



GAMP Good Practice Guide:

Electronic Data Archiving



Documents

Document Sharing Hub

Preface to the GAMP Good Practice Guide: Electronic Data Archiving

This document, the GAMP® Good Practice Guide for Electronic Data Archiving is intended as a supplement to the Guide to Validation of Automated Systems (GAMP Guide). It is recommended that this Guide be read in conjunction with the main GAMP Guide and the ISPE/GAMP® Good Practice Guide for Electronic Records and Signatures.

It seeks to provide a rational and scaleable approach to electronic data archiving through the development of an archiving strategy. The implementation of this strategy should help organizations to achieve and maintain regulatory compliance, and to more effectively manage electronic records over the long term.”

Disclaimer/Limitation of Liability

*While every effort has been made to ensure accuracy at time of publication, neither ISPE nor the GAMP COP can be held responsible for any errors or omissions in the text. **In no event shall ISPE or any of its affiliates (including the GAMP COP), or the officers, directors, employees, members, or agents of each of them, be liable for any damages of any kind, including without limitation any special, incidental, indirect, or consequential damages, whether or not advised of the possibility of such damages, and on any theory of liability whatsoever, arising out of or in connection with the use of this information.** Readers are strongly recommended to refer to current regulations when reviewing and deciding on policies, procedures, and processes.*

© Copyright ISPE 2007.

All rights reserved. No part of this document may be reproduced or copied in any form or by any means – graphic, electronic, or mechanical, including photocopying, taping, or information storage and retrieval systems – without written permission of ISPE.

All trademarks used are acknowledged.

Acknowledgements

The EDA SIG has been led by Per Olsson, ABB, who also has led the development of this Guide. A number of individuals have provided input to the work of the SIG, ranging from researching source information, providing information, attending SIG meetings, authoring sections of the Guide, reviewing text, editing text, etc. All contributions are gratefully acknowledged.

The following individuals constitute the ‘EDA SIG Core Group’ and have provided a significant input to the SIG over an extended period of time.

Andy Bearpark	Eli Lilly & Co.
Nick Blundell	AstraZeneca R&D
Anne Broadbent	AstraZeneca R&D
Peter Werner Christensen	Novo Nordisk Engineering A/S
John Dickson	Thermo Electron Corporation
John Farrell	The Synapse Partnership Ltd.
Philip Lord	The Digital Archiving Consultancy
Alison Macdonald	The Digital Archiving Consultancy
Steve Morrell	ABB Ltd.
Marian Mutch	Covance Inc.
Per Olsson	ABB Ltd.
Gary Rendle	VSS Ltd.
Bala Sreenivasan	Schering-Plough
Siegfried Schmitt	GE Healthcare
Frank Woods	GlaxoSmithKline R&D
Ann Wrightson	CSW Group Ltd.

Additionally, the following individuals have provided measurable and in some cases, substantial input to the work of the SIG.

Kate Baker	Daiichi Pharmaceuticals UK Ltd.
Stephen Boreham	ABB Ltd. (retired)
Sam Brooks	Novartis AG
Arthur Charlesworth	ABB Ltd.
Svend Martin Fransen	Novo Nordisk A/S
Mike Heyes	ABB Ltd.
Stephen Håkansson	GlaxoSmithKline R&D
Ewan Jones	Industrial Technology Systems Ltd.
Piera Lukes	IT's QA Consulting
Terry Thompson	Covance Laboratories (retired)
Wayne Lewis	MediSense UK Ltd.

All used sources are acknowledged and in particular those named in Appendices A, C, and J. Appendix B is based on the Open Archival Information System (OAIS) reference model as provided by Consultative Committee for Space Data Systems (CCSDS).

Table of Contents

1	Introduction	5
1.1	Overview	5
1.2	Purpose	5
1.3	Scope	6
1.4	Benefits	7
1.5	Objectives for this Guide	7
1.6	Key Concepts	8
1.7	Structure of this Guide	9
2	Overview of Archiving	10
2.1	Background to Archiving	10
2.2	Paper and Electronic Archiving	10
2.3	Driving Forces to Archive	11
2.4	Options for Archiving	11
2.5	Electronic Data Archive System Lifecycle	12
2.6	Main Archiving Processes	15
3	Creating and Implementing an Archiving Strategy	18
3.1	Creating and Implementing an Archive Strategy	18
3.2	Step 1 – Define Scope	19
3.3	Step 2 – Define Requirements	20
3.4	Step 3 – Investigate Available Solutions	30
3.5	Step 4 – Document Findings	32
3.6	Step 5 – Implement Strategy	33
4	Considerations When Archiving	35
4.1	Summary of Archiving Issues	35
4.2	Migration of Records/Data	38
4.3	Deletion of Records/Data	39
4.4	Exit Strategy	40
4.5	Retirement of the Archive System	41
4.6	Special Considerations	42
5	Appendix A: Example Regulatory Requirements	48
5.1	Specific Regulatory Topics	48
5.2	Regulatory References	55
6	Appendix B: The OAIS Reference Model	74
6.1	The OAIS Reference Model	74
6.2	Overview of Processes	78
7	Appendix C: Metadata	88
7.1	Overview	88
7.2	Kinds of Metadata Standards	89
7.3	Example of Metadata	90

8	Appendix D: Databases	92
8.1	Overview	92
8.2	Data Retention in the Production System	92
8.3	Reprocessing, Readability, and Information Loss	92
8.4	OLAP/Data Warehousing Technology	93
8.5	Other (Non-OLAP) Database Archiving Strategies	94
9	Appendix E: Laboratory Systems	100
9.1	Overview	100
9.2	Commercial Scientific Data Management Solutions (SDMS)	100
9.3	Conversion to a Technology/Vendor-Neutral Format	101
10	Appendix F: Archiving Approaches	104
10.1	Increase On-line Storage Capacity	104
10.2	Mothball/Time Capsule/Computer Museum	105
10.3	Format Migration	106
10.4	Archive to Standard Formats	106
10.5	Emulation and Virtual Computers	107
10.6	Transfer to Non-Electronic Media	108
10.7	Storage Media Considerations	109
10.8	Archiving Approaches Summary	110
11	Appendix G: Archiving Strategy Template	114
11.1	Introduction	114
11.2	Purpose	114
11.3	Scope	114
11.4	Roles and Responsibilities	115
11.5	Archive Content Requirements	116
11.6	System Requirements	120
11.7	Compliance Requirements	122
12	Appendix H: Good Practices	128
13	Appendix I: Glossary	132
13.1	Definitions	132
13.2	Acronyms and Abbreviations	138
14	Appendix J: References	146

1 Introduction

1.1 Overview

The subject of Electronic Data Archiving (EDA) is both large and complex, and one that is rapidly expanding and evolving.

Regulators are increasingly favoring the submission of information in electronic form. The cost of a failed inspection can be substantial, and an EDA mechanism is an important tool in ensuring continued compliance. This, together with the increasing use of computers and electronic data, has resulted in the need for compliant solutions for electronic data archiving.

In August 2003, the FDA issued a guidance document named “Part 11, Electronic Records; Electronic Signatures – Scope and Application.” This document contains brief guidance on record retention, reminding readers of the requirement to retain accessibility to records throughout the retention period and allowing the archiving onto non-electronic media, as long as predicate rules are met and the content and meaning of the record is preserved. At the same time, the PIC/S issued its guidance “Good Practices for Computerised Systems in Regulated GxP Environments,” which also reminded readers of the requirements to retain records in accessible form, and stressed the media independence of regulatory responsibilities with regard to records and record retention.

At the time of publication, it is generally accepted that there are no commercial off-the-shelf solutions that can comprehensively meet all the demands of the regulated life science sector, and meet security and confidentiality criteria over a number of software and technology platform upgrades and changes. Although such commercial solutions are in development, it is unlikely that future archiving needs can be fully addressed using off-the-shelf offerings.

The challenge is to plan and implement changes to processes and systems to enable compliant solutions to the regulatory requirements.

1.2 Purpose

This Guide is intended as a supplement to the main GAMP® Guide, which provides an introduction to record retention, archiving, and retrieval. It has been developed in accordance with the version of the main GAMP® Guide current at time of publication, but every effort has been made to align with planned revisions of that document. It is recommended that this Guide is read in conjunction with the main GAMP® Guide and the ISPE/GAMP® Good Practice Guide for Electronic Records and Signatures, which provide additional relevant information.

It seeks to provide a rational and scaleable approach to electronic data archiving through the development of an archiving strategy. The implementation of this strategy should help organizations to achieve and maintain regulatory compliance, and to more effectively manage electronic records over the long term.

More specifically, this Guide:

- Provides an introduction to the complex subject of electronic data archiving, recognizing the differences from the traditional paper archive
- Provides a process for creating and implementing an archiving strategy
- Highlights considerations in determining an archiving strategy, at an organizational, technical, and regulatory level

- Identifies those aspects of technology that have an impact on the selection of an archive solution, which are independent of specific technical solutions

The intended audience for this Guide consists of those individuals who:

- Need to archive electronic records (particularly regulated records)
- Have accountability for stored electronic records, including the Archivist or Archive Administrator, and those with management responsibility for providing records to both regulators and business users
- Need to access archived electronic records
- Are tasked with implementing an electronic archive (this includes IS/IT specialists and suppliers) and are responsible for the funding of the electronic archive

Therefore, the audience for this Guide will include individuals from many disciplines, as well as IS/IT specialists.

1.3 Scope

This Guide addresses the processes and issues around the long term preservation of electronic data. These include:

- Transfer of data from a live, on-line system to electronic archive storage
- Retrieval of data from archive
- Deletion of data
- Maintenance of an electronic archive system as it approaches obsolescence

Current GxP regulations related to archiving have been taken into account in developing this Guide. The Guide addresses management, planning, development, operational, and compliance issues. Good systems management and other matters covered by the main GAMP Guide are excluded.

The Guide does not cover the digitizing of paper records and the operation of a conventional paper archive. General rules about archiving and the role of the traditional Archivist also have been excluded since these are well-established concepts. Specific technical solutions and research material and concepts are not covered.

E-Mail

While the principles outlined in this document may apply to the retention of email records, their use and retention has not been specifically covered. Some of the issues relating to emails include:

- What constitutes appropriate use of email?
- When should email be considered a business record?
- Difficulties associated with retention of encoded electronic signatures and metadata

It is considered that reliance on emails as records exclusively for GxP purposes is now, and likely to remain, limited. Other regulatory requirements (e.g., Data Protection, Sarbanes-Oxley, Stock Market regulations) which go beyond the scope of GxP and differ from jurisdiction to jurisdiction, apply to email communications and should be taken

into account when considering the retention, destruction, copying, and forwarding of email communications. Therefore, comprehensive guidance in this matter is beyond the scope of this Guide. Regulated companies developing archives of such data should take specific and explicit legal advice concerning the management of those archives.

Web Sites

Given their special nature, Web sites are out of scope. The archiving and preservation of Web sites is difficult, as these often are dynamic in nature and do not have a final form. They are often temporary and they may contain links to other Web sites that are similarly temporary.

Data Warehouses

Data warehouses that are used for data mining purposes, and often contain copies of data held in archives, are not discussed in detail in this Guide, as their main purpose is commercial and their use often is not regulated. However, some of the issues surrounding data warehouses are covered in Appendix D of this Guide.

1.4 Benefits

This Guide aims to assist in the efficient, effective, and compliant development and implementation of electronic data archiving.

Particular benefits of such an implementation include:

- An enhancement of the asset management and retention of Intellectual Property, which represent a significant financial resource to the organization
- Improved business processes through the use of effective data access, enabling data and information reuse. These may assist in shortening the time-to-market for new products.
- Improved regulatory and legal compliance through enhanced search and retrieval facilities, traceability, integrity, and security of records
- Streamlined business processes for the retention of records and their maintenance, based on the use of automated processes
- Reduced business risk through the controlled use of advanced electronic tools for the safe-keeping of records

1.5 Objectives for this Guide

Methods by which electronic data archiving is conducted should be practical and efficient, in addition to meeting regulatory compliance expectations.

