) that "raw data were all those that could be saved and accessed later." The authors pointed out that chromatographers should consider the following scenario: two small peaks are detected eluting ahead of a peak of interest which are ignored in the current method. Later evidence finds that these are minor compounds, are toxic and the organization needs to know how much of these compounds was present in all batches of material analysed in the past year. Furman et al. then posed the question of which is preferable: reanalysing all batches of material or retrieving the old files of raw data and reintegrating them?

Furman's bottom line was very conservative: save all raw data, including the data files slices and the methods used to acquire the data, fit baselines and calculate results (1, 3).

The second main reference was the paper by the British Association of Research Quality Assurance (BARQA) (4) that outlined that raw data had four key factors:

- Original records or records of original observations.
- Recorded directly, promptly, accurately, legibly and indelibly with observer identified. Raw data generated by direct input should be identified by the individual responsible for data entry.
- Changes to raw data should not obscure the original.
- Wide range of media that could be defined as raw data.

The advantages of defining electronic raw data are that storage is compact and efficient, and, like the argument from Furman, the data are available for further processing and analysis. Data can be copied with ease to provide duplicates (back-up) to give increased security against loss or damage, providing it is documented and the copying is authenticated.

So that's were I left the debate at the end of 1996. As there are currently many debates about what constitutes electronic records for CDS, it's appropriate that I'm revisiting the subject and updating my thoughts and views.

The Intervening Years

What's happened since 1996? The publication of 21 CFR 11 in March 1997 has defined electronic records as, "any combination of text, graphics, data, audio, pictorial, or other information representation in digital form that is created, modified, maintained, archived, retrieved, or distributed by a computer" (2).

Furthermore, there is the need for efficient and effective archive and restore procedures for data systems in the chromatography laboratory as the regulations call for "Protection of records to enable their accurate and ready retrieval throughout the records retention period (§11.10c)" (2).

Departments in need of improvement: Poor practices still occur in many chromatography laboratories as the message to retain and preserve electronic data still has to sink in. Here is just a sample of those working practices. Of course, these never occur in your organization, do they?

Still deleting files? For instance, when working with stand-alone PCs that run CDS applications, when the hard disk gets full, the easiest way to resolve the problem is to delete those extraneous files that are clogging up the hard disk. A few commands or a couple of mouse clicks and your data storage space problems are solved. After all, we still have the printed paper to fall back on, don't we?

Overwriting method files? One of the features of some data systems is the ability to overwrite method files. Occasionally you may get a message that the system is going to do this or more likely the system just does it. It's a great feature, but how can you retrieve the same method that was used on the same day you acquired a particular set of data files? Not a hope.
No audit trail? You'll need to have an audit trail to help ensure that your records are acquired, manipulated and reported correctly with no falsification of results. The lack of one is a major requirement for compliance.

Culture Shock: Changing from Paper to Electronic Records

We all need to change our approach and mindset as we move from paper raw data to electronic records. Consider the paper record first. We have a pile of printed paper. Dependent on your working practices there may be a record of the

- sequence of injections
- instrument control parameters
- method used to acquire and process the data
- interpreted chromatograms with or without baselines drawn on them
- calibration method
- system suitability samples and calculations
- peak areas or heights
- any post-run calculations: (e.g., dilutions, mass or volume adjustments etc.)
- analyte amount or area normalization results.

This is to be found in the single pile of paper ejected from the printer. You'll pick this up, check it and then file it. If anyone wants to check it at some later date, some poor individual will go on an Indiana Jones expedition into the bowels of your organization's archive, moving cobwebs out of the way and digging out your data package. You are dealing with a tangible and physical medium that we all know and have used all our lives.

Moving to electronic record keeping means that we must have a different mindset. Electronic records must be considered in a much wider context than their paper counterparts. Note however, that the data retrieved tend to be similar to the list above, but you'll need the actual electronic version of the method file used to acquire the data, not a generic version. Here are working practices that must change to reflect the new electronic age. Similarly, the working practices and features of commercial chromatography data systems must also change; for instance, overwriting of files must not be allowed. However, before we can proceed much further, we have to consider the ghastly term "meta-data."

Meta-data/Schmeta-data?

