DS DASSAULT SYSTEMES

DS BIOVIA

ELECTRONIC ARCHIVES
FOR REGULATORY RETENTION COMPLIANCE IN LIFE SCIENCE INDUSTRIES AND CLINICAL TRIAL RESEARCH
White Paper By Teri Stokes, Ph.D.

ABSTRACT

"Digital born data" is found everywhere in regulated Good Practice environments for laboratory (GLP), clinical (GCP), and manufacturing (GMP) purposes, e.g., (GXP). Paper documents are also scanned into digital format (PDF) for storage and retention. This document explains the fallacy of thinking that system backup tapes suffice for electronic archive purposes. It then describes the characteristics of electronic archives and discusses the history and concepts supporting today's practical experience with digital preservation for long term retention compliance in organizations subject to US FDA, EU EMEA, and other global authorities.

INTRODUCTION

Today, the pharmaceutical, biotech, and medical device industries create and operate in a sea of electronic information. Much of this information is now "digital born data" which means it is created and recorded in electronic form at its source, e.g., as instrument output, electronic documents, spreadsheets, graphics, video, etc. It never goes through a paper phase. In addition a lot of paper information is being scanned into PDF format for ease of electronic sharing and storage. The need for electronic archive capacity has gone from kilobytes (KB) to megabytes (MB) to gigabytes (GB) to Terabytes (TB) to Petabytes (PB) and continues to Exabytes (EB) and beyond.

Storage Unit	Storage Capacity Equivalent
2 KB	Typewritten page
50 KB	Image of a document page, compressed
2 MB	Photograph, high resolution
5 MB	Complete works of Shakespeare; 30 seconds of broadcast-quality video
10 MB	Digital chest x-ray
50 MB	Digital mammogram
1 GB	Paper in the bed of a pickup; symphony in high fidelity sound; broadcast-quality movie
2 GB	20 yards of books on a shelf
1 TB	All x-ray films in a large technical hospital; 50,000 trees made into paper & printed
2 TB	Academic research library
2 PT	All US academic research libraries
5 EB	All words ever spoken by human beings

1 MB = 1024 KB; 1 GB = 1024 MB; 1 TB = 1024 GB; 1PB = 1024 TB; 1 EB = 1024 PT

Ref. James S. Huggins' Refrigerator Door. http://www.jamesshuggins.com/i/bas1/0_top_bar.jpg

SYSTEM BACKUP MEDIA VS. ELECTRONIC ARCHIVES

System level backups are performed daily, weekly, and monthly. They are indexed by server system identifier and date/time of backup. They contain software applications, data records, and database files for that specific date/time. Backup media is not indexed for the content and context of the files and records it contains. The focus for offsite backup media is to allow retrieval of a whole operational environment after a system failure. It can also be used to retrieve prior versions of lost or corrupt database files.

Trying to find specific data records on backup media requires knowing who was the record originator, what application and database were used and on what date it was created. Otherwise it is a "needle in haystack" archeological experience to search for a specific record on backup media from past months or years. This raises practical concerns for compliance to long term retention of data required by many authorities.

An electronic archive is a dedicated computerized system that uses a specific archive software application to index records deposited across time. An electronic archive is designed to operate like a reference library for long term retention and retrieval of electronic information that has enduring value to the organization. The archive application focus is for secure retrieval by searchable project and record level content to support the following:

- Protect and retrieve GXP records and data in context for reconstructing regulated actions and studies in audits and inspections.
- Protect and retrieve prior database files supporting the safety, efficacy, and quality of regulated products.
- Defend intellectual property rights and/or repurpose research data for new indications.

An electronic archive is a dedicated computerized system that uses a specific archive software application to index records deposited across time.

