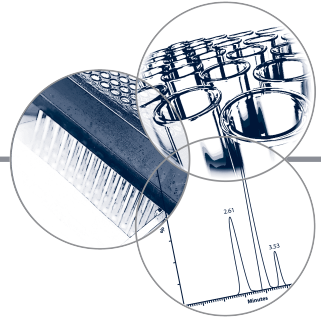

Challenges in archiving electronic bioanalytical data supporting GLP studies in a CRO

The purpose of this article is to articulate the fundamental issues of archiving electronic GLP data in a CRO environment. CROs struggle to address the absolute requirement of archiving GLP studies electronically. The difficulty of adhering to this is partly due to the wide variety of systems and types of electronic data. Often the end solution to this complicated issue is printing the data at the end of the study and in turn archiving the paper data, foregoing or ignoring the fact that scientific decisions were made while reviewing electronic data. Paper data and electronic data can be different. For example, chromatographic resolution or being able to focus in on detailed integration of chromatograms can significantly change the perspective of the data. While the core goal of archiving according to GLP principles are met, the reality is much different. Businesses purchasing CRO capabilities and regulatory agencies have been quite clear that when it comes to auditing the data that the electronic record is the preferred and often required source of the information for auditing and review purposes. The fact is that paper data do not always provide the flexibility, sensitivity and complete data context that an electronic record can provide. The ability to adhere to compliance standards due to the electronic data cannot be undervalued and by printing the data after all the acquisition, decisions, scientific judgments and review, do a complete disservice to the multidimensional (such as colors for deactivated or changed data, audit trails, 'zooming-in functionality', direct data links and snapshots) data in an electronic system.

Unprecedented number of electronic records

It is reported that more than 90% of the regulatory records being created today are electronic, and therefore the issue of archiving is a critical business and regulatory issue [101]. In August 2003 the US FDA issued a guidance document entitled 'Part 11, Electronic Records; Electronic Signatures – Scope and Application' [1] where it was very clear that all records governed by the predicate rules must follow a record retention policy and records must be available. This ongoing penetration of information technology enables many laboratory efficiencies but it also generates unprecedented amounts of data that must be properly managed and retained. In addition, printing electronic data for archiving purposes does not support the realities of today's digital laboratory environment. Paper data and electronic data can be different. For example, chromatographic resolution or being able to focus in on detailed integration of chromatograms can significantly change the perspective of the data. The paper record is generally printed at one resolution or height and unless all of the iterations of the chromatogram are printed, data is also lost in this method as the 'before' and 'after' pictures are not available. There are a variety of sources available on discussing the process of archiving electronic records. The most specific source was developed by the Swiss Medic based around the requirements of OECD No. 10 'OECD series on Principles of Good Laboratory Practice and Compliance Monitoring' and has very specific requirements and details surrounding archiving GLP data [2]. Good automated manufacturing practice (GAMP) has also published a reference on Electronic Data Archiving and delves into other types of records besides GLP records, which can aid companies in prioritization of data records for archiving.

Archiving strategy

An **archiving** strategy or policy must be determined for every company. This includes data types, data owner, retention period, regulation(s) governing that data type and a retention review date (a date the **data retention policy** for that data type or system is reviewed). This seems like a monumental task, but in the scheme of things this particular activity is well worthwhile and can save a bioanalytical laboratory a lot of time in the future. It can be accomplished by building a project with assigned timelines based on the data categorization and data owners (FIGURE 1). An archiving strategy

Merry Danley
Celerion, 621 Rose Street, Lincoln, NE 68502, USA
Tel.: +1 402 770 4317
E-mail: merry.danley@celerion.com

FUTURE SCIENCE part of fsg

Key Terms

Archiving: A collection of records held for official reasons or because of the status, role or value of the records. By extension, an archive is also often the physical or logical space independent of a production environment where records are held, protected from loss, alteration and deterioration so that they may be retrieved in the future, for example, to be used as trustworthy evidence.

Data retention policy: A policy that outlines the data owner, source, retention requirements and retention review period.

Electronic archive: Electronic records that have been identified as having to be archived.

GLP archiving: A collection of records that support a GLP study.

may also include yet another electronic system for archiving bioanalytical data. These systems are termed as electronic data management systems, or content management systems. These systems are easy to justify a return on investment, but yet are limited to the data feeding into them. These systems may provide one secure 'location' to archive data but it does not actually solve all the issues surrounding data retention, long-term data readability and data integrity. These systems can take flat or file types of data and sweep them into a secure database providing data protection and logical access controls. These electronic data management systems systems can also make data retrieval very quick with sophisticated searches, data mining capabilities and archival logs. And yes, many systems provide some meta-data and history as part of the archived records but unfortunately, these systems cannot always display the data in its native format (particularly for legacy or custom applications). Specifically, chromatography data systems where the original application is often required in order to re-evaluate the chromatography data for a bioanalytical study. The archiving of electronic raw data means the process of protection against loss, modification and unauthorized access [3]. It is important to make sure that these requirements are met when developing the archiving strategy.

