Questions of Quality

How Raw Are Your Data?

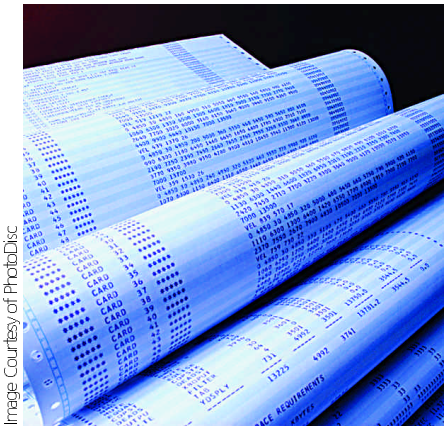

Image Courtesy of PhotoDisc

What are the options for defining raw data for a chromatography data system? Paper printout or electronic files?

R.D. McDowall

Introduction

If you want to start a healthy scientific discussion — how about: "How do you define your raw chromatographic data?" This is an essential step for chromatographers working in regulatory environments, such as Good Laboratory Practice (GLP) and Good Manufacturing Practice (GMP), or using voluntary guidelines such as ISO Guide 25.

If you are working with chart recorders or simple integrators it is a relatively short discussion: paper. However, with the use of personal computers (PCs), PC networks, or mini computer-based chromatography data systems, we now have two options: paper printout or electronic data files. It looks a relatively simple choice between two options, doesn't it? However, this apparently simple choice can lead to complex arguments and can result in actions not thought of at the time the decision was taken.

The aim of this month's "Questions of Quality" column is to review the recent regulatory moves regarding raw data and discuss the options for defining raw data for chromatography data systems. My contribution to the debate starts here. You may not agree with me and I would like your views and comments on this topic. I want to look at specific regulations and scientific opinions as to what constitutes "raw data" for chromatography data systems.

Exhibit 1: What Are Raw Data?

The first issue is to define what are raw data. The US Food and Drug Administration (FDA) and other regulatory agencies consider raw data to be the first record of an original observation, which may be collected on a paper form or on magnetic media. "Original observations" is the phrase that best describes raw data. It is usually left to the management of a laboratory or a corporation to interpret this phrase and define what constitutes raw data in their own environment.

In addition, GLP regulations provide for exact copies of raw data to be used interchangeably with the original raw data, which may include copies stored on magnetic media.

Exhibit 2: US GLP Regulations

The FDA GLP regulations (1) only refer to the term "computer" in two sections:

- Section 58.3(k), which provides the definition of raw data, mentions the terms computer printouts, magnetic media, and recorded data from automated instruments.
- Likewise, Section 58.130(e) describes the special features of data when automated data collection systems are used.

These sparse references notwithstanding, the FDA maintains that the GLP regulations provide sufficient guidance for the development, installation, validation, use, and maintenance of high quality automated systems. Otherwise computers are deemed to be equipment, and must be fit for this purpose and designed correctly.

Paul Lepore (2), a member of the FDA, states that laboratory observations, which are made so that there is direct data capture onto magnetic media, means the media are considered to be raw data. Observations made on paper, which are subsequently entered onto magnetic media, means that the paper constitutes the raw data. The GLP regulations confer raw data status to the magnetic media and to the computer printouts that accurately reflect the information stored on the media. Of course, the system that accomplished the printouts must be sound.

Exhibit 3: Environmental Protection Agency Report

In a report (3), that helped give rise to the Good Automated Laboratory Practice (GALP) guidelines, sub-part J (records and reports) discusses raw data. This is defined as original observations with the usual list of media, which includes magnetic media. Furthermore, it states that data that are captured through automated data acquisition or direct keyboard entry are raw data.

Back in the Lab...

In contrast, most chromatographers prefer raw data on paper. Even if they have the latest state-of-the-art data system, the easiest way is to define raw data as either the first printout from the data system or the final interpreted chromatograms. Also, the majority of inspectors or assessors seem to prefer paper. In a manual system the definition of raw data is clear. It is usually the first time the observation is written down: thus the written form is the closest to the original observation. This has to be protected from unauthorized and undocumented changes.

Look also at the benefits: paper is tangible. You can touch, feel, and see it. It is also permanent: well almost, excluding, of course, those minor mistakes where concentrated acid is spilt on the latest batch of chromatograms or you put the thermal paper next to a window on a bright day.

What about the ease of searching through a paper archive to find those elusive results? At times like this you begin to wonder just how many of your laboratory's chromatograms it would take to circle the globe. So paper may not be the complete answer in today's chromatography laboratory, as we shall see later.

Exhibit 4: Taylor's Definitions of Raw Data

To help, Taylor, an FDA inspector, attempted to elaborate the definition of raw data in 1984 (4). Although this is relatively old, it still offers some perspective on the definitions of raw data.