To accomplish its focus, an electronic archive is managed by an archivist and functions to a standard archiving business process. This is the 21st century alternative to the traditional paper archives process. In 2007 Paris, the OECD guidance for GLP archives stated the following: (2)

> "Electronic records are at risk without a preservation process to ensure that these records are available in the future. Procedures should be in place to ensure that essential information remains complete and retrievable throughout the specified retention period."

This OECD reference provides a clear discussion of general good archiving principles for regulated purposes. Section 8 discusses archiving electronic records specifically.

Archive Principles and Archivist Responsibilities

In 2006, the Society of American Archivists published the book by O'Toole and Cox titled Understanding Archives and Manuscripts. These authors described the three fundamental principles of archiving practice to be –

1. **Provenance**: a record creator determines the record content to be archived

2. **Original Order**: records come into existence with an order already imposed by the creator and the time of record creation

3. **Possible Uses**: how can the records be used based on the context of provenance and original order

Given the context of the principles above, it is the archivist's responsibility to establish an archive that defines the following: (3)

- Where and how records are kept, organized, and **made available**

- What records are kept and what records are **not kept** and why

- How records are maintained and media conversion is performed for **long term retrieval and readability**

O'Toole and Cox gave some clear advice on how to establish a successful archive. In planning an archive they said there must be a clear mission statement defining the policy and purpose for the Archive, the archivist role, and the records to be archived across all departments. This statement of policy and purpose must be fully endorsed by the parent organization and a clear reporting line to the organization's Chief Administrative Officer must be described.

The archive staff and services need to be structured so that measurable user results can be tracked, quantified, and reported. What number of records were deposited in a given timeframe, number of user requests for retrieval of archived information? How long did it take to review and prepare an average deposit package for different types of data records? What is the breadth and depth of the information catalog for available information of high priority to the organization? How many users were trained as record depositors and as record viewers, etc.? A successful archive uses standard service business metrics to show its contribution to the organization in fulfillment of its mission statement. What audits or inspections has the archive supported? What projects have repurposed data from archived records? What records have been decommissioned and destroyed from the archive?

The Good, the Bad, and the Ugly of Electronic Archiving

Jean-Luc Chatelain and Daniel B. Garrie wrote an article with the above title for the Journal of Legal Technology Risk Management in Vol. 2, 2007. In it they described three key differences for archiving an electronic document from archiving a paper document. (4)

1. **Scale**: the huge numbers of electronic documents add to archiving cost and complexity

2. **Mutability**: electronic documents are easily modified without leaving an immediately visible trace

3. **Readability**: it is hard to guarantee that an electronic document format will always be able to be decoded with content and format readable over the years

> The archive staff and services need to be structured so that measurable user results can be tracked, quantified, and reported.

Chatelain and Garrie identify <u>the Good thing</u> about electronic archives is that they can be used to keep corporate records secure and retrievable. They state that <u>the Bad thing</u> occurs when silo thinking in organizations result in archiving silos with duplicate or redundant information and no uniform retention management practices. They describe <u>the Ugly thing</u> as a high cost of ownership due to growing scale over time.

To counteract some of the bad and ugly things, Chatelain and Garrie gave some implementation advice for establishing an electronic archive. They advise that Step one is to have executive sponsorship committed to the success of the archive. Step two is to form a cross-functional team to include all relevant business areas, plus legal, regulatory affairs, IT and Quality Assurance. Step three is to use the cross-functional team to define organization-wide requirements for content indexing and classification, query performance at scale, etc. Given the growth of cloud storage services in the seven years since the writing of the article, the high cost of ownership has lowered per megabyte, but the data avalanche has continued to grow to new heights in every organization.

NASA's Model for Archive Systems

For many years the National Aeronautics and Space Administration (NASA) of the United States has been dealing with the challenge of securely storing and using enormous amounts of electronic information from its space missions. NASA has established a practical model for its archive system called the Open Archival Information System (OAIS). This model beaks an archiving system into six functions and has now been incorporated into the ISO Standard 14721:2003 as shown below.

