

# Archiving Requirements for Electronic Pharmaceutical Manufacturing Documents and Associated Executable Software

THOMAS QUINN

*The Hollis Group, Incorporated, Paoli, Pennsylvania*

---

*ABSTRACT: Many pharmaceutical manufacturers are currently evaluating the feasibility of electronic batch record (EBR) and electronic document management (EDM) systems. Considerable effort has been invested in the design of the batch record files and electronic signature devices and procedures. Much less consideration has been given to the potential need for pharmaceutical manufacturers to be able to re-create the operational software environment necessary to review archived documents at some future date. The paper discusses methods, policies, and equipment that can be used to fulfill this function.*

---

## Background

Among pharmaceutical manufacturers there is considerable interest in the transition from paper-based batch records to computerized batch record systems. Work has progressed to the point where several systems have demonstrated the dependable capability to support manufacturing operations. The FDA recently published a proposed rule, Title 21, Part 11 of the Code of Federal Regulations (21 CFR 11) (1) that is expected to affect the design of these computerized batch record systems.

One aspect of the "paperless factory" that has escaped rigorous treatment in current literature is the need for archiving systems to support electronic records. Section 10(b) of 21 CFR 11 requires "the ability to generate true copies of records in both human readable and electronic form suitable for inspection, review, and copying by the agency." The author believes it prudent to carefully consider the computer systems that are (and will be) required to maintain and reproduce such "true copies" of electronic records. This writing is intended to serve as a beginning of such a discussion.

## Scope

It is possible to consider the subject of computerized batch records as a set of three groups of functions, namely:

EDM = Electronic Document Management

EDD = Electronic Document Distribution

EBR = Electronic Batch Records

where each functional group includes a superset of the functions of the one(s) preceding it (2). This view

corresponds (roughly) to one FDA investigator's\* decomposition of automated systems' documentation requirements into the subsets of reviews, instructions, and events (3).

EDM systems automate the processes of document creation, editing, and approval, and provide the facilities for the review of the documentation necessary to demonstrate compliance with Current Good Manufacturing Practices (CGMPs). EDD systems include the requisite computer hardware and software to distribute documentation, such as electronic copies of standard operating procedures, (SOPs) and provide the facilities to demonstrate the use of current instructions. EDD systems require libraries of electronic documents and must include or be supported by EDM systems. EBR systems add the function of data acquisition and provide the facilities to record events and must include or be supported by EDM and EDD functions.

The scope of this writing is limited to an examination of the requirements of and technology available for an archiving system that would support an EDD system as described above. It is presumed that such an EDD system would be used to replace manufacturing area paper copies of reference documents, but that batch production and control records (BPCRs) would be completed manually. It is also presumed that such an EDD system would include an EDM subsystem that would be used in the design and maintenance of master production and control records (MPCRs), electronically distributed SOPs, and other documents as specified in 21 CFR 211. The discussion will focus on these two types of manufacturing documents (MPCRs and SOPs) as examples that suffice to represent documentation in general. Electronic batch production and control records are beyond the scope of this discussion.

---

Received November 4, 1993. Accepted for publication May 12, 1994. Presented at the Annual PDA Meeting, Orlando, FL, November 1993. Correspondence address: 19 East Central Avenue, Paoli, PA 19301.

---

\* DR. Tetzlaff, who was with FDA at the time of his writing (3) & (6), is now a private consultant.

## State of the Art

In current general practice, MPCR's and master copies of SOPs are maintained as paper records, and photostatic copies are distributed on an as-needed basis to manufacturing personnel (4, 5). Superseded versions of such documents are retained for some period of time as paper copies and are typically converted to microfilm or microfiche as they become more out of date. For this discussion, the term "hard copy" will be used to refer to paper, microfilm, or microfiche copies of documents.

The attendant costs, creation and distribution tardiness, and version control difficulties associated with such paper systems have prompted many pharmaceutical manufacturers to attempt to install computer hardware and software that will maintain "electronic libraries" of these master documents, in lieu of paper copy files.

One very common type of document management software provides a working environment where documentation personnel can "check out" copies of word processor files (via a local area network or LAN) from secure libraries, edit the files on personal computers running commercial word processing software, and "check in" the revised documents with the secure libraries. The document management software limits access to authorized users, tracks revision activities, and maintains versions of the original and edited document files.