- Where observations are converted to a written record, the written record becomes raw data.
- Where observations are converted to written records which are photocopied, microfilmed, or converted to magnetic media by input to a terminal, the raw data can be defined as the original material, microfilm, or photocopied records. The information recorded within the computer is not raw data.
- When a written record of observations is transcribed by terminal input into a computer and stored in a database, only the written record is raw data.
- When observations are input directly to a computer database, the database becomes raw data.
- Where observations are input directly to a database and reports are produced, the computer database can be raw data. Also, the resultant printouts must contain the collected raw data. However, if the printouts contain any derived data, they are not considered raw data. A key question is how can the laboratory demonstrate that the printouts faithfully contain the raw data?
- In a situation where observations are recorded by a chart recorder and entries are also made by the operator in a manual log, only the chart recording is considered raw data.
- Where the observations are recorded by a chart recorder and integrated or calculated values are produced on a printout, either the chart or the printout can be raw data, but not both. The decision as to which becomes raw data is left to the laboratory management and should be documented before any inspection.
- Where observations are accomplished by equipment that produces printouts with integrated or calculated values, then the printout is considered raw data.
- Where observations are entered directly into a computer, and the computer returns the input as incomplete or erroneous and the operator corrects or deletes data, and finally the observation is entered to the database, the database is considered raw data. Note that any changes to the data after they have been entered into the database must be documented. If the data are further manipulated, reduced, or plotted they are not considered raw data.

Back in the Lab...

OK, so where do chromatography data systems exist within these definitions? In the case of integrators that have no long term data storage capacity but do have paper printout, the printout is the raw data.

But, with PCs, networked and central data systems, the raw data, as defined above, are the electronic files. Of course, you do keep all your electronic data, don't you?

However, with chromatography data systems there is not always a clear definition of raw data. The data storage medium will usually be magnetic: a floppy or hard disk on a PC, network, or mini-computer. The raw data must be relatively permanent, protected against unauthorized change, and reflect the actual observation. However, magnetic media are not permanent: floppy disks are not suitable for an archive, hard disks may fail, and tapes can become brittle. The bottom line is that magnetic data must be copied regularly. Optical data storage, either CD-ROM or magneto-optical disks, may be more robust but that does not mean that you should not have duplicate copies of the disks. Copies of magnetic or optical media have to be certified and duplication procedures must be documented. Of course, regardless of the approach, it is implicit that you should occasionally test some of the old archives to see if the data can still be retrieved.

How long should you keep your chromatographic data? Most regulatory agencies specify how long raw data can be kept, however, many companies can keep this data indefinitely because of legal implications. Just think about backing up all those tapes and disks! And you thought jobs for life had disappeared!

Exhibit 5: Good Automated Laboratory Practices

The original Good Automated Laboratory Practices (GALP) was issued as a draft in December 1991 but was only issued in final form in August 1995 (5). The final document refers to Laboratory Information Management System (LIMS) raw data only. However, the definition of a LIMS is a little strange. The GALP guide uses the acronym LIMS to describe any automated laboratory system that collects and manages data, and there are a limitless range of possible configurations of automated data collection systems. Perhaps some enterprising marketing manager can make some money selling any instrument with a PC connected as a LIMS!

The GALP definition of "LIMS" raw data is "original observations recorded by the LIMS that are needed to verify, calculate, or derive data that are, or may be, reported." OK, much the same as before, however, GALP goes further and makes two further statements:

- LIMS raw data storage media: the media to which LIMS raw data are first recorded.
- Instrument transmitting LIMS raw data is uniquely identified when the data are recorded, and the time and date are documented.

Exhibit 6: Furman, Layloff, and Tetzlaff 1994

Written by three senior members of the FDA, this paper (6) is mainly concerned with the validation of computerized liquid chromatographic (LC) systems as a whole. There is a section on the definition and storage of "raw data". The usual definition of raw data in the Good Laboratory Practices 21 CFR 58.3 (k) (1) is quoted. Then, the authors propose a further definition of raw data: of the data captured by the data processor, raw data are all those that can be saved and accessed later.

To make the point, there then follows a very pragmatic debate. However, the authors point 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 minor compounds are toxic and the organization needs to know how much of these compounds were present in all batches of material analysed in the past year. Furman et al. then pose the question: Which is preferable, re-analysing all batches of material, or retrieving the old files of raw data and re-integrating them? The attitude of Furman et al. is very conservative: save all raw data. This includes the data file slices and the methods used to acquire the data, fit baselines and calculate results.

Back in the Lab...

However, before you all rush to redefine your chromatographic raw data as electronic. Furman et al. sound a note of caution about the paperless laboratory: there are several arguments for retaining printed copies of all data used to make decisions. For a given set of raw data it is unlikely that two qualified independent chromatographers will make exactly the same choices during interpretation: where to place the baselines and noise rejection level are the main areas here. The argument is that it may not be possible to fully reconstruct how the final decision was made. Second, the LC system may become obsolete, although with network Common Data Format (net CDF) files this may not be a major argument now. Third, review by internal and external staff is greatly facilitated.