1. **Ingest** – accept information and prepare it for storage

2. **Archival Storage** – pass information to repository and maintain until requested for retrieval

3. **Data Management** – populate, maintain, and access

    a. Descriptive information about archive holdings

    b. Administrative data to manage the archives

4. **Administration** – provide services for overall operation of the digital archive

5. **Preservation Planning** – system monitors environment and recommends actions to keep records accessible as hardware and software become obsolete

6. **Access** – system allows users to

    a. Determine the existence, description, location, and availability of information stored in the OAIS

   b. Request and receive information products stored

This NASA model and the ISO standard could provide an organized starting point for the cross-functional team suggested by Chatelain and Garrie to use in their planning phase.

This NASA model and the ISO standard could provide an organized starting point for the cross-functional team suggested by Chatelain and Garrie to use in their planning phase.

NARA's Electronic Records Archive (ERA)

The US National Archives and Records Administration (NARA) in the Washington, DC area has a "system of systems" with multiple components that perform different archival functions and that manage records governed by different legal frameworks. It is called the Electronic Records Archives (ERA) and it contains the permanent records of all the US federal agencies. Its major functions are pictured to the right.
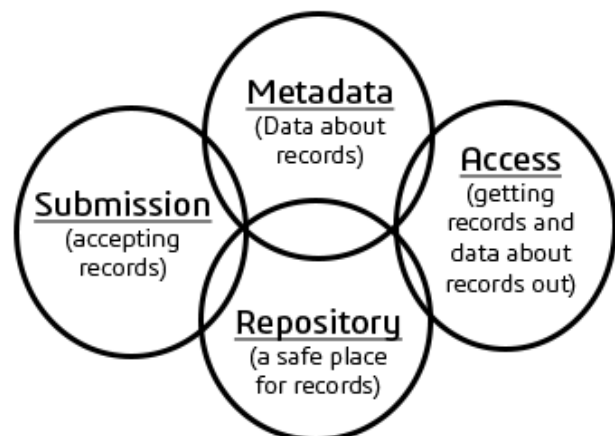
Government agencies use the **Submission** function to deliver records and metadata into ERA. Electronic records are preserved and reviewed in the **Repository**. The National Archives' knowledge of what those historically valuable records are, who created them and why, and what processes have been applied to them is all documented in the **Metadata**. Finally, the public uses the **Access** function to do research on records that have been screened by archivists to remove restricted information (there are separate, non-public kinds of access for restricted records, as well.) (5)

Other National Initiatives & Open Source Software

Other national libraries and archives programs include the PRONOM program in the UK, the VERS in Australia, and the Planets project in the Netherlands. These national programs have developed archiving tools and experience that is in the public domain for the benefit of others coming into the electronic archives arena.

Free open source software is also available for long term archiving of huge numbers of files and growing masses of data. OpenArchive can be used to make available petabytes of data – up to 100 billion files on hard drives and tape – from email, files, audio, video, PACS, engineering and other applications. It is the open source version of the commercial product Archive Manager by Grau Data AG. (6)

OECD GLP Archives Guidance (7)

This 2007 guidance document provides a valuable introduction for people who are new to the topic of archiving for regulated purposes. Most of its content is applicable to archiving for any of the GXP regulations and for both physical and electronic archives. It describes three archive roles:

1. **Archivist**: Person responsible for archive management, e.g., for operations and procedures in accordance with established company policies, standard operating procedures (SOPs), and relevant regulations.

2. **Archive Staff**: People working under archivist's supervision to perform routine archive operations.

3. **Third Party**: External archive services or IT support with archival responsibilities defined in a contract.

In discussing the delegation of technology archiving tasks for electronic records management, the OECD states, "Respective tasks, duties, and responsibilities have to be specified and detailed in SOPs." (7) (p.12)

The guidance describes the responsibilities of the archivist as follows:

- Ensure that access to the archive is controlled.