A representation of a typical personal computer-based EDM system is shown in Figure 1. A plurality of end-user personal computers are connected, via a local area network, to a document server. The end user computers each have a locally resident copy of the EDM client software and a copy of the word processing software. The server computer has the server portion of

the EDM software and the mass storage device for the electronic library of documents.

## Problem Statement

FDA regulations are currently being interpreted to require manufacturers to provide "... access to electronic records in a manner analogous to reviewing original hard-copy production or analytical records in a batch folder" (6). A requirement for any EDD system is the capability to retrieve copies of documents at some future date. It is fundamental to fulfilling this requirement that the system reproduce demonstrably accurate copies of the original documents.

When being used to reproduce copies of original documents, hard copy systems have the advantage of being based on a physical representation of the original (in the case of paper it is the original) document which can be directly viewed. An EDD system, however, depends upon a coordinated set of software entities, configuration files, and computer system components to reproduce a screen image or to print a copy of the original document based upon an archived computer file of that document.

For this discussion the term "operational configuration" will be used to refer to a specific set of software, files, and hardware.

In the author's experience, some operational configurations have been assembled from commercially available word processing and other software components. These "off the shelf" software components are subject to revision by their manufacturers on a regular basis.

Attempts to accommodate these software upgrades, as well as configuration tuning, and hardware replacement, frequently result in new operational configura-

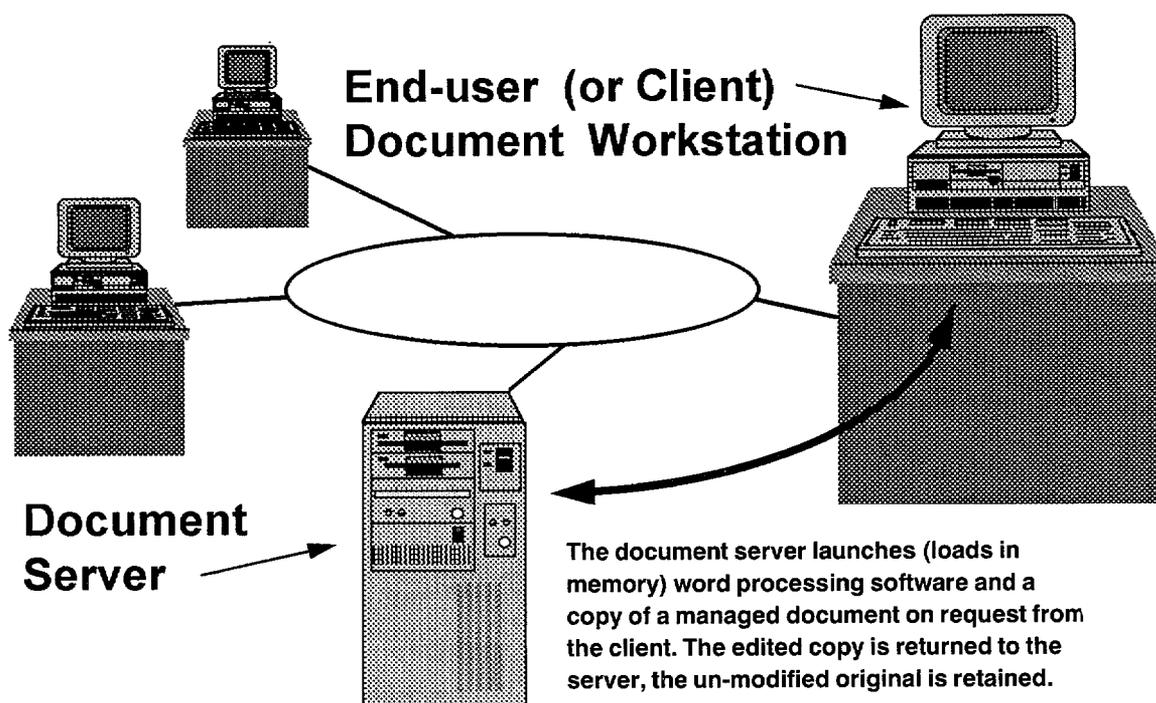


Figure 1—Typical computerized document management system

tions that are not capable of reproducing accurate copies of archived documents. In this situation, the archives of original document files may not be sufficient to comply with the retrieval requirements of the FDA. Such system can not be considered acceptable means of maintaining original records or copies of original records.