To this end, the following objectives applied to the development of this Guide:

- Provide a common understanding and awareness of the issues surrounding electronic data archiving
- Establish the principles of good electronic data archiving practice, supported within the framework of the main GAMP® Guide

- Provide a framework for the preservation of electronic records and data
- Assist the life science sector in reducing the regulatory, legal, and business risks associated with electronic data archiving

1.6 Key Concepts

This Guide covers a number of concepts, issues, and practices as they apply to the complex and developing subject of electronic data archiving.

This section highlights the fundamental role of the Archiving Strategy document and the key terminology used.

1.6.1 Archiving Strategy Document

The creation of an Archiving Strategy is central to this Guide. The Archiving Strategy sets out how the differing requirements will be defined and met, i.e., the approach for dealing with all identified archiving issues in a compliant manner.

The Archiving Strategy document can be applied to an organization, site, department, or an individual EDA.

The created Archiving Strategy should be practical and to the extent possible, based on available solutions. It should provide a concrete, practical strategy to reliably meet well-understood regulatory requirements.

The Archiving Strategy is similar to a high level Requirements Specification (RS), but with certain information missing, such as capacities (number of users, storage capacity, speed of response, etc.). The Archiving Strategy document should define the key principles and concepts for an EDA to be developed within scope, and form the basis of each RS for the particular EDAs to be implemented. Requirements that are particular to a specific EDA often are best contained in the relevant RS and not the Archiving Strategy document, unless the requirements are deemed to be of strategic importance.

Detailing the requirements in the Archiving Strategy will help to protect the EDA from future changes. The Archiving Strategy identifies and clarifies the key assumptions and requirements so that these can be taken into account when contemplating changes with a potential adverse affect on the EDA.

1.6.2 Key Terminology

The following key terms are used throughout this Guide.

Archive

Archiving is a formal process of taking a record off-line by moving it to a different location or system, often disabling it from any further changes. Protection from further updates is particularly important for the life science sector, as there is a danger that archived data can be improperly re-used for a regulated activity. The archived record is deemed to be the master record.

Backup

Backups are copies of records, created on a regular basis as security against loss. The backup copy of a record is just that. The original record that resides on the system is deemed to be the master record (the form of the record to be relied upon in the performance of a regulated function) until such time as it is replaced by the backup copy following loss.

Data

Data is, in archiving terms, the smallest piece of information that is handled. Data can represent content, information about context, or contain structural information, etc. A single datum is unlikely to convey any meaning. Data relating to the context and structure of a particular datum or data, 'data about the data', is often referred to as metadata.

Record

Record is, in archiving terms, the smallest collection of data that conveys content, context, and structure.

Retention

Retention is maintaining records in a secure, accessible, and reliable form for a period of time, as set in regulations or other mandate. The requirements for data retention usually are defined by Regulatory Authorities with the associated retention period (the length of time specified for data on a data medium to be preserved) varying from a few to many years.

Storage

Storage is the process of keeping an electronic record on a given medium, e.g., magnetic tape, CD, or hard disk. Therefore, this term applies to on-line records, as well as archived records and backup copies.

Retaining records is not the same as retaining data. A record is essentially a collection of data presented in a particular manner. The same data could be presented or related in a number of ways and convey completely different information. In a computer system, Users see the data expressed as records or information. They would have little or no control over how to select or arrange the data, and are constrained to viewing records rather than random fields from a database. Hence, the meaning and context of the data is protected. One of the major challenges associated with record retention is that of retaining the original meaning or context of the record outside of the original application.

1.7 Structure of this Guide

This Guide consists of a main body and a set of supporting appendices.

The main body consists of:

- Introduction
- Overview of archiving
- Creating and implementing an archive strategy
- Considerations when archiving

2 Overview of Archiving

This section covers:

- Background to archiving
- Paper and electronic archiving
- Driving forces to archive
- Options for archiving
- EDA system lifecycle
- Main archiving processes

2.1 Background to Archiving

Data is being generated at a huge rate, and much is of significant potential value. The collection, organization, storage, and retrieval of information are central to the operation of the life science sector. The archive has a central function in the modern business model. The role of the traditional Archivist is moving to that of Knowledge Manager.

Much of the information generated today is in digital format. An archiving strategy for electronic records is an important tool in addressing the need for adequate methods for the long-term preservation of this digital information.

2.2 Paper and Electronic Archiving

The development of the electronic archive is at a very early stage compared with the well-established paper archive.

Early electronic archives were largely replicating what had been learned from their paper equivalent. Many early electronic records were stored as paper records or as individual magnetic tapes, as you would store a book. Much of the content of these archives is now lost, as the solutions developed for the paper archive have proved inadequate for the electronic record.

Initially, much raw data was generated in paper form and only subsequently translated into electronic form. However, this is changing rapidly, because the majority of data is now generated electronically with no equivalent paper record. Furthermore, thanks to its commercial and business advantages, electronic archiving is increasingly used for the storage of non-electronic data: this data is migrated into electronic format and the original data deleted. These trends impose the urgent need to develop robust electronic solutions for the electronic archive.

There are some fundamental differences between the paper and the electronic record and the required attributes of the archive:

- Electronic representations of data are subject to frequent changes or updates and can become obsolete within a few years with no guarantee of proven migration paths. This can occur at the level of the media, the storage format, metadata needed for interpretation, or even the character sets used. Written languages, on the other hand, generally evolve slowly, and dictionaries record words and idioms, and translate between languages and formats.

- Computer technology is developing rapidly. Paper technology and archiving methodologies, on the other hand, are well-established and stable.
- The electronic archive and access to its records require the use of computer technology. The archived paper record can be read without any such tools.
- The electronic record is often not self-contained, i.e., to understand its meaning and context requires additional resources, such as metadata, orientation memoranda, databases, or laboratory files. Archived paper documents are frequently understandable without the need for other resources.

The confirmation for the electronic archive design is to enable the long-term retention and controlled accessibility of information (as opposed to records and data) over many computer technology and platform changes.

2.3 Driving Forces to Archive

The principal drives for archiving electronic records include business reasons and regulatory demands. An archive is central to being able to retain and reuse knowledge. This knowledge may be used to:

- Further the business
- Enhance performance
- Save costs by preventing repetition of work
- Facilitate business processes such as mergers and acquisitions
- Reduce risks by applying previous learning

Regulatory demands, as expressed by local, regional, and central governmental bodies, as well as business sector organizations, often require the retention of large amounts of data. This extends into all aspects of business life, such as employment legislation, environmental controls, and tax law. The ability to safely locate this information in a timely manner is becoming increasingly important, as the financial penalties and cost to reputation can be severe.

Underpinning both of these drivers is the need to keep data securely accessible, which a well-designed and managed archive can help to achieve.

2.4 Options for Archiving

The three main options for retention of electronic information are:

- On the live computer system that generated the records
- In non-electronic archives
- In one or several electronic archives

Computer technology changes will eventually make the option of live on-line storage untenable. The migration issues through upgrades and platform changes will make this option impractical and uneconomical. A continually expanding on-line archive also has the potential to degrade the performance of the live system.

A non-electronic archive has the attraction of low initial cost and the use of established methods. However, this may not allow the necessary interpretation and use of the data. In migrating to non-electronic media, much of the context and meaning of the electronic records may be lost. Whether this loss directly impacts the ability of the record to meet a particular regulated purpose will depend on the design of the archive in question; however, a wholly non-electronic archive is unlikely to retain all of the capabilities of a properly designed electronic archive.

Electronic archiving offers the prospect of long-term preservation of the required electronic information. However, there are factors such as different technology systems, archiving demands, and business areas or locations that can make a single archiving system difficult to manage. In such cases, the optimal solution may be to use a range of different systems or methods.

Archiving options are further explored in Appendix F of this Guide.

2.5 Electronic Data Archive System Lifecycle

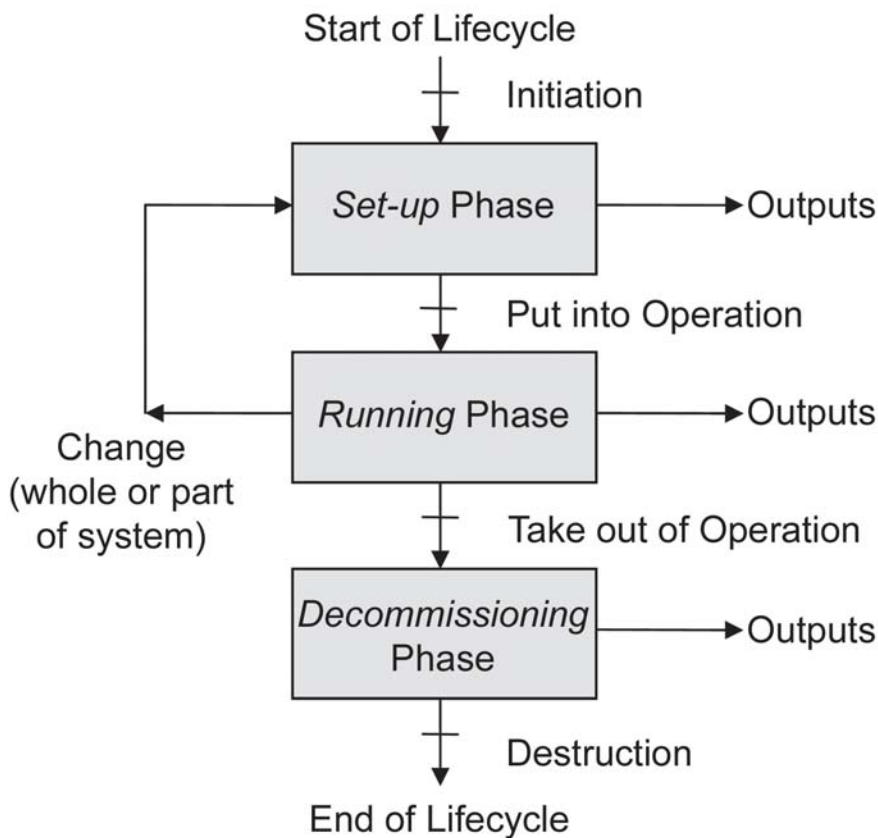
If the chosen approach is to use an EDA, this requires an EDA system to be specified and implemented.

An EDA system will move through a lifecycle that consists of three key phases:

1. The Set-up Phase relates to the specification, development, testing, and verification of the system.
2. The Running Phase relates to the day-to-day operation and maintenance of the system. During this phase, data will be accepted for storage in the EDA, migrated to other hardware platforms within the archive period, and deleted once the retention period has expired.
3. The Decommissioning Phase relates to the retirement of the system once the system has reached the end of its useful life.

This is represented in Figure 2.1.

Figure 2.1: System Lifecycle Phases



The transition from one phase to another should be controlled and each phase will generate outputs relating to the way in which the system is specified, designed, implemented, and operated. Given the intended long life of an EDA system, and the inevitable progression of technical change, it is unlikely that the system would move through the three phases in a single unbroken pass. It is likely to be subjected to changes, in which case the system, or part of the system, will move back to the Set-up Phase to take account of the necessary changes. Hence, different parts of a system can be in different phases at any point in time; however, the outputs from an executed phase may remain relevant throughout the system lifecycle.

The Set-up Phase is critical, as anything done (or not done) in this phase will have direct consequence on the subsequent lifecycle phases. This is aggravated by the long-term nature of archiving. To facilitate this phase, it is recommended that an Archiving Strategy document is prepared early and that agreement is sought from all the key stakeholders, e.g., Business Operations, R&D, QA, IS/IT, and regulators. (This list will be specific to each organization.)