The electronic records debate is becoming interesting because what should be defined as an electronic record is widening as the impact of 21 CFR 11 is being understood in more detail. Some bright spark has coined the term meta-data to describe the additional files needed to initially process and then reprocess the data. Unfortunately, meta-data is not defined in any regulations or guidelines and, in my view, causes confusion especially as it is a computer term that tends to be unfamiliar to chromatographers and other scientists.

However, to help understand this and the wider context that electronic record keeping brings, recall the definition of Furman et al. above when they wrote all data should be saved including fitting peaks etc. We must go further along these lines to understand what we must now save as electronic records for a CDS. I'll describe this in terms that will be familiar to us all in the chromatography laboratory.

Back in the Lab…

Let's look in more detail at what we can define as electronic records for an individual chromatographic run. The following descriptions you'll have to look at and modify for the specific CDS that you are using within your individual laboratories; some functions may be called something else or the system may work slightly differently.

I'll assume that you have an automated chromatograph controlled by a CDS. The system is an isocratic high performance liquid chromatograph with a variable wavelength ultraviolet detector that is operating in many chromatography laboratories worldwide. For those readers that only have gas chromatographs, the principles will be the same. However, you'll need to adjust the approach for the differences between liquid chromatography (LC) and gas chromatography (GC).

The basic files that should constitute electronic records are shown in Figure 1 with the exception being the method of storing the data files that I'll describe now.
Data organization: Before you embark on collecting data, you'll need to consider how you are going to store them: usually this will involve using a directory structure defined using the operating system or a database (dependent on the CDS you use). Either way, you'll need to ensure the storage is set up for easy archive to remove all the electronic records associated with your analysis from the system. Currently some data systems may aid you in this process but most won't, so plan and consider all the aspects.

Organizing the data will be a key aspect here as you'll need to ensure the directory or data table you want is defined to locate the files generated from the chromatographic run. The larger the work package, the more thought you may have to put in to manage all data, and all files must be named appropriately.

Naming conventions for data files, methods, sequences and reporting methods may also be important and should not be underestimated. Again this is dependent on the data system you have.

Get this right now — otherwise you'll be in trouble later in the archiving phase of your work, and an army of Indiana Joneses won't dig you out of this particular hole.
Instrument control and calibration: Again, this is dependent on the type and complexity of the data system and the instrumentation interfaced to it. If you have a chromatograph controlled by the data system, there is usually a file for instrument control within the data system that is either part of the method or linked to it somehow.

If the CDS can be used to check that the instrument is working correctly, then the data files for each analytical run must be saved as part of the electronic record. For some mass spectrometers used as GC or LC detectors the data system can control and calibrate the instrument. Data files from calibration runs should also be available to demonstrate that the instrument worked acceptably before an analytical run was started.

Setting up an analytical run: You'll have a method file that will describe how the instrument will be controlled and how the data will be acquired over the time of the analytical run. There will also be an associated file detailing the sequence of samples to be injected into the chromatograph from the autosampler or injector and an integrator file that will determine how the data system will interpret the files and place baselines and perform any post-run calculations.