Archiving can also be performed as part of the system in an on- or off-line fashion. Logically, data can be archived offline on different media types. Electronic raw data stored off-line should be well protected against accidental data changes and physically separated from the productive environment [3]. This particular method is very dependent on the media selected to archive the data and a risk assessment should be focused in this area. Due to the physical separation of the data media, the maintaining of the overall index and the amending procedures may be difficult to handle. In addition, direct access to the data for any form of data warehousing or data mining is not feasible. In this instance, media and storage along with validation of those components must be scheduled on a regular basis. This can be a resource-heavy task with large amounts of data and with different media/timeframes.

Online archiving data has additional controls that must be put in place. The data must be 'frozen' or 'locked' to ensure that it does not change over time. This poses various risks such as software versions and data components may be upgraded, company names are changed and other features that are in the application that were not available during the time of the original data during the operational phase of its life cycle. Online archiving can be achieved either using a dedicated **electronic archive** system (physically separated) or for designation on a productive environment, by marking the electronic raw data explicitly as archived (logically separated) and readonly [3].

Furthermore, any of these solutions can be outsourced to other vendors (data archiving or warehouse companies). It is important to ensure that those vendors have appropriate service level contracts set up with the CRO clearly describing the responsibility of the service provider and the CRO. The service provider should also comply with GLP requirements surrounding training, proper job description documentation and current CVs.

Archiving format & dilemma based on the life cycle of data

It is well known that computerized systems to support GLP related studies should be developed, validated, operated and maintained [4]. This system implementation process should include specific testing for **GLP archiving** of data. Scripts should be written and executed to adhere to clear archiving requirements relating to the CRO's archiving policy. Care should be taken to ensure that the archived data meets the company requirements paying particular attention to data legibility, accuracy

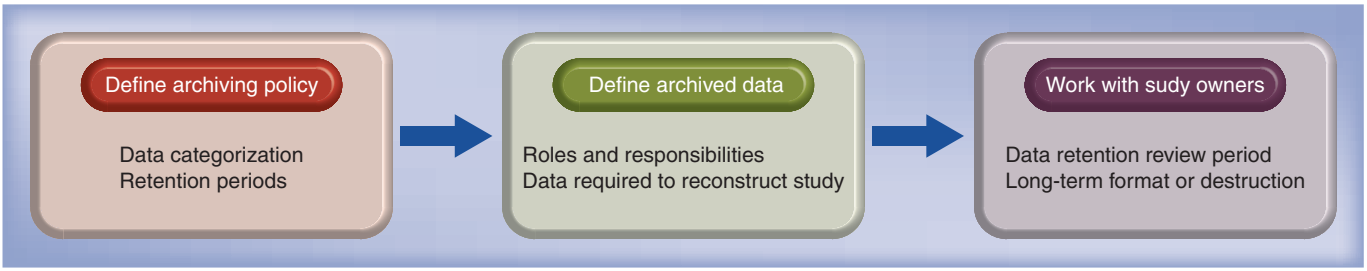

Figure 1. Archiving policy roadmap.

and formats. Testing should include verification that all data selected to archive is present in the archived form (preferably in human readable format) along with the ability to de-archive the data in the event that data must be re-processed or -evaluated, or reviewed. In addition, specific controls and responsibilities must be put in place to adhere to GLP regulations for test facility management and study directors. According to the Organisation for Economic Co-operation and Development (OECD), the study director's responsibility is the same for data recorded electronically as well as on paper. The Test facility management has the overall responsibility of the system to comply with the GLP principles [4]. The Federal Office of Public Health (AGIT) guidelines are even more specific indicating that "The study director should ensure the proper archiving of all data related to the study, including electronic raw data", along with receiving confirmation that data has been archived, and any interactions between the archivist and the study director be documented. In addition, they must ensure that appropriate facilities, equipment, materials and SOPs are available and that electronic archiving systems are suitable for their intended purpose and are validated, operated and maintained in accordance with the principles of GLP. The test facility management must establish procedures for all aspects of the life cycle of archived electronic raw data (i.e., archiving, amending, maintaining, migration, reformatting, conversion, retrieval and deletion). These roles must be clearly defined by the laboratory [3].