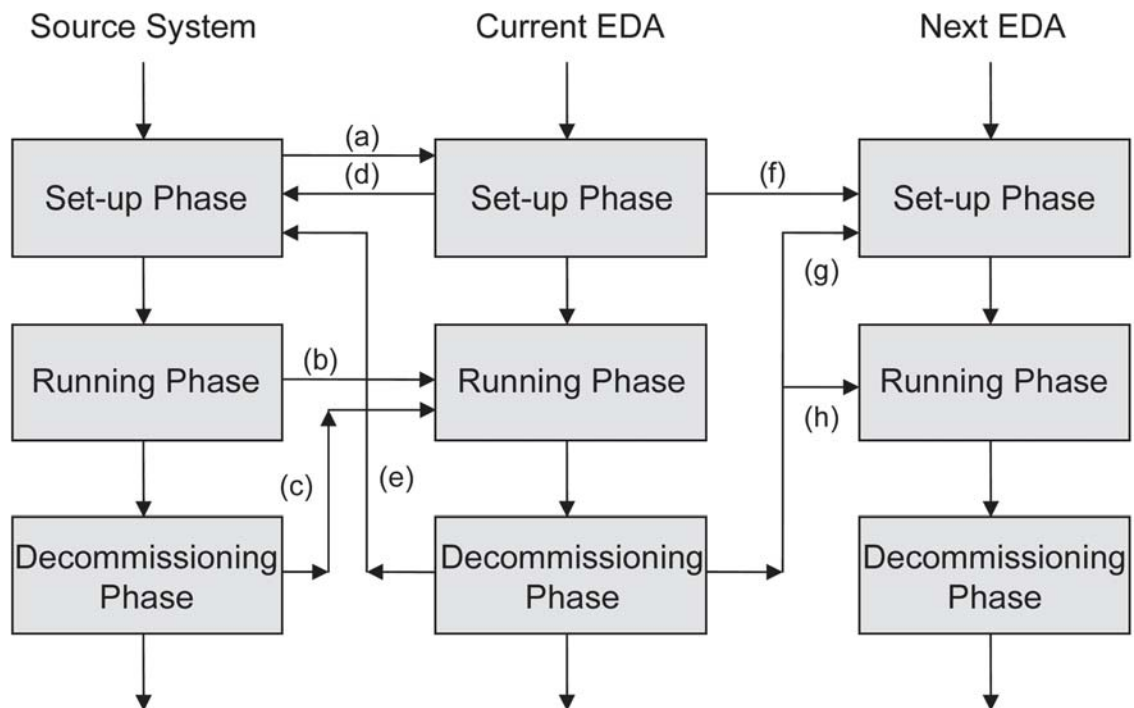
The purpose of the Archiving Strategy document is to attempt to document all necessary considerations for an EDA. It is recognized that there are unlikely to be commercial solutions for all identified issues, but by formulating them, an understanding of the demands and their implications can be gained. This Archiving Strategy document can then be used as a road map for the introduction of one or several EDAs.

Archiving often tends to be considered as an end-of-life process, e.g., when a system is retired or becomes obsolete. The danger of this approach is that the archiving process may be rushed and a large quantity of data dumped into storage, without considering future retrievals and retention periods. A better approach is to adopt a record lifecycle methodology, whereby data is archived in a secure and continually accessible form on an on-going basis as a normal part of its life. The Archiving Strategy document should help to define the considerations to be taken into account.

Section 3 of this Guide contains guidance on the creation and realization of an archiving strategy, and a template can be found in Appendix G of this Guide.

The simplified model shown in Figure 2.1 can be extended to take into account that the EDA will receive data from several source systems, each of which will, independently, move through a similar lifecycle. At some point, the EDA will be replaced by another EDA that also moves through a lifecycle; these two cycles are likely to be connected to each other. This extended model is shown in Figure 2.2. For simplicity, only one source system is shown, with no changes or outputs.

Figure 2.2: Extended Lifecycle Model



For clarity, only the major interdependencies between systems have been shown.

Each interdependency shown is explained below:

- When a new source system is set-up, it may have an impact on the set-up of the EDA or cause the EDA to be updated.
- Source system passing data to the EDA in the course of normal running.
- When the source system is decommissioned, final data is passed to the EDA, as well as informing the EDA that this data link is no longer valid.
- During the Set-up Phase of the EDA, it may affect the current source system with a need for the latter to be reconfigured to allow archiving of its data.
- Decommissioning of the EDA will impact the source system as the source system will no longer be able to archive its data.

- f. The set-up information for the current EDA is transferred (at least in part) to the next EDA.
- g. At decommissioning of the EDA, the set-up of the next EDA is likely to be affected.
- h. At decommissioning of the EDA, current archive records are transferred to the new EDA.

The above concepts can and should be specifically refined and extended for each proposed EDA system by defining the input and output data flows in each phase, followed by analyzing the dependencies to clarify cause and effect dependencies between systems.

2.6 Main Archiving Processes

Electronic data archiving is an important activity within the overall lifecycle of data, which ranges from data creation through to deletion. These high-level archiving processes are shown in Figure 2.3. This Guide covers the shaded areas in Figure 2.3.

Figure 2.3: Electronic Archiving Processes within the Data Lifecycle

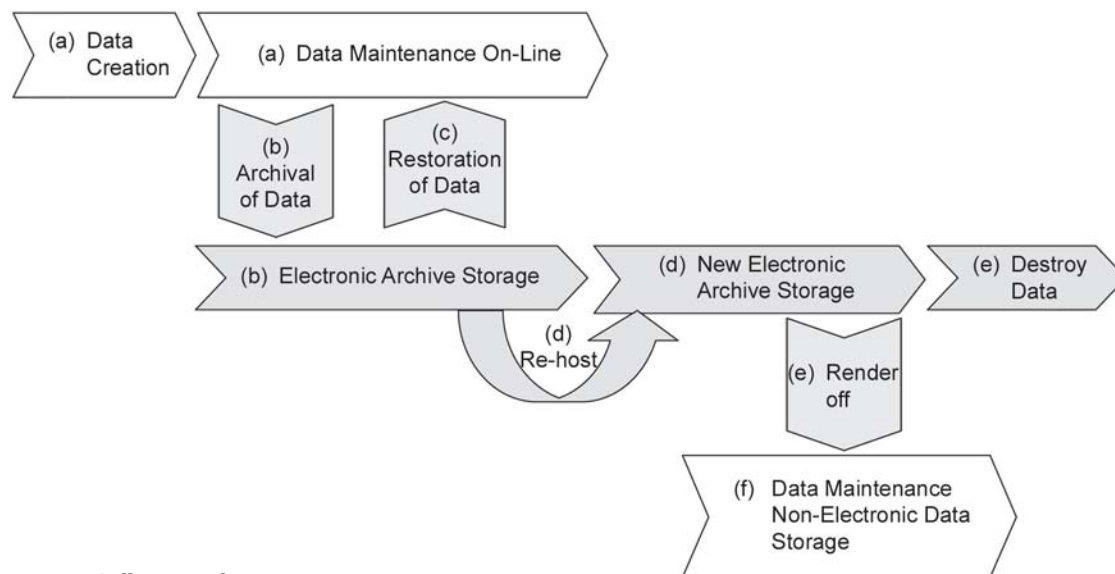


Figure 2.3 illustrates that:

- a. The creation of data and its maintenance in the on-line electronic system are already addressed within the scope of the main GAMP® Guide.
- b. Some of the data is archived off the on-line system to electronic archive storage.
- c. Some of this archived data may be restored from the electronic archive storage.
- d. It may be necessary to replace (re-host) the electronic archive storage, because of the longevity of data in the archive.
- e. Eventually, some of the data may be deleted or rendered off into non-electronic archive storage (e.g., printed or copied to microfilm).
- f. There are well-established methods for the maintenance of non-electronic archive storage.

Figure 2.4 shows the main archiving processes for a single archive. In addition, the figure identifies key roles. See Appendix G of this Guide for definitions of the roles.

Figure 2.4: Main Archive Processes

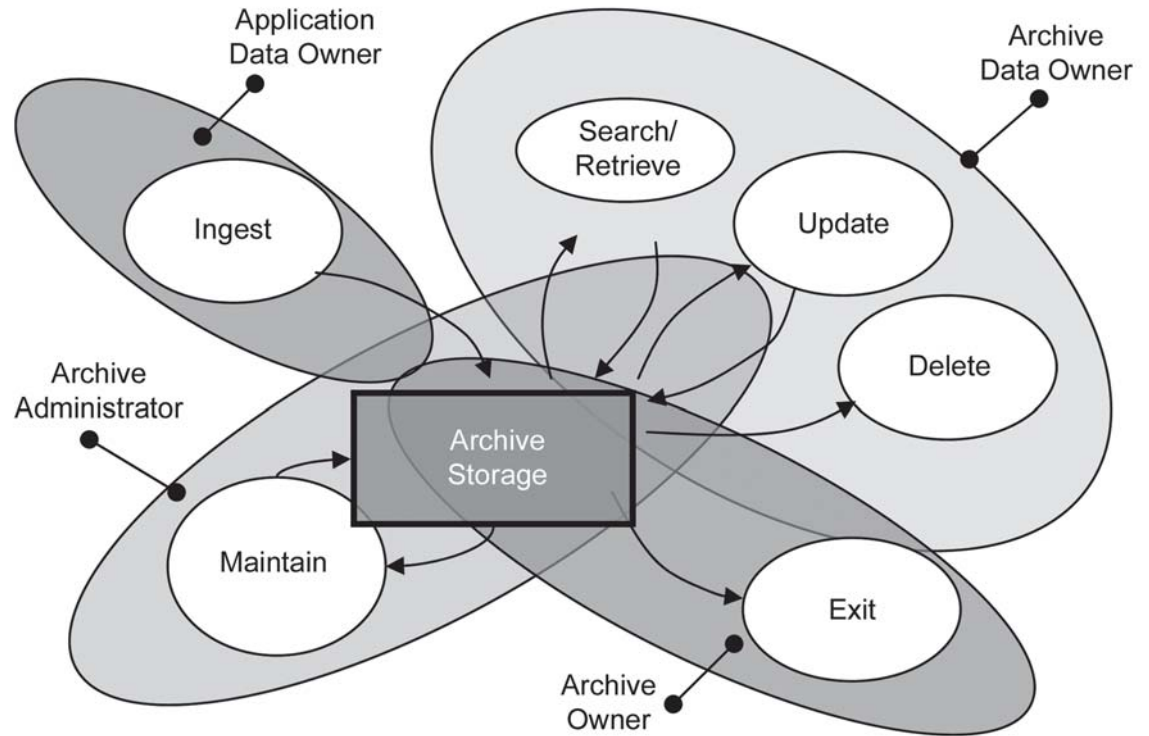


Table 2.1 defines the identified archiving processes.

Table 2.1: Main Archiving Processes

Process	Definition (for purposes of this Guide)	Responsible
Ingest	Process for moving data from source system into the archive.	Application Data Owner (supported by Archive Administrator and Quality)
Archive Storage	Process for keeping the stored data safe, i.e., protected from unauthorized and accidental modifications. This process is supported by the Backup/Media Refresh and Update processes.	Archive Administrator (supported by Archive Owner, EDA Technology Owner, and Quality)
Search/ Retrieve	Process for authorized users being able to interrogate the archive for stored data and to access the stored data.	Archive Data Owner of relevant business area (supported by Archive Administrator)
Maintenance Functions	Process for ensuring protection against system failure (backup and disaster recovery) and for ensuring the data integrity from a technical perspective, i.e., protection against archive media failure (media refresh).	Archive Administrator (supported by Technology Owner)
Update	Process for modifying archived data so it remains relevant, secure, and accessible for the current operation. Updates may be required as a result of personnel changes, software upgrades, technology changes, operational modifications, legal requirements to redact personal data, etc.	Archive Data Owner (supported by Archive Administrator)
Delete	Process for permanently erasing the data from the archive.	Archive Data Owner (supported by Archive Administrator and Quality)
Exit (Migration)	Process for transforming (migrating) the archived data to another repository at the end of the lifetime of the archive.	Archive Owner (supported by Archive Technology Owner, Archive Data Owner, and Quality)

Note that the processes have been shown only as one-dimensional, i.e., as applying equally to all data. This may not be so, e.g., for certain critical data additional verification processes may be applied.