- Ensure that the orderly storage and retrieval of records and materials is facilitated by a system of indexing.

- Ensure that movement of records and materials in and out of the archives is properly controlled and documented.

This last bullet refers to the concept called "Chain of Custody" where logs are kept for who touched what record, when, why, and with required approvals as per SOP. This traceability becomes very important in long term retention when defending intellectual property rights or during product quality disputes.

OpenArchive can be used to make available petabytes of data – up to 100 billion files on hard drives and tape – from email, files, audio, video, PACS, engineering and other applications.

The OECD recommends that SOPs for GLP regulated archives should address the following issues, where applicable:

Access to the archives	Definition and description of the archive
Indexing procedures, including electronic records	Conditions under which records and materials should be stored
Procedures for the receipt of records and materials to be archived	Procedures for accessing, removal, and return of records and materials
Responsibilities of the archivist and archiving staff	Security of the archive facility and the records and materials retained
Climate control	Retention period
Disposal of archive records and materials	Contract archiving services, if applicable
Disaster recovery	Training requirements for the archivist and archiving staff
Frequency of archiving non-study specific records	Periodic refreshing of electronic records

Ref. OECD GLP Archive Guidance (7) p. 15

GAMP Guide Electronic Archiving Strategy Template

Another very helpful resource of practical support for developing an electronic archive in regulated situations is the *GAMP Good Practice Guide for Electronic Data Archiving* published by the International Society of Pharmaceutical Engineering (ISPE) in 2007. Appendix G of this volume provides a template and tutorial for creating an Archive Strategy document. The sections and topics for such a strategy document are shown below. (8)

Thinking through and documenting all sections of this template by a cross-functional team is an essential first step before creating a request for proposal or user requirements specification for any electronic archive system.

Access to the archives	Definition and description of the archive
1. Introduction	Clearly state the authority establishing the archive, summarize the strategy and put it in context with company and external standards and regulations.
2. Purpose	To identify and define the boundaries and key requirements of electronic data archiving to guide implementation and maintenance of e-data archival solutions to ensure continued compliance of archives records in the future.
3. Scope	Identify key dimensions and define boundaries of strategy. Consider physical location(s); organizational unit(s); data sources; data owners; data types; data content, regulatory/legal requirements.
4. Role & Responsibilities	Identify archiving activities and assign who is responsible for each: original data owner; archived data owner; archive administrator; archive owner; technology owner; Quality Roles; Management; Sponsor of third party work.
5. Archive Content Requirements	Define archive processes; identify supported data and any limitations; describe required metadata (data identity, type, content, source, context, data owner, access profile, retention period); identify quality safeguards for a record deletion process; define an exit strategy for end of tech support; define how data and metadata is to be migrated from one platform to another.
6. System Requirements	Specify the broad technology platform; identify key data and user interfaces to/from the archive; describe location/environment and key security requirements; define a core set of required operations, describe a test environment for initial and ongoing changes to the archive system software.
7. Compliance Requirements	Define a risk assessment process; specify applicable predicate rules; describe how to achieve and maintain compliance; describe how to validate the archive; define archive's performance verification criteria; address change management and business continuity; define SOPs to be written; define how archive will be maintained and evaluated.

This documented archive strategy approach clearly addresses the fundamental success criteria referred to earlier by Jean-Luc Chatelain and Daniel B. Garrie in their Good, Bad, and Ugly journal article. Thinking through and documenting all sections of this template by a cross-functional team is an essential first step before creating a request for proposal or user requirements specification for any electronic archive system.