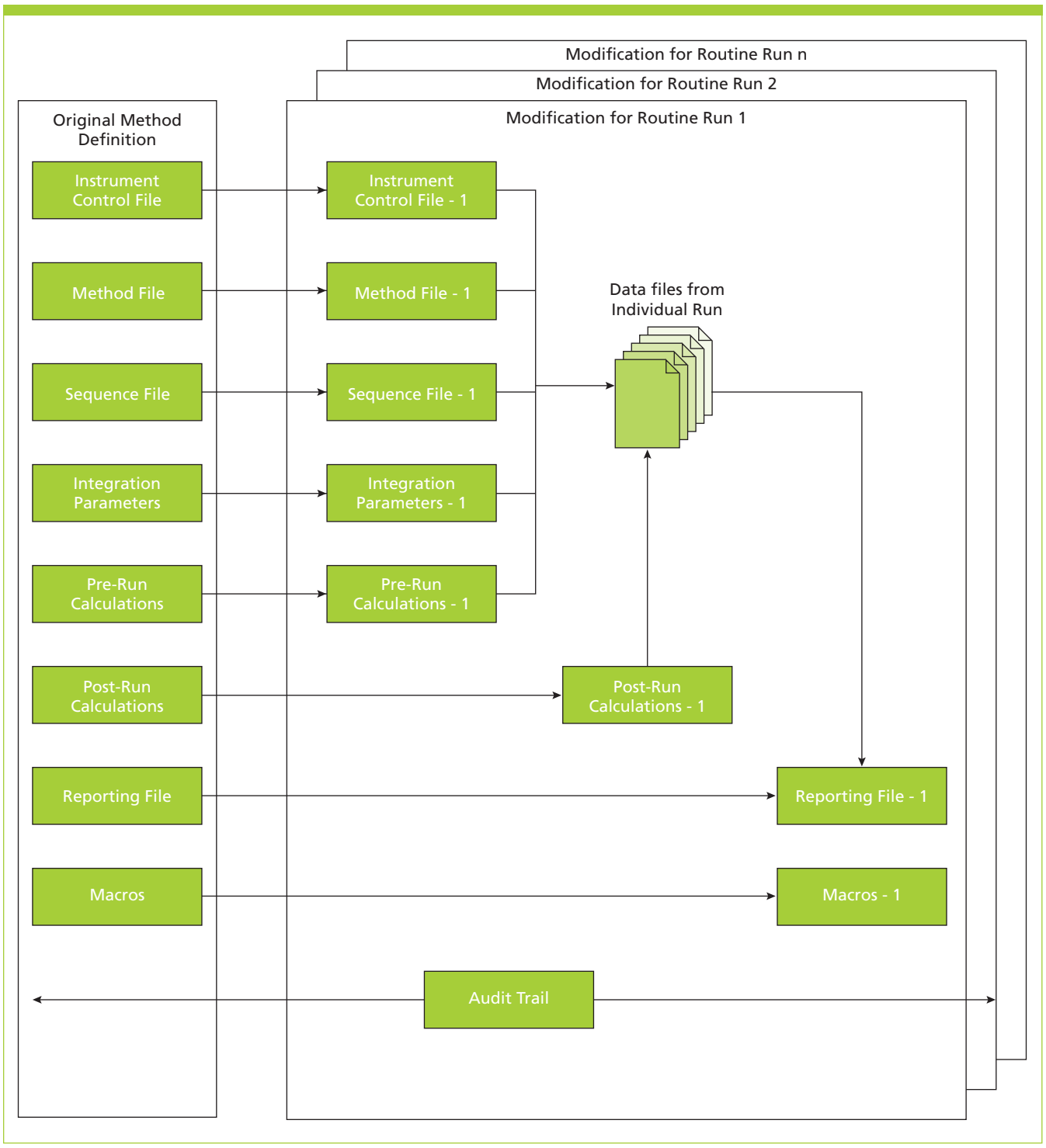

Figure 1: *Basic files that constitute electronic records.*

These files will be stored on the data system as a master set and used for each analytical run. However, what happens in practice is that you'll fine-tune individual files on the basis of the chromatographic parameters found each day you run the method. Reasons for change may be one or more of the following:

- retention times will vary slightly depending on the condition of the column
- how well a chromatographer has made up the mobile phase
- ambient temperature of the laboratory (if the column is not temperature controlled)
- robustness of the method
- pump seals and condition of the check valves.

Each version of your method, sequence file and integration file that is used with a specific analytical run must be saved. Some data systems are good about doing this, others are less so and only the latest version is saved.

Therefore, additional safeguards need to be considered from the data system and reinforced by training and documented working practices: especially version control of these files with no overwriting allowed. These features on your data system will need to be linked to the security profiles of your users: who is allowed to create, modify and delete files?

The sequence file will contain the run sequence of samples, quality controls, blanks and standards and the number of replicate injections to be made for each sample. There may be some pre- or post-run calculations required (e.g., entry of sample weights, dilution factors). There may be download of information from a spreadsheet or laboratory information management systems containing sample identity. This file is very important and must be saved otherwise *electronic* reprocessing is not feasible.

Run samples and acquire data: We are now ready to run the chromatograph and collect our data from our sample injections — press go, check everything is OK and go home for the night. Of course, the instrument will perform perfectly and when you come in the next day, you'll be ready to check the data. Note well: the chromatographic run data files we have just generated are traditionally considered as the "raw data"; I hope that you'll be considering a wider definition of electronic records after reading this column.

Interpreting the data files: Ideally, if your data system has been set up correctly, you'll be interpreting the data automatically and then checking that baselines and the like have been correctly positioned by the system. However real life does not always work on autopilot and in some (most?) instances, if the chromatography has moved over the time of the run, there will be more manual intervention. The data system should record whether the baselines have been set by the software or manually reset by the chromatographer. Again, the interpretation is an important part of the analytical record and must be captured by the data system, either as a different integration file version even to an individual injection or associated with the original file from the injection. Guess what? You have more electronic records!