For a more comprehensive description of archiving processes, please refer to Appendix B of this Guide for a description of the Open Archival Information System (OAIS) reference model.

3 Creating and Implementing an Archiving Strategy

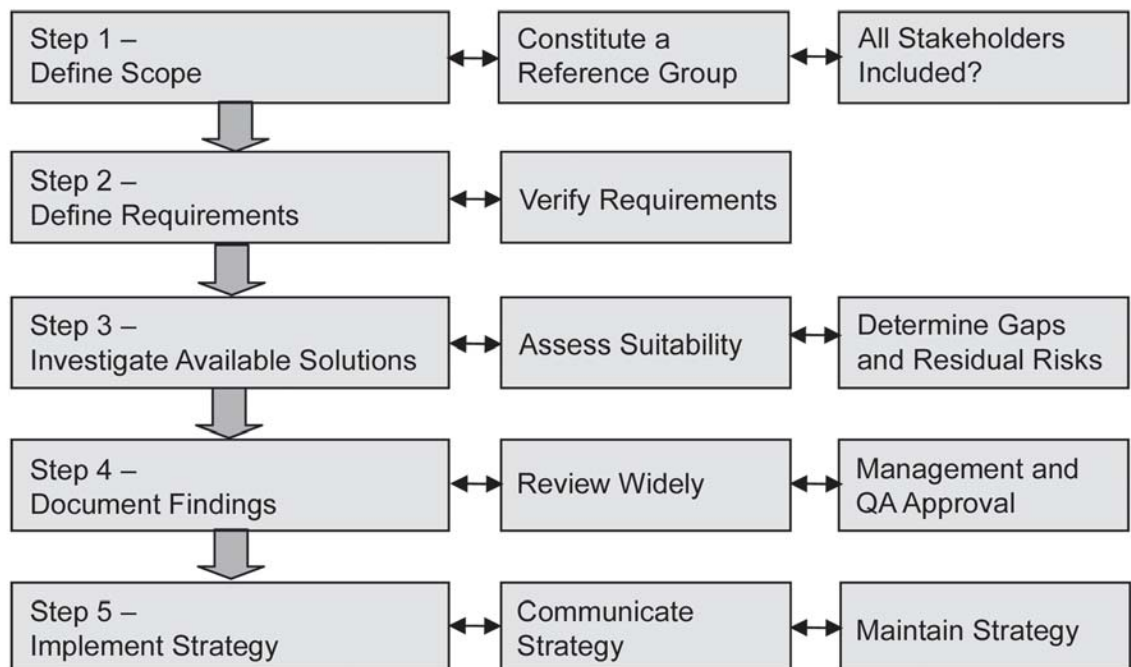
3.1 Overview

Creating and implementing an archiving strategy may be difficult for many reasons:

- It affects a number of disciplines.
- It has high regulatory criticality.
- It is likely to have an impact for a long period of time.
- It may involve use of new and unfamiliar technical solutions.
- It may involve judgmental decisions based on a balanced risk assessment and practicality of implementation, which arise from uncertainty as to what is achievable both in the short and longer term.

For these reasons, a defined project approach should be adopted. A suitable model is shown in Figure 3.1.

Figure 3.1: Model for Creating and Implementing an Archiving Strategy



Step 1 – Define Scope

Although the scope of the Archiving Strategy document would be defined at the beginning of the process, subsequent tasks are verified against, and may force changes to, the scope. Step 1 will be preceded by identification of the requirement, which will be further analyzed in Step 2. A reference group should be chosen to represent key decision makers and all stakeholders. Management backing for the project, as well as long-term commitment to the EDA systems and the supporting processes, should be sought.

Step 2 – Define Requirements

Requirements should be analyzed and classified, e.g., as essential, important, or non-essential. Requirements may affect the scope; if a particular requirement with a large impact is essential in only one particular area or user group, then the decision may be taken to develop this requirement separately.

Step 3 – Investigate Available Solutions

This step aligns the strategy to what is technically possible and commercially available; in essence a ‘reality check’ is performed. Available systems may come from the open market, but may come from internal systems (some of which may not have been known to the project team, initially) and these should be assessed for suitability. It is possible that by carrying out a gap analysis and determining the residual risks, an iterative process may be required to fine tune the strategy and scope. Resource requirements for the implementation should be estimated.

Step 4 – Document Findings

The development strategy should be fully documented and circulated among all the stakeholders and decision makers. Once re-drafted and acceptable, the document should, as a minimum, be approved by Management. Archiving of GxP relevant data also may require QA to approve the document. Due to the technical nature of an EDA, approval by IS/IT is usually essential.

Step 5 – Implement Strategy

In order to be effective, the strategy should be vigorously implemented and articulated. The purpose of the strategy should be well understood. This may involve specific communication on its applicability. Management support should be visible. No strategy, however well designed, will remain unchanged. Therefore, it is recommended that the strategy is reviewed at least every two years, possibly as part of a Periodic Review or similar activity. In addition, reviews of the strategy may be required during an EDA implementation.

Various considerations for each of the five steps are detailed in the following sub-sections. No further specific guidance is provided on how to conduct these steps; however, as normal project techniques would be applicable.

3.2 Step 1 - Define Scope

It is desirable to have a uniform strategy that covers all electronic data archiving demands within an organization, but this may not be straightforward or possible, as described below.

A single Archiving Strategy document will enforce a consistent approach to archiving, but possibly at the expense of making the Archiving Strategy document somewhat unwieldy. At a minimum, the archiving strategy would apply only to regulated areas, and could even be restricted to apply only to regulated data within those areas, if such data could be identified. Such identification of data is often impractical; however, it may be more effective for the document to cover whole groups of systems with related archiving requirements. It is possible to extend it to cover all archiving requirements within an organization, but this would remove boundaries between regulated and unregulated areas, which may be undesirable from an inspection and auditing perspective.

The Archiving Strategy document should define the complete data archiving scope. Data that may be assessed as not within the regulated area and not within the strategy may have an impact on other regulated data and might not be available for inspection or auditing. This aspect should be considered for larger systems.

It is unlikely that a large comprehensive EDA solution is rolled out across several functional and organizational boundaries in one go. It is more common for these large projects to be implemented in a controlled staged manner, perhaps with initially reduced functionality in a few selected locations, before being more widely applied across an organization.

It may be difficult to generate a common strategy. For various reasons, some areas, applications, and data may have special considerations that do not fit into a common strategy. In this case, the answer may be to develop an Archiving Strategy document for each solution/area. It is recommended that all such instances are captured and documented with justification and appropriate cross-referencing within each of the Archiving Strategy documents. A clear statement as to why the scope of the EDA has been split the way it has should ensure that these factors can be properly considered in any future revisions of the archiving strategy.

The scope should be clearly defined. The template document in Appendix G of this Guide provides considerations for defining the scope.

3.3 Step 2 - Define Requirements

For the purposes of this Guide, the archiving requirements have been split into the following Sections:

- Roles and responsibilities
- Process mapping
- Data requirements
- Compliance requirements

System requirements and commercial considerations are considered in Step 3.

3.3.1 Roles and Responsibilities

In defining the roles involved in archiving, consideration may be taken of existing defined roles (rather than creating several new ones) and also the size and complexity of the organization. In a small organization, it may be feasible to combine roles, and this is acceptable as long as the responsibilities are clearly defined. However, regulations may demand that there are separations of duties, e.g., SOX requires that there is no conflict of interest, and GLP requires that a separate archivist is identified and made responsible for the safe keeping of archived records. As a rule, the Management and Quality functions should be separate from operational roles.

When allocating roles and responsibilities to individuals, their qualification, training, and experience should be taken into account. Individuals should have documentary evidence that they can perform the allocated role with their given background and capabilities, and the tools, systems, and procedures that are made available. Systems already in place for the selection of personnel and documentation of their background, experience, and training can be used.

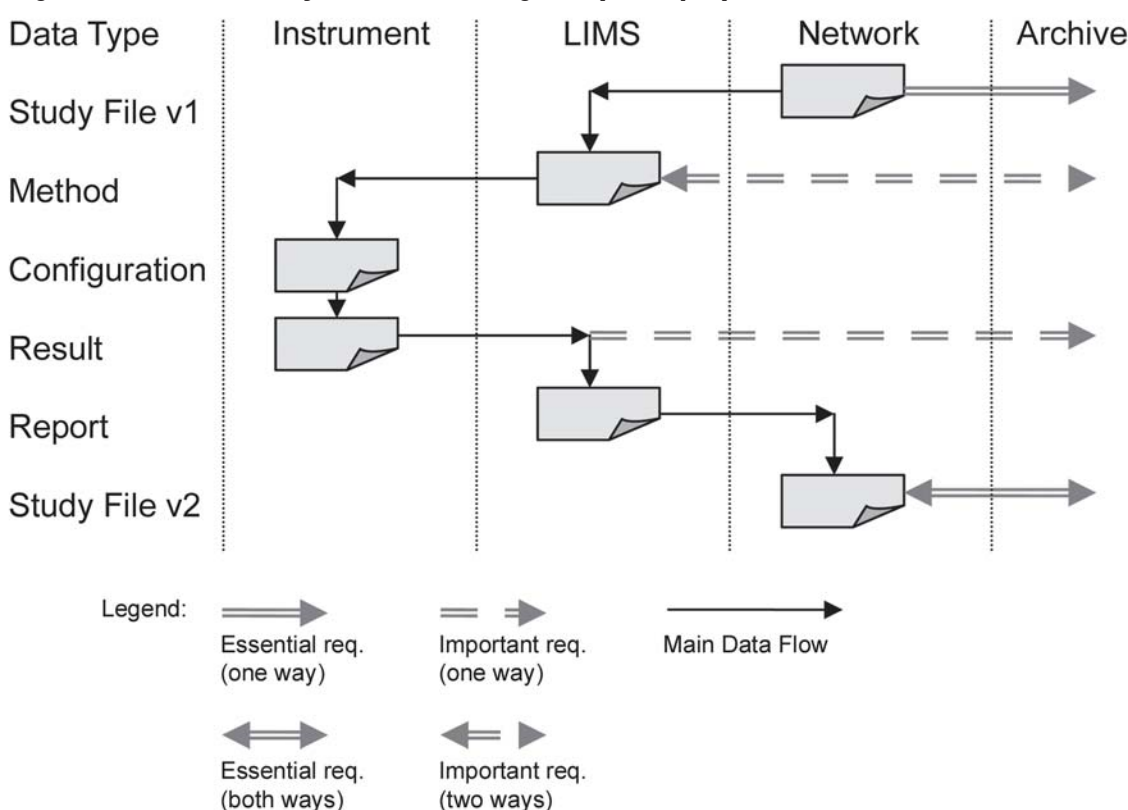
The strategy may be one that involves several transitional stages, as the envisaged technology and resources become available. When allocating roles and responsibilities, consideration should be taken of likely changes in the short and medium term, to prevent unnecessary changes in personnel due to inadequate ability to handle the future requirements.

Apart from the roles identified in Appendix G of this Guide, there may be legal considerations to be taken into account. The strategy may need to include a defined role for Regulatory Affairs and Legal departments. In addition, an implementation project should include an IS/IT Specialist with knowledge of the technicalities of long term secure storage of electronic data and of secure migration of data.

3.3.2 Process Mapping

In determining the various archive processes, and what data to archive (see Appendix G of this Guide) it may be helpful initially to generate data flow diagrams that depict how data is moved around the organization. An example from a GLP laboratory is shown in Figure 3.2:

Figure 3.2: Laboratory Data Flow Diagram (Example)



The figure shown is simplified, but can be extended with additional data, such as access rights, audit trail, and sample table. For each of these instances, the archive requirement can be stated as:

- Essential
- Important
- Non-essential

In the example shown in Figure 3.2 the requirement is to be able to archive the initial Study File v1 and final Study File v2. It also must be possible to restore the Study File v2 to the network drive so it can be further developed should the need arise. Archiving and retrieval of the Method and archiving of the Result are seen as desirable, but archiving of the Configuration File and Report has been excluded from scope in this example. This should be justified with a rationale, usually based on the GxP-relevance of the data.