The life cycle of data must be reviewed while determining the archiving philosophy of a particular system (Figure 1). During the operational phase data is created, protected and reviewed. For a period of time after data is created it is in the operational state and it is needed to complete ongoing laboratory operations and study completion [102]. Then, after some additional period of time, the data moves into an area where it is no longer needed for completing laboratory transactions and during this phase, it is only required for querying and reporting. Generally, during this phase there are very few associates that require access to the data and it is generally locked or readonly. However, the data still needs to be retained for regulatory compliance at which point it moves into the archived phase. These records generally are not planned to be retrieved again, but if the event should occur they should be easily accessible, reflect original data authenticity and accuracy, be under the control of the archivist, and easily retrievable.

One of the challenges the bioanalytical laboratory has to face is the fact that the archiving strategy for one particular system can be a very time-consuming and resource-heavy activity. Bioanalytical laboratories should follow their respective company's archiving philosophy or policy, but it is well known that not all systems are created equally. In addition, in the bioanalytical laboratory environment often solutions for instrumentation or data acquisition are not always purchased based on the ability to archive data following that policy. Generally this type of system is purchased because of the ability to acquire different types of data. Perhaps a more sensitive instrument to achieve LLOQ, or it might be the ability to reduce carryover issues with analytical methods. The system's archiving ability or format is not the most often asked question or even a requirement for a scientist who wants to solve an analytical problem. For this type of system, it often falls to the validation team to determine how to archive the data once the system has already been purchased and implemented. Therefore, it is very difficult to have one archiving format that applies to multiple systems. In addition, systems that have been around for a long time were not designed with long-term archiving as part of the design requirements. It is suggested to build the data archiving components into each study workflow as much as possible. An activity performed only once a year that takes 4 weeks can easily be postponed for one week, then another, then another. Before you know it your data will not have been archived for 18 months. It is also recommended that archiving specialists, the IT team or validation team, as appropriate, be involved in system selection even if the system is selected for analytical performance only. Data archiving formats aside, a large amount of money and resources can be wasted on analytical systems that do not meet a company's security, architecture or Part 11 requirements alone.

For each computerized system, the electronic raw data have to be defined with respect to the measured values, their meta-data, audit trail specifications and electronic signatures [3]. Meanwhile, both the capabilities and connectivity of the internet have grown at such a remarkable pace that critical laboratory applications have moved from thick-client application to

thin-client use of the web browser front ends. Vast quantities of data are being generated and stored in databases in various incompatible formats. Therefore, each type of data need their own data archiving requirements owing to the fact that the data resides in many different types of formats and many different locations. As data becomes older, generally the migration of data from obsolete systems is more cost effective then maintaining and preserving them.

Data retention periods/life cycle of data

For CROs in particular, this topic can provide hours of discussions. Often many CROs are involved in large multisite studies and data may reside all over the world. In addition, countries may have various rules regarding data retention periods. The retention period defines the minimal period of time that data must be retained and must be available for review if the safety studies that support the registration of new products or marketed products need to be verified. CROs that reside in countries that rely on reviewing study data to obtain GLP certificates also have a vested interest in maintaining study data in the CRO archives. According to the OECD, it is highly recommended that records and materials should be retained for at least three inspection cycles so that inspectors can evaluate the compliance of the test facility with the principles of GLP. For those studies that will not be submitted to regulatory authorities it may be acceptable (if justified) to dispose of the study-specific records and materials after this period [5]. The FDA states that records can be destroyed if management authorization has been obtained [1]. In addition, different countries have regulations regarding the data retention period, and it is most often subjected to the amount of time after a study is submitted to the regulatory agency for regulatory approval.

From a CRO standpoint this is a record management issue that has ongoing cost over time, long after the study has been archived. Electronic records originate for a drug during its early stages of drug development throughout the drug-approval process. The drug gets approved by the governing agency and the drug remains on the market as the innovator for a number of years. The drug-development company, in addition, runs other clinical trials for various formulations during this period. The drug is then submitted to other governing agencies for marketing and drug approval in other countries. Meanwhile the original drug-development company then gets purchased by another one. If you base the data retention period on the national regulations set up by the GLP authority in that country, the data may be at the end of the required duration of archiving and could be eligible for destruction after the original drug approval date. However, the requirements for data retention might fall under the country specifications and the retention period should be as long as the marketing-permit is held and therefore would not be eligible for destruction [6]. Therefore, it is important to maintain communication with the drug-development company and assign a record review date to data types in order to re-evaluate proactively. An example of details supporting an archiving policy is located in Table 1.