EDM/EDD systems that are based (Fig. 1) on a client-server topology and off-the-shelf computer software are particularly vulnerable to problems with document retrieval. An exhaustive list of difficulties is beyond the scope of this discussion, but several general classes of problems warrant mention.

**Automatic File Updating:** Most commercial word processing software includes the provision for viewing and editing document files that have been created on previous versions of the software. Some word processing software automatically converts older file formats when users, who are using a newer release of the software, open such files for review. Often this document conversion is poorly announced to the users.

For example, with one popular word processor, it is not possible to stop file conversions from occurring, although a properly trained and attentive user will know not to overwrite the existing file with the converted one. Unfortunately, the author has observed actual overwrites of original archive file copies because of this "feature" of the software.

One FDA investigator expressed concerns about the acceptability of equivalent backup copies (6) of original files. If accurate copies of archived files are not to be trusted, it is also reasonable to expect that modified source files would not be accepted as "original files" per the regulations.

**Layout Incompatibility:** Special layout information, including page positioning of text, underlines and boxes, paragraph and page numbering, is usually coded within files as control code sequences that follow schemes that are specific for and proprietary to individual software programs. It is often not possible to duplicate layouts on later software versions. It is rarely possible to duplicate layout between different vendors' software.

One of the more drastic examples of layout incompatibility is that one of the most popular word processing software programs deletes header and footer information from files that are saved in ASCII (American Standard Code for Information Interchange, a generic representation) format. Clearly, if an original document file had included a header block, containing title, part number, or lot number information, and that header had been deleted in this manner, it would not be an acceptable archive copy.

**Font Incompatibility:** More advanced personal computer systems tend to have operational configurations that include windowing software. Most windowing software supports a multitude of printer and display options. Configuration files allow the substitution of various fonts by the display drivers of windowing environment, the print driver software, and multiple word processors.

Dated files, especially ones that were developed for simpler dot-matrix printers, tend to not have font definition declarations within the files. When these files are displayed or printed on windowing systems, unexpected font definitions are used. Line and page lengths can vary due to proportional character spacing, resulting in page numbering differences. If retrieved documents have page number references within their text, these references will be incorrect, making the retrieved documents appear to have had errors in their original form.

### Approaches to Archiving Operational Configurations

Let us consider four possible approaches to an archiving system that retrieve accurate copies of outdated documents, and avoid some of the common problems listed above.

**"Parallel Pathing":** In this system the manufacturer retains signed, paper copies of all CGMP documents. An EDM system is used to increase the efficiency of document development and change, but official copies of all documents are still reviewed, signed, and retained.

**The "Environment Archive" Scheme:** The basic philosophy of this scheme is to retain all of the computer hardware, software, and configuration information necessary to reproduce, at some future date, an equivalent operational environment in which to view the archived documents. Archive copies of all of the versions of the software components of the operational environment, including operating systems, word processors, windowing software, and network software, are retained in secure archives. Also, a facility is included to archive a configuration map of all of these components. Typically, this is an ASCII text file description of the environment, and a procedure to re-establish the environment. The author recommends that this text file be duplicated, and retained, on paper.

In order to implement an environment archive system a number of coordinated policies and procedures are required.

The word processor-compatible files that contain the master records must be securely archived. All executable software entities and configuration files must also be archived. Records of the installed configuration (namely all of the software, hardware, versions, etc.) associated with each archived master must be recorded and archived. Representative computer hardware must be retained.

A detailed configuration management plan must be established to ensure that the software environment of installed and operating executables is as expected. The configuration management plan must also include hardware elements. Regular audits must be conducted to demonstrate compliance with this plan. Such audit records are the basis for the trustworthiness of the archive retrievals.

A set of "test vector" documents must be developed so that the correct re-installation of the various versions of the operating environment may be demonstrated,

independent of viewing archived documents. Versions of these test document sets must be developed for new executable environment configurations as they are brought into production.

An alternative to preserving physical records or the original systems used to produce electronic records is to preserve electronic records in a generic form that can be consistently retrieved. An archive system that retrieves the electronic document essentially emulates and replaces the operational configuration of the original system.

It is possible to describe two broad categories of these generic formats. One can be described as a raster, or bitmapped format that is based upon capturing a computer-readable image of a computer screen. The second format can be described as a vector format, or one where drawing objects, text, and text descriptors are stored in a computer-readable file.