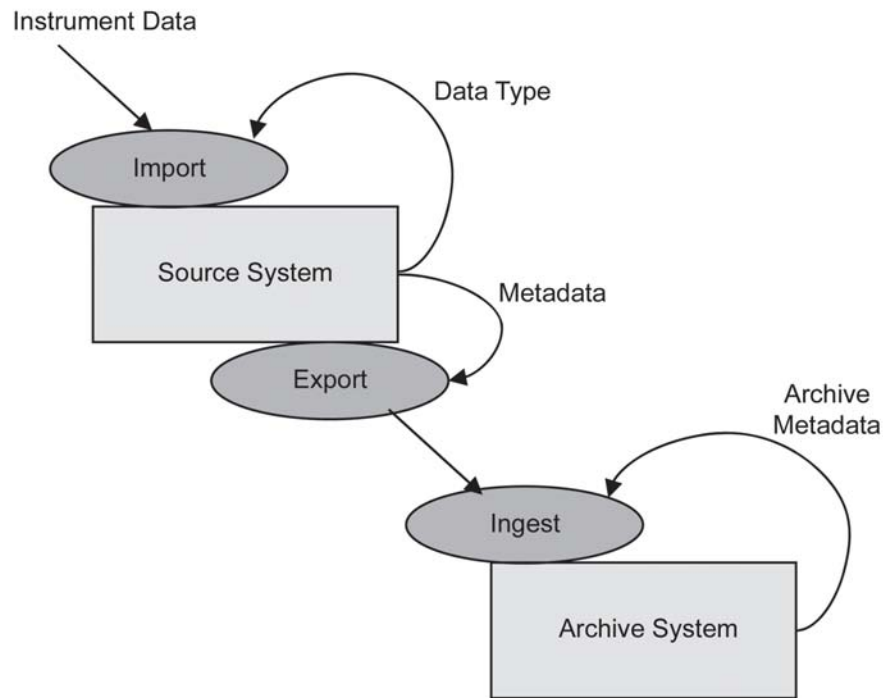
The data flow diagram can be further enhanced by adding job roles that control or approve the data. This may highlight potential change of ownership issues that need to be addressed. For example, in Figure 3.2, ownership of data may remain with the Study Director throughout the process, but also may pass from the Study Director to the Laboratory Manager and then to the Analyst. It may then return to the Laboratory Manager and the Study Director, and finally the Archive Data Owner. Each change of ownership should be controlled through defined processes.

3.3.3 Data Requirements

3.3.3.1 Data Definitions

It is important to define the data being processed by the EDA. Figure 3.3 illustrates the concept, taking the ingest process as an example.

Figure 3.3: Example of Ingest Process



Data definitions are considered as having four attributes:

- Data content
- Data type or the format of the data
- Metadata
- Archive metadata

Data type and metadata are two of the dimensions that can be used for defining the scope of the EDA; the strategy should define the data types and metadata that are supported and that can be archived in accordance with the strategy. Classification of records will assist in specifying the archive requirements, rather than looking at individual records.

3.3.3.2 *Metadata*

Metadata gives data meaning and context. There are some key attributes that should be stored with the data when it is archived or converted to another format during data migration. Required metadata may differ for the different types of data and content, but addressing the key attributes will ensure a minimum and consistent level of context information.

Some typical examples of metadata that present difficulties during archiving or conversion to other formats include:

- Security information and access permissions
- Font/color of artwork text or graphics where an alternative system may not support the color format
- Additional information such as virtual sticky-notes

The data context may be captured in an orientation memorandum. This would provide the full context of the data, including:

- Why and how was the data generated?
- How the data could be used?
- Why has it been archived?
- How the data relates to other data?

The key objective of the orientation memorandum is to turn data into information, i.e., the archived data becomes something that can be useful. One orientation memorandum could cover a large number of related archived files and would need to be indexed and linked accordingly.

In addition, the archive itself is likely to impose a set of archive metadata. This is metadata, e.g., search criteria, that is required for realizing archive processes. Apart from the supported archive processes, the archive metadata is likely to be dependent on the archive platform solution, i.e., the technical realization of the archive. Some of this archive metadata may be automatically assigned by the EDA, but other data may be required to be stored with the data content, in which case, it should be clearly specified. The amount of manually entered data should be reduced to a minimum for ease of archiving.

3.3.3.3 *Retention Periods*

From a regulatory perspective, the retention period for the archived data is of key importance. GxP regulations define how long different types of records are required to be retained. Similarly, business and legal considerations also will dictate retention periods, e.g., data relating to patents. In most cases, it is expected that the organization already has well-defined record retention schedules, and this Section provides only general guidance.

The retention period should be assigned to the archived data content. This may be done in one of two ways:

- Individually with each record, e.g., by the use of metadata
- By defining a retention period for each class of data, where class may refer to:
 - Physical location

- Organizational unit
- Data source
- Data owner
- Data type
- Data content or regulatory requirements

In the first instance, the required retention period for each record or datum is assessed individually. This is likely to be practical only where the diversity of archive data is low.

In the second instance, generic rules based on data content are used for assigning a retention period. Such generic rules should be set on a worst case basis, i.e., the retention period is set to be at least as long as that for the individual record requiring the longest storage period. Table 3.1 illustrates how generic retention periods can be applied. (Note that the table provides examples only.)

Table 3.1: Retention Periods

Property: Retention Period	Consideration
Manufacturing Data	Retain for five years after production of batch. This assumes that batch expiration date is no longer than four years.
Laboratory Data for Manufacturing	As for manufacturing data
GLP Data	Retain for 30 years (or forever).
Unregulated Studies	Retain for five years (e.g., to allow data mining).

Retention of data, and particularly deletion of data, is also affected by commercial and legal considerations. The company, regulator, or courts may impose a litigation or legal hold, which will override any defined retention period and prevent the data from being deleted.

Note that the above applies equally to the associated metadata.

3.3.4 Compliance Requirements

3.3.4.1 Risk Assessment

Risk assessments are particularly applicable to an EDA. Applying a risk-based approach through a structured risk analysis is seen as a practical approach that is acceptable to the regulatory bodies.

The GAMP® Good Practice Guide for a Risk-Based Approach to Compliant Electronic Records and Signatures gives further information on methods for conducting risk analysis, including risk identification, assessment, and prioritization.

One aspect of a risk-based approach to compliance is that in low risk areas, a lower level of documentation, controls, or verification may be applied, and additional effort expended in areas identified as high risk. Therefore, the risk analysis should be properly documented. In particular, the rationale for conclusions reached with regard to residual risks requires justification that the residual risks are deemed to be acceptable. The rationale should be realistic and consistent with good practice, and appropriate to the regulated area and the computer system concerned.

3.3.4.2 *Regulatory Requirements*

Example archiving requirements by various regulatory authorities is included in Appendix A of this Guide. It is not intended to be complete or comprehensive.

Records should be maintained according to the predicate rules, a justified and documented risk assessment, and a determination of the value of the records over time.

Off-loading archived records to a non-electronic medium (e.g., paper or microfilm) or to a standard electronic file format (e.g., PDF, XML, SGML) is acceptable, provided that all predicate rule requirements are complied with and that the content and meaning of the records are preserved. Hybrid records (a combination of paper and electronic records) are generally permitted. Archives also should comply with relevant national and international consent and confidentiality legislation.

3.3.4.3 *Achieving and Maintaining Compliance of the Archive System*

The EDA should be compliant with all applicable regulatory requirements. The Archiving Strategy document should detail all applicable requirements and describe how these will be met.

All relevant approved company policies and procedures should be followed and as the EDA is fundamentally a computer system, appropriate good practices as defined in the main GAMP Guide may be used. A blanket reference to GAMP in company procedures or the Archiving Strategy document is not recommended as GAMP is written for a wide range of systems, situations and audiences, and not all are likely to be applicable to a specific EDA.

Guidelines found in the main GAMP Guide have not been repeated in this Guide. Table 3.2 deals with items that may merit special consideration when developing the Archive Strategy.

Table 3.2: Special Considerations

Aspect	Consideration
GAMP Categorization	The EDA is likely to contain several software components and be a mixture of both standard and custom software. See Note 1.
System Security	Applies to all computer systems, but of particular importance to the EDA. The role and responsibilities of the archive administrator and super-users should be addressed. See Note 2.
Data Integrity	Of key importance to the EDA. This includes unique identification of records so that there is no risk of incorrect or duplicated identity. Record identity must be robust to changes. For example, a record that relies solely on the folder hierarchy and path name for its identity is probably susceptible to changes.
Data Cut-over	As opposed to most business systems, cut-over can be a long period, and requires a managed process that addresses archiving in parallel to 'old' and 'new' systems. See Note 3.
Validation of Archive Processes	See Section 3.3.4.4.
Data Migration including deletion of records	See Section 4.2.
Performance Verification	See Section 3.3.4.5.
21 CFR Part 11	See Section 5 for regulatory requirements.
Change Control and Business Continuity	See Section 3.3.4.6.
SOPs	See Section 3.3.4.7.
Maintenance	See Section 3.3.4.8.
Periodic Reviews	Applies to all GxP relevant computer systems, but of particular importance to the EDA, due to the expected long period the EDA is used for. Periodic reviews are essential for confirming compliance and continued fitness for intended use.

Some of the table entries have been elaborated in Note 1 to Note 3.

Note 1: It is recommended that effort is concentrated on any configuration and custom code. This approach presupposes that documentary evidence exists that code considered to be standard is not specific to the actual installation. This evidence may be collected through auditing or investigation of the code and supplier of the code.

Note 2: Control of the System Administrator and Super-User accounts are a known problem for most computer systems, and particularly for the EDA. One way of addressing this is by assessing the opportunities for unauthorized (or fraudulent) use, weighed against the inclination to perform these operations. Where an unacceptable risk level is determined, risk mitigation through a combination of additional security and supervision, tighter procedural controls and screening of individuals should be implemented.

Note 3: Data cut-over may require managing two parallel archive systems. For example, cut-over may be made on a project/study or location/system basis. This means the cut-over process may be lengthy, and it is recommended that a separate cut-over plan is prepared that details how this process will be managed, controlled, and monitored. A particular concern is how the routing of data is managed during this period, i.e., what controls are in place for ensuring the data is directed to the correct archive.

3.3.4.4 Validating the Archive Processes

The validation of the various archiving processes outlined in Section 2 of this Guide is likely to be a complex task, as it is expected that there will be many parameters associated with each process.

Table 3.3: Example Parameters

Parameter	Examples
Interfaces to EDA	Location, type
Users	Number of users, user role, and profile
Source Data	Type, content, volume, metadata
Purpose of Process	Required checks, approval, archive location

Dealing with these parameters and the large number of possible test permutations make complete test coverage impractical, and will require a structured test strategy. Guidance on testing is available in main GAMP® Guide and the GAMP® Good Practice Guide for Testing of GxP Systems.

The test strategy should address the following:

- Test coverage to be specified (e.g., as a percentage of all possible permutations) with a mapping of planned tests against possible test cases
- A risk analysis that justifies the chosen tests and test coverage. This may be based on regulatory criticality of the processes and data, business criticality of the processes and data, plus level of criticality determined by software standardization and rigor of applied design rules, checking, and testing.
- Challenge testing of the archive process that includes boundary verification of critical parameters, negative testing using outside boundary values, worst case testing (using elements of simulation), error handling, and failure recovery
- Test validity based on the chosen test data. As system implementation progresses and the system expands, the validity of the used test data may need to be revisited, and may result in additional testing in the future. Testing should, where possible, closely simulate all the intended use-cases of the archive data.