Unless the CRO is sending invoices to the original drug-development companies, inquiries as to the status of historical studies and/or drugs often go unanswered and require constant follow-up. Unfortunately, in the competitive market today archiving costs are generally not even discussed during the business development processes. There are a variety of alternatives on how to manage the data retention period for a CRO, all of which are resource heavy and require communication and commitment from the drug-development company to solve the data retention issues. This can be difficult as drug-development companies also struggle with their own data retention policies. Often it is quicker for a drug-development company to contact the CRO to retrieve data associated with an older drug, and the CRO then has the whole burden of retrieving historical (and hopefully archived) data. In addition,

Table 1. Data retention policy example.

Data type	Typical timeline	Archive timeframe	Applicable regulation	Record retention period	Risk over time	Record retention review date	Data owner	Records required for
Chromatograms	10 years from date of submission	At final report	GLP	10 years	Decreases	2 February 2013	Test facility management	Study reconstruction

drug-development companies often find themselves being acquired by other companies and resources involved with the original drug were transitioned to other projects or companies. Drug development may also be terminated anywhere along the way; this communication is rarely passed onto the CRO and is rarely in the framework of the contract that is followed up on by both the drug-development company and the CRO. As a CRO this type of communication has never been received from a drug-development company stating that: "This drug has been discontinued in the R&D pipeline, please destroy all related data".

New drugs may also sit on the R&D pipeline schedule for a number of years. A drug may then surface on the priority list or be sold to other drug-development companies without knowledge of the CRO. Long after the data retention period has passed, a request to access the data will surface and it is the expectation that the CRO be able to fulfill the request. To solve this dilemma CROs generally become data warehouses of archived data and therefore the data is retained indefinitely. This is a burden on the CRO, especially those that have been in business for a long period of time. The cost of archiving not only the paper data but the electronic data over time can be a labor and resource burden. It is also important that the archiving philosophy and procedure do not get lost with personnel changes, again another reason to re-evaluate the data and formats proactively.

Regardless, dealing and complying with these diverse requirements is difficult because there are mixtures of GLP requirements and requirements based on commercial laws with the additional complexity of the drug-development process and drug-development priorities from the drug-development companies. A clear distinction is often not possible. Setting up tracking mechanisms for all projects requires yet another electronic system (generally speaking) and requires constant supervision and followup. In the competitive CRO environment, running clinical and GLP studies is increasingly a cut-throat environment and an easy negotiating point is data archiving to reduce overall costs of running and executing the studies. CROs may choose to maintain all the data indefinitely and therefore taking on the costs of archiving the electronic data and systems can be quite cumbersome over time. CROs may work with the drug-development companies on data retention period and 'send' the data back to the drug-development company after a period of time, as described by the contract. This may again influence the archiving format of the data and can be especially complex for electronic data. In what format should the CRO provide the data back to the drug-development company? Often what happens in this situation is the data is printed and sent to the client. This may not be as useful to reconstruct the study long-term for the drug-development company and in turn reduces the value in finding long-term electronic archiving formats and solutions for the drug-development company and CRO. Is it logical, and does it provide authenticity, integrity, and usability in that format [7]? If the CRO declared the records electronic and validated the systems to that standard, are the paper records considered equivalent? Regardless, the CRO should work with the study directors to assist in determining the archive policy along with client interaction and feedback to determine when it is appropriate to return data to the client and/or destroy the records (Table 1).

Backups for disaster recovery & archives

Electronic data backups should not be confused with archived data. Regular backups of all relevant data should be done on a regular basis to protect the electronic data from disaster. This is generally done to reduce the risk to data integrity during the operational phase of the data life cycle and is used to protect data in the case of inadvertent deletion, data integrity/database corruption, or other electronic data risks. Integrity and accuracy of backup data and the ability to restore the data should be checked during validation and monitored periodically, and should not be confused with the official archived data. A classic backup application takes periodic images of active data in order to provide a method of recovering records that have been deleted or destroyed. Since the archived records are static, there is no reason to include them in periodic backups other than for disaster recovery planning/testing. Sometimes IT administrators will archive a backup session (i.e., keep monthly backup tapes for 5 years). This is not the same as an archive. A backup is designed as a short-term 'insurance policy' to facilitate disaster recovery, whereas an archive is designed to provide ongoing rapid access to decades of GLP information. Therefore, there is no need to do ongoing tests of older backup media as that is not its intended purpose. The backup schedule and

the length of time the backup material should be retainined should be defined. However, validating of testing of this system can be reduced with a proper risk assessment based on the backup schedule and backup material retained along with a proper archiving policy for the system(s). The retention requirements for backups and archives are much different. Backups should have high and large media capacity, have high-performance read/write capability and should have low storage cost for large amounts of data. The archives format should provide data authenticity, have extended lifetime capability, have high-performance random read access and have low total cost of ownership. The backup GLP data format(s), creation frequency and testing should be clearly defined by related procedures and defined by data type and location, but are only used in support of a disaster.