It is important to note that within these categories many files types and formats still exist. Also, the categories themselves can not be considered as black-and-white (pun intended) separations because many software vendors embed elements of both in storage formats.

**The "Snapshot" Scheme:** The term "snapshot" refers to computer file that is an image of the electronic document that is captured while it is being used in its execution environment. This image is stored in a format that is not dependent upon any specific, executable computer program for its retrieval and display. A good example of this type of image is a bit-mapped image of a display screen.

The snapshot method is analogous to an electronic microfiche of documents. One way to visualize such a system would be to consider a document scanning system, that records digitized scanned images of archived documents. If the documents in such a system have originated as word processor files that are subsequently printed and then scanned, a snapshot system simply removes the printing and scanning steps from the process.

The snapshot scheme has several requirements: The snapshot format must be defined. Several choices exist for this, including many for black-and-white (pun not intended) and color bitmaps. PCX, or PC Paintbrush, which is produced by many commercial "paint" software programs is a typical example of such a format. Many digital image storage subsystems are commercially available to process and store these files. Public domain and proprietary data compression and recovery algorithms are also available to improve the performance of these systems.

In order to use this method, a pharmaceutical manufacturer must develop a document retrieval and viewing system, capable of reading the document image files. This system must be validated, and its accurate retrieval of documents must be maintained for the duration of the document retention period. Also, policies and procedures must be developed, instituted, and followed to test the reproduction fidelity of new versions of image file

retrieval and viewing systems with older versions of document image files. In the author's experience, the expense of this task of ongoing fidelity testing and validation is frequently underestimated.

**The "Least Common Denominator" Scheme:** With this kind of archiving system, a "least common denominator" (LCD) document file rendition language or format is developed or adopted. This LCD is a vector format in that the LCD is a computer file of the description of a document, or a page of a document, in some form of descriptive code or language. PostScript, and SGML are typical examples of such an LCD.

Such a system requires that all renditions of the document, pages, or screens (depending on the LCD used) are processed through this LCD at run time. In other words, a policy is established that all application or rendition software must operate with the LCD file definition of a document during the processing of the file.

Requirements for the least common denominator scheme include: The LCD must be chosen, and all application software must support and conform to LCD. Testing this support would be part of the acquisition process for software.

A document retrieval and viewing system, capable of reading the LCD-formatted document files, must be developed, validated, and maintained for the duration of the document retention period. Policies and procedures, similar to those for a raster system must be developed and followed.

## Conclusion

EDM/EDD systems manage computer files that contain electronic copies of manufacturing documents. These files require the intervention of a computer system to reproduce copies of the original documents.

It is reasonable to require that an archive retrieval system be able to reproduce exact copies of archived documents. It is also reasonable to expect that an EDM/EDD system's operational configuration will change over time. Some of the most common software used with these systems produces dissimilar reproductions of original document files with different versions of the software. Such systems can not be considered to fulfill the records retention requirements of the CGMPs.

Several approaches to CGMP-compliant archiving of electronic records are possible. It is necessary to demonstrate that any archive retrieval system reproduces accurate copies of documents that have been archived under all versions of an EDD/EDM system's operational configuration. Policies and procedures, test plans, and test document sets need to be developed and used to verify the accuracy of document retrieval. These tasks are complex and it is prudent to expect that this will require considerable design and operational effort.

## References

1. FDA, 59 *Federal Register* 45160 No. 168, 31 August 1994, Docket No. 92N-0251, "Department of Health and Human Services, Food

- and Drug Administration, 21 CFR PART 11" "Electronic Signatures: Electronic Records," "Proposed Rule."
2. T. Quinn. "Workflow redesign for pharmaceutical batch record electronic document management systems." *Proceedings, ISA '93*, Instrument Society of America, #93-408, (1993).
  3. R. F. Tetzlaff, "GMP documentation requirements for automated systems: part II." *Pharmaceutical Technology*, **16** (4), 60-72, (1992).
  4. C. V. DeSain. "Documentation basics: master production batch records." *BioPharm* **4** (10), 20-24, (1991).
  5. C. V. DeSain. "Documentation basics: standard operating procedures and data collection forms." *BioPharm* **4** (9), 22-29, (1991).
  6. R. F. Tetzlaff. "GMP documentation requirements for automated systems: part I." *Pharmaceutical Technology*, **16** (3), 112-124, (1992).