Some processes will be more difficult to verify than others, e.g., the search process, where a search criterion is specified and the process returns all the instances that fulfill that criterion. In this case, the difficulty may be alleviated by:

- Using known test data (so the result is known in advance)
- Making a comparison between multiple searches for signs of inconsistency in results
- Defining and restricting the search rules (such as used data input format)

- Using only well-respected standard search tools

3.3.4.5 *Performance Verification*

Performance verification, or an equivalent, is a key activity in ensuring the EDA is fit for its intended use because:

- It provides the opportunity for performance testing of the EDA in a representative operating environment using typical expected data traffic levels and database loadings, including testing by formally trained users in accordance with approved operating procedures.
- The often large user-base and many system interfaces and processes will put emphasis on the verification that the EDA is operating consistently as an entity; this is usually difficult to demonstrate during previous testing.
- It is likely that the EDA is put into operation in stages, i.e., users and system interfaces are added on an on-going basis. This will, over time, invalidate initial performance verification and require renewed confirmation.
- The EDA is at risk from the data avalanche, i.e., through technological advances increased data through-put rates can be expected. This will require a sustained allocation of technical resources.

As the EDA is likely to evolve over time, there is expected to be a program of verifications over time, linked to the maturity of the EDA. This program should be linked to reviews of system usage and change management to ensure its timely execution.

Verification should include tests of critical performance parameters, such as the limits on number of files per dataset, size of files, size of dataset, number of simultaneous users submitting datasets to the EDA. A key element in this aspect is to demonstrate the ability of the EDA and the network infrastructure to handle the required volume of data and requests, as well as provide retrieval of information to the consumer in the specified time and format. Speed of retrieval of information will be affected by the critical performance parameters as well as the implemented technical solution.

3.3.4.6 *Change Management and Business Continuity*

Change management and business continuity take on additional importance for the EDA, because of the likelihood of a long period of use of both the system and its data. Additional measures may need to be taken.

The EDA may interface with and receive data from a large and diverse user base of source systems. Each of these poses a potential threat to the integrity of the EDA through uncontrolled changes. Changes to the source systems may jeopardize how the data is migrated to the EDA and potentially invalidate the ingest process. Such changes may include:

- Software upgrades
- New work patterns
- Additional templates
- New methods

The necessity of managing these changes appropriately implies the need for change management that spans the whole user-base of the EDA. It is recommended that the system owner demonstrates how change management is applied in a coordinated manner across what may often be organizational, operational, and physical boundaries within an organization. The GAMP® Good Practice Guide for Global Information System Control and Compliance contains further information on these topics.

It is good practice to develop procedures and guidance to assure business continuity. In addition, due to the EDA's longevity, it is recommended that an exit strategy is developed. This is covered in Section 4.4 of this Guide.

Closely related to business continuity is disaster recovery, which includes being able to adequately handle disasters such as fire, floods, and criminal damage. The realization of the EDA is often a centralized repository. When designing its backup system, these factors should be taken into account by, for instance, having remote backup on a disparate system and in a separate location. The level of advanced planning and protection should be adequate for the most business critical systems.

3.3.4.7 *Standard Operating Procedures*

The existence of well-written, applicable, and proven standard operating procedures applies in equal strength to the EDA as it does to any GxP critical system. One particular aspect is the potentially large and diverse user-base of the archive, in particular for the ingest process. The archive processes need to be controlled through SOPs.

Table 3.4 summarizes which processes are expected to be covered by SOPs. The actual number and naming of SOPs is determined by the organization - SOPs can be combined or split to suit the business model used.

Table 3.4: Typical SOPs

Process	Typical Scope of SOPs
Archive Processes	Routine Operations: Ingest (moving data into the archive); Search and Retrieve; Data Update
	Non-Routine Operations: Data Deletion; Data Migration (to/from another repository)
Maintenance Processes	IS/IT processes such as backup and restore, maintenance, software upgrades, virus protection, disaster recovery, security, etc.
Management Processes	Administration of users and user groups, system incident reporting, business continuity, management controls, etc.
Quality, Regulatory, and Validation	Document management, required documents to be maintained, change control, regulatory requirements, quality checks and approvals, validation procedure, etc.

3.3.4.8 *Maintenance*

Some archive processes are run infrequently and the result of an archive process malfunction may not become obvious for a long period of time. For example, apart from self-verification of the ingest process itself, an archived record may not be accessed for several years, at which time it may cause considerable problems in trying to rectify any mistakes.

Implications from maintenance activities may not become obvious for some time. Consequently, tasks need to be clearly defined with a risk assessment to determine their potential impact so that reasonable precautions can be established.

Maintenance also should include operational monitoring to establish that the archive is functioning as per specifications and procedures. This should be done in addition to Periodic Reviews.

Where external resources are being used for executing and managing maintenance activities, the same concerns apply as for System Administrators, see Note 2 (see Section 3.3.4.3 of this Guide).

3.4 Step 3 - Investigate Available Solutions

This step is related to the realization of the actual EDA system, rather than the content of the archive. The aim is to define the desirable platform solution, based on what is achievable, rather than specifying the system requirements in detail. This Section addresses considerations arising from the technology platform, security requirements, location of the EDA, and commercial matters. In addition, each organization may have additional considerations driven by corporate and strategic demands.

3.4.1 Technology

A broad technology platform that will underpin the EDA and support the archive requirements should be defined. The technology platform is likely to have a high impact on the business and would require a substantial investment to change; therefore, the platform solution is likely to be in place for some time. The GAMP® Good Practice Guide for IT Infrastructure Control and Compliance contains further guidance on platform related issues.

Examples of technology platforms are given in Table 3.5.

Table 3.5: Example Technology Platforms

Technology Platform	Examples
Operating System	MS Windows Mac OS Unix
Architecture	PC-based client-server configuration Centralized main frame server with dumb terminals
Operator Interface	Open Web-enabled solutions Closed Local Area Network
Supplier Solutions	Servers: Compaq (HP), Dell PC clients: Compaq (HP), Dell Tape library: HP CD-R jukebox: JVC Network: Cisco, HP Database: Oracle, Microsoft

Reasons for defining these technology platforms in the Archiving Strategy include:

- Once a key element of the technology solution has been selected, any future changes are likely to be both costly and disruptive. For example, the transitions from the existing solution to the new solution may not have been accurately mapped, leading to lengthy verification of functional, operational, and content migration processes, such as ensuring no loss of data or information when moving from one solution to another.
- Technology development is, in part, driven by individual companies, resulting in proprietary solutions that may not enable a mix-and-match approach. Regulatory compliance to a large extent is driven by having readily available technology solutions. End Users are extremely reliant on key suppliers to deliver these solutions. Therefore, selecting the right supplier has an impact on the ability to provide compliant solutions.
- There are no right or wrong answers to the various technology platform decisions. No single commercially available solution is likely to be wholly compliant with regulatory demands. The chosen platform together with a brief rationale for the decision should be specified.

There is a general question regarding regulatory compliant solutions and selecting suppliers - should the most compliant technology solution be selected, or is it acceptable to use the solution supplied by the market leader?

The solution supplied by the market leader may not be the most compliant or best technological solution. End Users may wish to move away from using the leading supplier and seek custom solutions from smaller providers in an effort to enhance technical compliance. However, the overall compliance and business risks from such a move should be carefully considered. Compliance of the total solution may be a combination of technical and procedural compliance. A best-fit technical solution within the operations and philosophy (e.g., buy versus build) of the organization should be sought.

The ability of a supplier to provide sustained development and support is highly desirable. An established market leader is likely to remain so, at least until there is a technological change. In the context of archiving, this is an important consideration with regard to the often extended archiving periods that require stable solutions.

The following steps should be taken when selecting a provider of EDA solutions:

- Any regulatory compliance gaps are identified and their impact assessed
- Reasonable steps are taken to mitigate the implications of such gaps to an acceptable level
- There is a plan for dealing with the residual risk after mitigation of the implications of such gaps

The last point may entail communicating the requirements to the supplier, and pursuing a technical up-grade solution once this becomes available.

When deciding on the broad technology platform, consideration should be given to the scalability of the selected platform. The optimal platform should ideally be able to cope with future, and as yet un-quantified, demands. While this may not be fully achievable, the scalability of the technology platform should be assessed.

3.4.2 Security

Security covers the following three areas for both the EDA and the surrounding IT infrastructure:

- Physical security
- System logical security
- Application (logical) security

Physical security includes security of locations that house computer equipment that gives access to the EDA. For example; it is common practice that critical servers are housed in physically secure locations, e.g., using of coded door locks.

System logical security includes the logical controls for accessing the computer equipment, e.g., gaining access to the company network and starting the application.

Application security includes the logical controls that determine what each user group is allowed to do within the application.

The security requirements should be specified both in regard to the EDA, e.g., user groups, and for any wider impact on other systems or processes. System requirement also should specify potential, material impact on the EDA by other systems or processes. Examples of impact considerations for the three areas of security include:

- Physical security requirements may have an impact on the location/environmental requirements that go beyond the EDA itself.
- System logical security requirements are likely to be impacted by the IS/IT infrastructure and corporate security policies and standards.
- Application security may be impacted by organizational changes that modify the defined user groups. Regulatory demands may impose changes of access rights for user groups. For example, a change in corporate procedure for approvals may have material impact where the EDA is not sufficiently flexible or technically sophisticated to accommodate such changes, e.g., the introduction of additional signatures and verification procedures.

3.4.3 EDA Location

An EDA does not have to be located and operated in-house. It is possible to use a third party supplier for hosting the archiving services. This situation is similar to outsourcing of IS/IT infrastructure and services, and similar considerations would apply:

- Audit the potential supplier of the outsourced services for suitability based on factors such as competence, capacity, continuity of service, cost effectiveness, financial viability, trustworthiness and integrity.
- Define the roles and responsibilities of key positions, both internal and external.
- Define the procedures to be used and how these will be controlled.
- Define the management and quality controls that will be used to ensure that the service is delivered to the set parameters.

These requirements should be documented in a formal service level agreement (or similar contractual document) and the agreement should be approved by all parties involved. It should be demonstrated that the agreement is in place, applicable to the scope it covers (including an appropriate level of monitoring and management of the third party by the Sponsor organization throughout), and is being applied satisfactorily.

From a regulatory perspective, the Sponsor organization remains accountable for the archive. It is possible to delegate the responsibility, but not accountability. For example, a key consideration of outsourcing the archiving of GLP data is that the Sponsor organization is responsible for providing GLP QA oversight of the outsourced archiving operations as part of its accountability.

3.4.4 Commercial Considerations

Commercial issues should be carefully considered, but are outside the scope of this Guide.

It should be noted that the initial project cost does not represent the total cost of the EDA over its lifecycle. The on-going operation and maintenance costs and the retirement costs can be substantial, and should be properly allowed for when calculating the cost of the EDA.

3.5 Step 4 - Document Findings

At this point, needs should have been analyzed and requirements should have been defined and compared with available solutions. The conclusions of this process should be documented. A suitable template for this purpose is provided in Appendix G of this Guide.

It is also good practice to document any basis for the conclusions reached. Examples may include information on why the selected strategy was chosen, why alternatives have been discarded, and which considerations have been given the highest priority and why. Such information is useful when updating the strategy or when the strategy details are questioned.

Approval of the Archiving Strategy document is deemed sufficient evidence of due diligence review and agreement of the signatories. It is good practice to have this document signed by all key stakeholders.

3.6 Step 5 - Implement Strategy

Simplistically, there are two main ways to implement the Archiving Strategy for a particular EDA:

Option 1: control and manage the process in-house. Develop detailed User and Functional Requirement Specifications (RS/FS) that are then used to select a delivery solution.

Option 2: leverage the available market solutions. Develop a high level RS for use in selecting the delivery partner(s), and then jointly develop the detailed FS with the selected delivery partner(s). These partners may be external or in-house resources.