After the interpretation, there will be calculation of system suitability parameters and acceptance or rejection of the run, checking that the method is within calibration using the injected standards and any quality control samples are within acceptable values. This process of evaluation of the run must be captured by the data system. Yes, more e-records.

Usually a supervisor will check the results and may reposition baselines or interpret a sample different to the original analyst. This will also be captured in the records of the data system. Got the message yet? And I've not even mentioned e-signatures!

Post-run calculations and reporting: After the initial results are calculated any post-run calculations or manipulation of the results will be undertaken such as adjusting for sample weight, dilutions etc. Some of these can be applied using the original method definitions or may be based upon run-specific data. In either instance, these figures need to be retained for any reanalysis of the work.

Reporting can be undertaken in some laboratories using a predefined method that is applied to every batch of analytical results, alternatively each run requires an ad hoc report dependent on the specific requirements of the sample submitter. Again, in either instance, the reporting file or template will need to be considered as part of the electronic records of the work.

Do you use macros? Some data systems allow you to record keystrokes or pre-program some functions to perform data analysis or manipulation using macros. Macros need to be designed, tested, validated and documented as to their correct operation before they can be used but if used they also constitute part of the electronic records of the analysis and need to be associated with the work package.

Did I Forget Something?

By the way, I haven't forgotten the audit trail records for this system; I just didn't want to get the debate out of hand. The audit trail must monitor all the data files we have discussed above and needs to be archived with the electronic records. For more discussion on the audit trail, see a previous "Questions of Quality" column (5).

The other issue to consider here is what happens in the admittedly very rare situations when you need to reinject a sample or reanalyse the whole run? The data need to be collected and stored separately; there must be no overwriting of files and again the work is recorded in the audit trail.

Home Free?

Phew, I bet you're glad that debate is out of the way aren't you? We've got everything covered and are out of the woods, home free and we can put our feet up and relax, can't we?

Er, yes and no.

If you have the equipment we discussed above then we can now be relatively relaxed about our approach to electronic records. However, do any of you use a diode array detector (DAD)? Ah didn't think about that did we? What is the impact of electronic records here?

Let's look at this in more detail. DADs can be used in a variety of detection modes:

- single wavelength
- dual wavelength
- spectral scans
- spectral libraries and compound identification.

The first situation is already covered in our debate above and all we may have to consider is that the file to set up the DAD is included in the collection of electronic records we archive. The second is also relatively simple, as you will usually get two files from a single injection that are linked to each wavelength monitored.

The most interesting situations, from an electronic records perspective, are the last two situations where spectral data are collected and libraries are used for identification of compounds. Here the impact of electronic records can hit home and you may want to buy shares or stock in data storage companies to benefit financially from some of the opportunities that this change in approach offers.

Consider that most data files in a CDS will be in the region of about 30–150 KB size compared with a DAD spectral scan that can be up to 5 MB. You only need a small run of samples collecting spectral data to realize the size of hard disk space you'll need. This is why several CDS systems will have the option to delete the original data file. Of course you'll now understand that this should not happen as you'll be destroying original records. The traceability of a sample from the original spectral scan to the final report must be available and audit trailed as well.

Huber has suggested that spectral scans are limited to the region of the chromatogram of analytical interest (6). This is a very practical solution to the looming electronic records problem. Put in a very blunt way, just how many void volume, methanol or acetonitrile spectra do you want to store for posterity, especially if you are analysing complex organic molecules?

Of more than passing intellectual interest is the situation where a laboratory uses a DAD to identify compounds by comparing the spectrum of the eluting peak with the reference sample in the spectral library. Here the spectral library becomes part of the electronic records as it is the reference point where the decision was made that peak *A* is compound *X*.

More complex is the issue where a laboratory is adding to the spectral library over time; the instance of the library at the time of the analysis is part of the whole electronic record for the analysis. See what I mean about data storage company shares? Of course, I won't bother to mention the audit trail…

As an aside, those chromatographers who have been working with MS detectors in some form or other are also in the same boat as the DAD users, but you had realized that already, hadn't you?