Archive-media dilemma

There is tension in any discussion surrounding the philosophy of digital preservation issues. It is also true that storage materials are fragile and their lifespan is unknown. In addition, storage environments can change rapidly and create a reading environment incapable of working with older materials. The fact is, digital or electronic archives cannot be left unattended (sometimes for even as a few years) and still be readable. Digital information is also more susceptible to changes in the technologies of access and retrieval. As mentioned earlier, the information is often so closely linked to the software or other technology that it cannot be used outside these proprietary environments. In addition, it often takes years to determine a company's digital archiving media philosophy, at which point technology has probably changed and the decision no longer has commercial or long-term value. Regardless, the long-term archive plan should be consistent with being able to provide assurance the data can be read for up to a certain period of years according to the data being archived and the CRO archiving policy. The archivist must be responsive to the rapid changes in technologies including hardware and software for the archiving period.

The media and/or software selected for the archives should be regularly tested for readable content preferably with automated systems (software used to restore the electronic data) and real users (who would be familiar with the content of the data). Records may be migrated from a computerized system onto a storage medium, for example, magnetic tape, diskette, CD or optical disk that can be placed in a physical archive. Archive procedures should include the consideration of additional controls for the migration of electronic records from old to new media of these records if applicable. There may be a need for special storage conditions, for example, protection from magnetic fields depending on the media selected [5]. Another media option is a storage-area device. There are a variety of different levels of redundancy available in this environment (e.g., disk arrays, tape libraries or optical drives) accessible to servers. This in addition to file systems (and associated security) can provide a limited and secure area in which to house data. These systems can be more expensive by residing on high availability devices (providing faster access to data) or on less complex systems which have lower overall cost but quick data access may be impacted.

Conclusion

There does not appear to be a single technology and philosophy that can fully satisfy the data storage requirements throughout the GLP data life cycle (operational, reference and archived) for a CRO and it must be evaluated carefully for each system. It is suggested that the CRO provides a thorough risk assessment of the data types (including non-GLP data archiving requirements) and justifies the decisions based on the following:

- GLP requirements;
- Data retention requirements;
- Customer requirements;
- Market requirements;
- Cost requirements;
- Technology requirements;
- Physical space;
- Resources (internal or external);
- Legal requirements.

An archiving philosophy must also be prioritized allowing for various types of archiving solutions. Each computerized system is then evaluated against these priorities for determination on the archiving procedure for the application. The archiving policy should be reviewed often to ensure that the priorities and requirements evaluated for the archiving policy have not significantly changed for the CRO.

Future perspective

It is the author's opinion that as electronic data continues to evolve along with the discussions and challenges of archiving electronic (and paper) data continue, there will be more concerted effort to pinpoint data retention periods and data archiving formats across systems. Scientific systems have had a propensity to be years behind actual regulation and/or discussion in this area, but will be required to comply as customers become more familiar and able to quantify costs of archiving long term. As media for archiving advances with technology it would be expected that costs and alternatives for storage space in the 'computing cloud' will also be more widely available and accepted. Electronic archiving systems will also adjust to help solve the archiving challenges in a bioanalytical laboratory by becoming more cost effective, providing a larger return on investment, and force a more generic or industry-accepted format.

Financial & competing interests disclosure

The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Bibliography

1 US FDA. Guidance for Industry Part 11, Electronic Records, Electronic Signatures – Scope and Application. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (2003).

2 The Organisation for Economic Co-operation and Development. *OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 10.* Paris, France (1995).

3 Federal Office of Public Health (AGIT), Working Group on Information Technology. *Guidelines for the Archiving of Electronic Raw Data in a GLP Environment.* 2003.

4 OECD. *OECD Consensus document: the application of the principles of GLP to computerized systems (OECD Series on Principles of GLP and Compliance).* Paris (1995).

5 The Organisation for Economic Co-operation and Development. *OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 15. Establishment and Control of Archives that Operate in Compliance with the Principles of GLP.* Paris, France (2007).

6 *Code of Federal Regulations Title 40, Part 160, EPA 40 CFR 160.195 – Retention of Records.* s.l.: Enviornmental Protection Agency (1983).

7 GAMP. *GAMP Good Practice Guide: Electronic Data Archiving.* s.l. : ISPE, 2007.

Websites

101 ARMA International. Electronic records and e-discovery. www.arma.org/erecords/index.cfm

102 Mullins C. The data administration newsletter (2006). www.tdan.com/view-articles/4591