In both cases, the Archiving Strategy document would be used as a basis for the EDA implementation. The main characteristics of the two options have been summarized in Table 3.6:

Table 3.6: Project Management Aspects (Generalized)

Project Management Aspects	Option 1 (Detailed RS/FS)	Option 2 (High Level RS)
RS developed in-house.	Detailed RS/FS document	Abbreviated
FS	Detailed RS/FS document developed in-house.	Developed jointly with key supplier(s)
Applicable solution for the actual application	Could lead to more custom solutions that are tailored for the actual application	Likely to lead to more standardized solutions that may require adapting or compromising for the actual application.
Project Risk	Reduces risks from changes by fixing requirements early	Reduces technical risk by using supplier's standard solutions Increases technical risks by tying to single supplier solutions
Project Timescales	Lengthens timescales through initial detailed work on RS/FS and requirement for possible design development to meet custom demands	Reduces timescales by engaging key supplier(s) early
Project Costs	Project costs are determined early Custom requirements may not leverage supplier's full functionality leading to under-utilization	Supplier standard solution likely to have cost advantages Full project cost is unknown until later in the project Reduction in competitive tendering could lead to cost increases for changes

Due to the criticality of the implementation of the strategy, it is recommended that a documented and proven project model is used and that reasonable documentary evidence is created and collected, which will demonstrate that a controlled process has been applied. The more prolonged the project time scales become, the more important this becomes.

4 Considerations When Archiving

This Section contains a number of archiving considerations, good practices, and items to avoid or be aware of when dealing with EDA. The learning points can be only general in nature, and their applicability to specific situations should be carefully judged. Care should be taken to adhere to company standards, and these may not be consistent with some suggestions made below.

A risk-based approach to all these aspects is recommended.

4.1 Summary of Archiving Issues

What follows is a brief overview of various issues related to the electronic archive. They are not arranged in order of importance and have been included to illustrate the complexity of the subject. The rest of the Guide, including appendices, discusses many of these issues in more detail.

4.1.1 Regulatory Compliance

An awareness of applicable regulatory requirements is essential. In general, these regulations address the need to preserve the data being archived in a form that allows it to be reliably and completely recovered and inspected over a pre-defined period.

In many cases, the requirements of regulations are consistent between different regulators, and on-going efforts are being made to bring requirements into greater practical alignment. However, the specific requirements may not be the same in all cases. Specific record retention requirements, as well as data protection and privacy requirements in cases involving identifiable personal data, may apply.

Please refer to Section 3.3.4.2 of this Guide for an overview of relevant regulatory requirements with a more comprehensive listing in Appendix A of this Guide.

4.1.2 Trustworthy Records

A trustworthy record has its content, context, and sometimes its structure maintained in a secure manner and has:

- Integrity, i.e., the record is archived following defined secure processes and stored under safe conditions. The record also can be identified with its associated metadata.
- Authenticity, i.e., the record has some quality attributes that enables the user to ascertain its status, e.g., who created it, any changes, electronic certification.
- Usability, i.e., the associated metadata enables the record to be read, the content understood, and the efficient use of the record.

4.1.3 Ownership and Responsibility

In some circumstances, more than one owner may be responsible for, or has ownership of, a record. An example is in the area of records related to clinical trials. Subsequently, there should be an understanding of the requirements placed upon the different entities involved with the trial to maintain records. Regulators may require that some records are independently maintained by more than one party, and particularly that some records are maintained outside of the control of the Sponsor. In developing an archiving strategy, it is important that these matters of separation of control over archives between sponsors and investigators are carefully addressed, alongside the need to maintain a coherence and comparability of the overall data.

4.1.4 Data Privacy and Confidentiality

Increasingly, the concept of personal data being the property of the individual to whom they relate is gaining status in national and international law. When designing archives these rights should be taken into account. The individual may have right of access to data, the right to have the data corrected or even deleted, and can in certain cases withdraw consent to the use of the data for particular purposes.

Additionally, identifiable personal data obtained in one jurisdiction may be restricted in its transfer into other jurisdictions. This is particularly the case with data obtained in the EU being transferred to the USA. Therefore, great care should be taken in designing any archive which contains, or may contain, identifiable personal data, particularly where systems support multiple locations.

4.1.5 Access Control

Data contained within archives has similar requirements applied to it as data within a live system. The ability to readily transfer access permissions from a live system to the archive can present a challenge. Access control systems within live applications are typically based upon identified individuals, job roles, and organizational structures. However, none of these categories are likely to be completely stable over time, potentially requiring different categories to be used for the archive. Access rights may be part of the record to facilitate access management.

4.1.6 Data Migration and Rendering

In many cases, and particularly in dealing with records retained for regulatory purposes, it is important both that data can be reliably recovered and that it can be presented in a form that, as closely as possible, replicates the content and context of the original information, and if possible, also the format in which it was originally created. Due to software and technology up-grades, archived data often should be migrated and rendered. These processes pose potential problems to the extent that they can alter the appearance and content of records. Please refer to Section 4.2 of this Guide for further considerations on migration of records and data.

4.1.7 Evolving Data Types

The electronic archive is likely to be required to deal with a growing number of data formats, ranging from text files to relational databases, complex instrument file formats, graphics, video, sound, etc. The challenges associated with long-term management of these data types, and in particular of making them available to search and explore, should not be underestimated.

4.1.8 Metadata

As identified above, metadata is critical to the value of any archive. It is also one of the easiest aspects of the development of the archive to overlook. In determining the metadata for a particular archive, it is often tempting to select a particular set of internally developed standards that can be applied. However, these need to be kept under constant review, both because of the changing scope of data to be archived, and also because standards evolve over time. To a lesser extent, this applies also to external standards.

4.1.9 Longevity

At a physical level, many of the electronic storage media that are used have shelf lives of less than ten years, and the data formats used also are often obsolete within a decade. Additionally, the ability to use the devices capable of reading these obsolete media is itself being lost, as new operating systems cease to support many of the older peripheral devices. There is a very real need to develop strategies that can withstand the test of time.

In a number of cases, the required retention period will exceed the economical and technical life of the archive media, necessitating the migration of archived data, possibly several times during its life. To enable these migrations to be carried out safely and efficiently, presupposes that at the point of archiving the data structure, format classification and associated metadata all facilitate future migrations and eventual deletion. These migrations and deletions are likely to be undertaken by persons that would have had no involvement in, or knowledge of, the original data. A further consideration is that the archiving requirements may be subjected to future changes, requiring that the data classification and metadata contain sufficient details to enable the correct actions to be taken.

4.1.10 The Data Avalanche

The rate at which electronic records are generated is exponential. Even allowing for the rate of growth of available mass storage capacity, the ability to manage the resulting archives, and in particular, to filter what data will go into the archive and to manage its eventual removal or deletion, clearly presents considerable challenges. A possibly greater challenge is to turn this mass of data into meaningful and useful information.

4.1.11 Data Criticality over Time

The value of any given record is likely to decrease with time. The rate of this will vary according to the nature of the record. The majority of records have little value beyond the immediate circumstances in which they are created. These records are unlikely to ever justify being passed from the live system into any form of long-term archive, but present the challenge of being able to be filtered out from any archiving process. For those records that are transferred into an archive, the challenge is to develop mechanisms to allow for these records to be removed from the archive as their usefulness decreases and regulations permit.

A risk analysis may be required when determining what to archive and what is feasible and acceptable to do. In this context, the criticality of the data will be central. From a regulatory perspective, data is either critical (i.e., required by regulation or used for regulatory purposes) or not, and its criticality will remain constant over the required retention period.

From a business perspective, the criticality of the stored data is likely to decline over the retention period. The older the data, the less likely it is that the data will be needed and its value also is likely to decline. For example, the benefit from data mining will become less as the data ages and newer, more representative, multi-faceted, and accurate data becomes available. The concept of decreasing criticality over time may be used as a component of a risk-based approach to compliance.

Clinical trial data may be used as an example. Drug recalls are more likely in the earlier stages of the licensed use of the drug. While the need remains to securely maintain the relevant information for the required period, it could be argued that certain loss of less critical metadata after a period of time represents an acceptable risk. After a long period of use of the drug, the value of clinical trial data is likely to diminish, as data from patient use of the drug, through adverse event reporting, provides more comprehensive information on a wider scale.

It is likely that a record may need to be migrated at least once during its retention period. Each migration carries a risk to the content and structure of a record and potential loss of metadata. Taking the concept of reduced usefulness over time into account enables the migration risk to be set against the usefulness of the record, perhaps making rendition to, e.g., PDF more acceptable.

4.2 Migration of Records/Data

Migration of records or data involves the movement (usually in large scale) of records or data between systems, and this typically comprises two key elements:

1. Data Migration, i.e., the applied control process for executing the movement of records or data.
2. Format Migration, i.e., the transformation of various data formats (see Section 10.3 of this Guide).

Both elements need to be validated. Similar validation considerations apply to the migration of records or data as to the validation of archive processes. It is recommended that established processes for data and format migration are used. Even so, such processes are likely to need to be validated, unless there is substantial documentary evidence that the processes work under the defined operating conditions that apply to the EDA and its interfacing systems.

Format migration will be required when the EDA interfaces to another system for the purpose of archiving/retrieving data and the data formats between the EDA and interfacing system differ. The following scenarios could give rise to a need for format migration:

- Archiving data into the original EDA where the formats differ
- Data sources are upgraded/replaced
- The EDA is changed/upgraded
- External demands requiring data formats to be changed

Specific considerations for format migration have been listed in Table 4.1.

Table 4.1: Format Migration Elements

Element	Consideration
Method	Definition, repeatability, robustness, and verification of process
Content	Securely protected from unintentional changes in data format or platform
Metadata	Retention of vital functionality (search, calculate, trend, etc.)
Context	Retention of vital metadata, data memoranda, and links that clarify context of data

Data and format migration are often unavoidable and may require some compromises. In transforming the data, some metadata may be affected, meaning a loss of, e.g., the ability to perform calculations using the data. Format migration also may result in the addition of data elements from the new environment. These will usually be allocated with their default settings, and it may be relevant that the default values are reviewed. It is important that such considerations are addressed prior to any data migration, and are captured in the migration method. The Data Owner is central to approving the migration method to be used.

4.3 Deletion of Records/Data

Deletion of archived data at the end of the retention period is just one of the archive processes. However, because this process is, in many cases, irrevocable and involves the deletion of GxP critical data, it warrants special consideration.

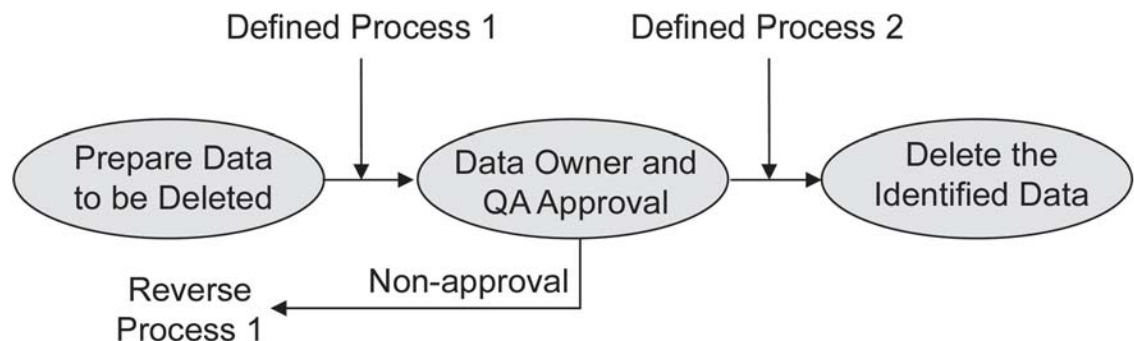
There are several driving forces for wanting to delete data:

- End of regulated retention period
- Non-regulated data no longer applicable
- To remove the management burden of records that no longer need to be retained. The cost of discovering each electronic record during litigation processes is significant, and increases in proportion to the volume of records stored. Therefore, limiting the volume of records in a controlled manner saves considerable cost, and also optimizes the time taken to retrieve records.
- To save disk space and to speed up data processing, e.g., backup and search functions
- To remove confidential or financial records that are no longer required for regulatory purposes, and which present a business risk to the organization if they were to be retained and accessed by unauthorized individuals.