Exhibit 7: OECD GLP for Computerized Systems

The new GLP consensus document from the Organization for Economic Cooperation and Development (OECD) on computerized systems (7) is very instructive. The document defines raw data as all original laboratory records and documentation, including data entered directly into a computer through an instrument interface, which are the results of original observations and activities in a study, and which are necessary for the reconstruction and evaluation of the report of that study. Computerized systems operating in compliance with GLP principles may be associated with raw data in a variety of forms, for example, electronic storage media, computer or instrument printouts, and microfilm/fiche copies. It is necessary that raw data are defined for each computerized system.

Where computerized systems are used to capture, process, report, or store raw data electronically, system design should always provide for the retention of full audit trails to show all changes to raw data without obscuring original data. It should be possible to associate all changes to data with the persons making those changes by use of timed and dated (electronic) signatures. Reasons for change should be given.

Think carefully, can your data system provide full audit trail facilities? Such as:

- old and new data values
- date and time the change was made
- who made the change
- the reason for the change.

It is the last point where a chromatography data system will fail this regulatory requirement. Mainly because we, as users, do not ask or demand of vendors that we should have such features.

If you are not already depressed, you will be after reading the next section! "No electronically stored data should be destroyed without management authorization and relevant documentation." In some cases, laboratories that define "raw data" as paper will delete the electronic files after a certain time. Does anyone want to bet a beer that there is no procedure for this?

Exhibit 8: Quality Assurance Perspective

It is fitting that (nearly) the last word in this column is left to that group of individuals that often have to sort out the mess that the chromatography laboratory produces. Yes, it's quality assurance! The British Association of Research Quality Assurance (BARQA) have produced a report on the definition of raw data (8).

Raw data has 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.

Raw data is the first record of an observation. Therefore, according to this document, "It is unwise to define chromatographic raw data as a calculated peak height or area. The raw data are either the signal which allows the peak to be displayed or the image of that peak." Now, that's a surprise or a shock for some of you!

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. If raw data are defined as paper, there is an increased risk of the electronic data being modified or edited whilst the raw data on paper remains unaltered. Therefore, procedures must be set up to restrict editing of data files on the data system and the paper is updated as necessary. If paper is defined as the raw data, what happens to the data files? In some laboratories these data files are deleted after a set time has passed or the report of the work has been circulated. Does this mean that the original raw data have been destroyed?

The "Ideal" System

- During the specification of the chromatographic data system, raw data were thoroughly considered and defined as electronic and resident on the system (with back-up copies for security).
- Associated with each data file, is the time and date when it was recorded and the identity of the chromatographer who was responsible for collecting the data.
- Electronic signatures (personal identity number, barcode plus password) for unique identity of the operators and any supervisors involved with approving the interpretation and results generation.
- Audit trails exist to record any change to the method. Most chromatographic data files cannot be modified, therefore, the audit trail of the method used to interpret the data file should be modified.

Back in the Lab...

I have looked at the issue of raw data from several viewpoints. Are you going to try anything different? As the final word, does your system match the ideal one as described by BARQA? If not, what are you going to do about it?

References

(1) United States Food and Drug Administration, Title 21 Code of Federal Regulations Part 58. Good Laboratory Practice for Non-Clinical Laboratory Studies, 1994.

(2) P. Lepore, Laboratory Information Management, Chemometrics and Intelligent Laboratory Systems, 17 (1992) 283–287.

(3) D.W. Taylor, Drug Information Journal, 18 (1984) 189–194.

(4) B.A. Hamilton, Automated Laboratory Standards: Evaluation of Good Laboratory Practices for EPA Programs, EPA Draft report 1990.

(5) Good Automated Laboratory Practices (GALP) EPA 2185 (Principles and Guidance to Regulations For Ensuring Data Integrity in Automated Laboratory Operations with Implementation Guidance) 1995 Edition.

(6) W. Furman, T. Layloff, and R. Tetzlaff, JAOAC International, 77 (1994) 1314–1318.

(7) OECD, Good Laboratory Practice Consensus Document: Application of the Principles of GLP to Computerized Systems, Paris, France, December 1995.
(8) S.H. Gamble, D.L.M. Weller, and P. Withers: Definition of Raw Data, November 1994, British Association of Research Quality Assurance.

"Questions of Quality" editor Bob McDowall is visiting senior lecturer in the Department of Chemistry at the University of Surrey, and Principal of McDowall Consulting, Bromley, UK. Direct correspondence about this column to "Questions of Quality", LC•GC International, Advanstar House, Park West, Sealand Road, Chester CH1 4RN, UK.
e-mail: lcgc@advanstar.co.uk