Computer Controlled Chromatograph with Separate Data System

Up to now we have been discussing a data system that can acquire data from an automated chromatograph and the data system itself possibly controlling the instrument. However, consider a common configuration in many laboratories: there is separate computer control of the instrument, and the data system just acquires the detector signal. What are the electronic records in this instance?

An interesting question that will generate much debate; here is my view. We now have hybrid electronic systems as the records are stored in two separate systems; usually incompatible, which means that records cannot be stored in a central location. Let's explore this situation in a little more detail. Within the data system you'll have the same e-records that we have discussed above *plus* the instrument control files with the corresponding audit trail records held on the separate PC on the computer controlling the chromatograph. This separation of the e-records between two separate systems will present problems, as some users will not think to consider that the additional PC actually has e-records. The view here may be that the method file in the data system is sufficient but this argument falls flat on its face when there is no instrument control file available from the data system.

Remember the discussion we had at the start of this column? The move from paper to electronic records will cause a number of problems because of the breadth of what can constitute an electronic record. This column has been aimed at broadening your vision of this topic.

Standards for E-Records?

The key for a solution to ensure long term archive and retrieval of electronic records is to extend the ASTM (American Society for Testing and Materials) format for chromatography data files (ex-ANDI protocols) to include at least the additional electronic records outlined in Figure 1. The aim is that any CDS application can decipher the system's operational parameters versus time data and not just the detector output versus time data. Some of the keys for electronic records are that you must have the ability to make the data human readable and can retrieve and review the data throughout the life time of the electronic records.

Again, it is worth emphasizing that you must think in a wider context than you have been used to with paper records when interpreting 21 CFR 11.

Electronic Records: The Bottom Line?

The key issue when defining your electronic records for your chromatographic data systems is to look carefully at your software application and the associated equipment.

Consider the following questions to help you reach your electronic nirvana:

- Which files in your data system are needed to set up the system to acquire data and control the equipment? These will include any methods and modifications specifically for that run and the sequence of samples that include the sample identities, volumes, dilution factors and any post-run calculations.
- Are all the required files on the same or separate computers? If the latter, how will you be able to archive them efficiently?
- Are any calibration methods and run data available to demonstrate that the chromatograph was within specification or calibration at the time of the analysis?
- Which data files were produced from each run? Are they correctly labelled and cross-referenced to the specific version of the other files used to interpret them?
- What happens with reinjected samples? Are there any overwritten files or do you have a second version of the file?

- Which files are needed to produce results from the run? Here you'll need to think in more detail about how you interpret the data files, interpret the peaks (including any manual override from the automatic operation of the system), check the system suitability results, calibrate the run with the method, calculate the results in the unknowns, apply any correction factors or post-run calculations, and approve the results and report them. Don't forget the files from the repeat samples or any dilutions.
- Have you considered all eventualities? Don't forget to include any special situations such as the use of DADs, mass spectrometers and spectral libraries.
- Have you included the audit trail entries for the run as well?
- Check that if you can archive all these files, you can restore them as well (i.e., a two-way process). Remember that the archive medium you are starting with now may not be the one at the end of your record retention period. This area, along with the rest of the computer hardware, is purely technology driven. If you start with CD-ROM as your medium now, you'll be onto DVD or magneto-optical or some medium we have not even heard of. Your archive problem is only just starting but more of that in a later "Pharmaceutical File" column. There is also the problem that the hardware (either instrumentation or computer platform) is not available, if these are also required for reprocessing of the electronic records.

There you have it, conceptually simple but the difficulty is a consistent and practical implementation. Concentrate on the electronic process and what the data system and you are both doing to identify the electronic records used or created during an analysis — then document it.

Acknowledgment

I would like to thank Kate Townsend and Mark Mercer for constructive comments and input to this column.

References

(1) R.D. McDowall, *LC•GC Int.*, **9**(12), 790–793 (1996).
(2) Enforcement Policy 21 CFR Part 11: Electronic records; electronic signatures (Compliance Policy Guide 7153.17) Federal Register 64 (1999) 41442–41443.
(3) W. Furman, T.Layloff and R.Tetzlaff, *J. AOAC Int.*, **77**, 1314–1318 (1994).
(4) Gamble, Weller and Withers: *Definition of Raw Data, British Association of Research Quality Assurance*, November 1994.
(5) R.D.McDowall, *LC•GC Europe*, **13**(2), 79–86 (2000).
(6) L.Huber, *Biopharm*, in press.

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.