For section 7 of the template there are two predicate rules related to electronic archives:

- 21 CFR Part 11.10 Electronic Records (c): Controls for protection of records to enable their accurate and ready retrieval throughout the record retention period. (FDA 1997)

- EU GMP Guide Annex 11 Computerized Systems:

  a. 7.1   ...Access to data should be ensured throughout the retention period.

  b. 17. Archiving – Data may be archived. This data should be checked for accessibility, readability, and integrity. (EU 2011)

SUMMARY

The regulated industries have filled mountains with paper born records that have been carefully archived over the last century. Now in this century we are beginning to address the transition to archival of digital born data and digitized paper records. Advances in storage and cloud computer technology have made it practical to retain larger amounts of records for less cost. The challenge is to establish the electronic archive function with the same executive commitment, corporate focus, and appropriate organizational structure as previously given to the paper archives being replaced. This transition starts with using an industry standard electronic document management system (eDMS) that communicates with an archival function or can be equipped to communicate with separate archival technology. The future for GXP electronic archives is NOW.

About The Author

Dr. Teri Stokes, Ph.D., Director GXP International

Concord, MA, USA

https://gxpinternational.com/contact/

Book Title: Computer Validation: A Common Sense Guide ISBN: 1-93372253-3

Hard copy available at **Amazon.com**

Access a PDF version of two chapters highlighted in whitepaper from **GXPInternational.com**:

- Chapter 9 - The Archivist Role for Electronic Archives, Appendix 9.1

- Chapter 10 - The Corporate Role for Electronic Records Quality, Appendix 10.1-10.4

About BIOVIA

Dassault Systèmes Biovia Corp. (BIOVIA) provides global, collaborative product lifecycle experiences to transform scientific innovation. Powered by Dassault Systèmes' 3DEXPERIENCE® Platform, BIOVIA solutions create an unmatched scientific management environment that can help science- and processdriven companies create and connect biological, chemical, and material innovations to improve the way we live. The industryleading BIOVIA portfolio integrates the diversity of science, experimental processes and information requirements across research, development, QA/QC and manufacturing. Capabilities include scientific data management; biological, chemical, and materials modeling and simulation; open collaborative discovery; scientific pipelining; enterprise laboratory management; enterprise quality management; environmental health and safety; and operations intelligence. BIOVIA solutions are used by more than 2,000 companies in the pharmaceutical,biotechnology, energy, chemicals, aerospace, consumer packaged goods and industrial products industries, as well as academic and government entities.

References

1. James S. Huggins, James S. *Huggins' Refrigerator Door*. 2014 URL: http://www.jamesshuggins.com/i/bas1/0_top_bar.jpg

2. OECD, *Series No. 15. Establishment & Control of Archives that Operate in Compliance with GLP Principles,* Organization for Economic Cooperation and Development (OECD), Paris 2007.

3. James M. O'Toole & Richard J. Cox, *Understanding Archives & Manuscripts,* Society of American Archivists, 2006.

4. Jean-luc Chatelain, Daniel B. Garrie, *The Good, the Bad, and the Ugly of Electronic Archiving,* Journal of Legal Technology Risk Management, vol./ 2, Fall 2007.

5. Nara's ERA website at 2014 URL: www.archives.gov/records-mgmt/era/

6. OpenArchive open source website at 2014 URL: www.openarchive.net/Start

7. OECD website GLP Archives Guidance free download at 2014 URL: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)10

8. ISPE. *GAMP Good Practice Guide: Electronic Data Archiving,* International Society for Pharmaceutical Engineering, 2007. pp. 114-125. 2014 URL: www.ispe.org

Our **3D**EXPERIENCE Platform powers our brand applications, serving 12 industries, and provides a rich portfolio of industry solution experiences.

Dassault Systèmes, the **3D**EXPERIENCE Company, provides business and people with virtual universes to imagine sustainable innovations. Its world-leading solutions transform the way products are designed, produced, and supported. Dassault Systèmes' collaborative solutions foster social innovation, expanding possibilities for the virtual world to improve the real world. The group brings value to over 170,000 customers of all sizes in all industries in more than 140 countries. For more information, visit **www.3ds.com**.

3DEXPERIENCE®

Dassault Systèmes | The **3D**EXPERIENCE® Company