The process used for the deletion of data may be automated or contain a manual element and should take account of the following factors:

- A reverse step method is recommended to enable the deletion to be reversed in the case of serious failure.
- Irrespective of whether the deletion process is automated or manual, it should be validated to include test cases that prove that only the appropriate records have been deleted. Where an organization operates a fully automated deletion process, stringent validation of the process should be carried out. In addition, the automated process should contain in-built verification of the correct operation of the process with error notification. The performance of the deletion process should further be regularly reviewed for correct operation.
- Any 'litigation or legal holds' must prevent the data from being deleted, and a check of any legal hold status must be part of the process for the deletion of data.
- The Data Owner is responsible for ensuring that the correct data is deleted, i.e., that data still required by regulation or the business is not deleted. This means that the Data Owner has a central role to play in the deletion of data.

Due to the criticality of the process, it is recommended that in addition to approval by the Data Owner that there is a QA approval for deletion of GxP data. Additional approval by the legal department may also be put in place. Such approvals should be at a stage in the deletion process before the data is irrevocably lost. This may imply a two-stage process, as depicted in Figure 4.1.

Figure 4.1: Delete Process

In Process 1, the data to be deleted is identified from the rest of the archive, and the case for deletion (rationale) is prepared.

In Process 2, the data deletion is carried out and verified as complete. This stage cannot be executed without a Data Owner and QA approval. Validation will ensure that the process is carried out safely and consistently.

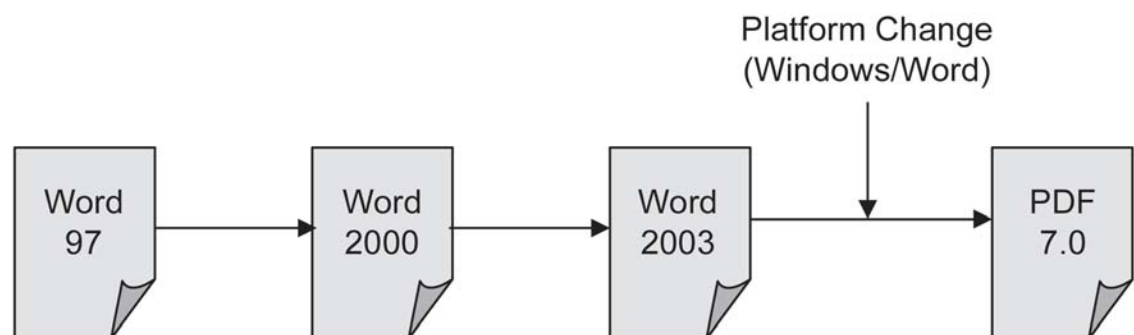
Bear in mind that it may not be possible to technically delete some data. In this scenario, the risks from retaining the data should be assessed. This risk may extend through to the decommissioning of the archive, and may call on the physical destruction of the storage media to prevent unauthorized access to the data.

4.4 Exit Strategy

Presently, there is no commercial realization of an EDA that can archive data with a high degree of certainty for several decades. The reason for this has already been discussed, i.e., changing technology making today's solutions redundant in years to come, thus making it difficult to fully access and understand all archived data, e.g., database records.

The driving forces for using an EDA are considerable, thus making a delay in implementation undesirable and probably not feasible. Taking these two contradictory factors into account, it is recommended that an exit strategy is defined for each EDA. An exit strategy can be defined as the fall-back position when the record comes to its end of technological support.

An example to illustrate what is meant by an exit strategy is outlined in Figure 4.2.

Figure 4.2: Exit Strategy (Example)

An EDA is used for storing documents generated using MS Word97® that also incorporate graphics and screen shots. During the archiving retention period the application will be up-dated to the latest version of the same family of applications (Windows/Word). It is conceivable that sometime in the future, there will be a platform change that does not allow comprehensive up-date of legacy data without corruption and/or unforeseen consequences. In this case, the exit strategy would be to transform the data to a PDF file. Such a file will continue to retain the content of the Word file, but will lose some of the attributes.

Having an exit strategy means thinking ahead to a worst case scenario, where the archived data can no longer be supported. In some cases, this may mean converting to PDF for example, which is likely to be able to be supported for a long period of time. The exit strategy for PDF is to print to paper although this is undesirable for obvious reasons.

Any exit strategy is likely to have drawbacks, such as loss of data properties. Therefore, it is important that the exit strategy is defined at the time of archiving so the potential consequences of EDA are understood before a commitment is made.

One way of simplifying the exit strategy, and minimizing the risk of losing technological support, is to avoid proprietary formats. Data content and type that conform to international standards and/or industry practice are likely to enjoy longer support. There is also an increased likelihood of more comprehensive solutions for transforming the data once the end of the current platform is reached.

While an exit strategy may mean a certain degree of loss of information, it should be set against the advantages of EDA. The extreme alternative to electronic archiving is to commit to paper with an immediate loss of information such as a significant amount of the metadata (although some of this may be printed). In addition, the exit strategy is a worst case scenario, i.e., it is possible that future technology is developed that makes platform changes both feasible and practical to use.

4.5 Retirement of the Archive System

Once the EDA comes to the end of its lifecycle, the archive data will need to be migrated. This process is irreversible and it is recommended that:

- The retirement process is defined and validated. The EDA is likely to be connected to other systems, and its retirement could potentially affect the operation of those systems.
- Verification that archive data has been migrated as required is performed.
- The archive is investigated for any remaining data, and this is either migrated or deleted. The archive may contain metadata that does not need to be retained.
- Documentation of the final configuration of the archive is retained.
- Approvals for retirement are obtained from all stakeholders, e.g., Management, Data Owners, and Quality.
- The retirement process is documented and affected systems are verified for correct operation.

4.6 Special Considerations

This Section discusses other aspects of archiving to consider, including:

- Raw data
- Hybrid records
- Audit trails
- Electronic signatures
- Databases
- Laboratory systems

4.6.1 Raw Data

Careful evaluation is necessary if the data both constitutes raw data and is required to be kept by a Predicate rule, e.g., 21 CFR 58.190 “Storage and Retrieval of Records and Data.” This does not preclude transfer and deletion of the raw data. However, the measures taken to ensure data integrity should be evaluated and assessed for the potential risks. The evaluation should ensure that:

- Modification of the data would not be possible by normal means
- Any modification of the data would be detectable

Measures should include:

- Automation of the raw data transfer/deletion
- Verification that the transferred copy is accurate and complete
- A secure transit mechanism (to preclude modification during transit)

For example, during a clinical study, the study director identifies all raw data (including electronic data) generated during the study. At the end of the study, access to all raw data generated during the study is archived. Access to, and responsibility for, the raw data is transferred from the Study Director to another responsible person (or organization) and that this individual (or organization) are custodians of the data on behalf of the test facility/site.

4.6.2 Hybrid Records

Hybrid records contain links between the two parts of the record (e.g., in the case of a paper/electronic hybrid, the electronic part and the paper part). These enable the integrity of the record to be maintained. It is essential that these links are retained after archiving. For the paper/electronic hybrid, a possible solution might be to convert one of the records so that both are in either electronic or paper format. So, the paper record could be converted to PDF and stored as metadata, or the electronic version to paper, subject to content and meaning being preserved. Wet-ink signatures on paper should be retained and if the related electronic record should be archived as such, procedures are required to ensure that the link is retained.

Note: Wet-ink signed paper can be replaced by a scanned image, and for high risk documents, a checksum of the scanned image can be applied or it can be re-signed digitally.

4.6.3 Audit Trails

An electronic audit trail typically captures changes to the data and includes new data details, old data details, who initiated/approved the change, the date and time of change, and the reason/context for the change. Normally, the change to an individual piece of data is not isolated, but relates also to other data and events. When the data is archived or migrated, this raises some potential issues with regard to retaining the integrity of the audit trail and its full context and interaction with the other data and events. When deciding what to archive, a risk analysis should be performed to determine the criticality of the audit trail items.

When addressing these issues, it may be helpful to first analyze the following different scenarios that an audit trail may be required to cover:

1. Authorized scheduled events such as entries in a batch record
2. Authorized unscheduled events such as software bug fixes
3. Unauthorized events such as inadvertently changing a measured value or fraudulent changes

Scenario 1 should be covered by an automated recording of the GxP-critical events. This may be data that is presented in a batch report. Where it is not feasible to have an automated audit trail, manual recording of these events may be acceptable, i.e., a hybrid system.

Scenario 2 is often handled through a manual change control system, where changes are recorded either by hand or through various electronic applications.

Scenario 3 is the one that is least suitable for manual recording, particularly to deter fraud. However, an electronic audit trail will never prevent fraud, only record what took place; therefore, stringent access controls may be more applicable in addressing scenario 3.

It is recommended that the chosen method for capturing changes, whether through an automated audit trail, manual procedural change control, enhanced access controls, or a combination of these, should be commensurate with risk.

Where an action originally was deemed sufficiently critical to audit trail, then that audit trail should be archived. Otherwise, formal justification for not archiving the audit trail should be included within project documentation. However, consideration should be given to whether any decision not to archive the audit trail would impact on the perceived integrity of the record to which the audit trail relates.

Where the audit trail is part of the Preservation Description Information, as discussed in Appendix B, Section 6.1.3 of this Guide, it would be archived with the Content Information as an Archival Information Package. In this respect, the audit trail is no different to other Preservation Description Information.

The problem comes where the audit trail is not fully contained within the data to be archived, i.e., the audit trail is generated on a system level taking in a number of data objects. In this case, the Data Management function should contain the necessary information to link the various Archival Information Packages to the relevant audit trail.

Where the audit trail is a hybrid record, e.g., a mixture of electronic and paper records, the Data Management function also may need to reference data held in paper form outside the EDA, which is difficult to manage. Alternatively, the paper records may be scanned into the EDA to enable it to be held electronically.

If the EDA itself permits the modification of the archived data, this too may need to be captured in an audit trail and many of the foregoing considerations will be equally applicable to this particular audit trail.

4.6.4 Electronic Signatures

Where data to be archived is held under an electronic signature, the archiving process should maintain the integrity of this electronic signature. The issues associated with this will vary with the mechanisms employed to secure the original signature. For example, signatures that are designed to be incapable of being copied/transferred from one environment to another may be extremely difficult to archive.

There are special considerations when dealing with the preservation of the content, context, and structure of records that are augmented by Electronic Signatures:

Content: the electronic signature or signatures in a record are part of the content. They indicate who signed a record, the meaning of the signature and when the signature was applied. Multiple signatures can indicate initial approval and subsequent concurrences. In order to meet regulatory requirements, signatures should include the identity, date/time and meaning. All of this is part of the content of the record and should be preserved. Lack of this information may affect a document's reliability and authenticity, and hence, its acceptance by regulatory authorities.

Context: some electronic signature technologies rely on individual identifiers that are not embedded in the content of the record, trust paths, and other means to create and verify the validity of an electronic signature. This information is outside the content of the record, but is important to the context of the record, as it provides additional evidence to support the reliability and authenticity of the record. Lack of these contextual records prevents the ability to verify the validity of the signed content.

Structure: preserving the structure of a record means its physical and logical format, and the relationships between the data elements comprising the record, remain physically and logically intact to ensure that the context and meaning of the record are retained. This applies equally to any associated electronic signatures.

Technically complex electronic signatures will not migrate well if the decryption or digital signature-processing services have been decommissioned. Such signatures can become unreadable, but are not a regulatory requirement or expectation. A simple electronic signature would be executed by ID/password and involve the change of a status field with capture of name, date/time, and purpose. Appropriate security would prevent the misuse of this field. This simple signature would easily be migrated to new formats and could be expressed in XML to maintain its readability.

Options for archiving those records with an electronic signature that cannot be migrated into the archive include:

- Using a trusted third party or a certification authority to verify the signature on the original record(s) and to apply a new signature to the migrated record(s).
- In the case of documents, delegate authentication of documents to the document management system so that the system routinely verifies the authenticity of a signed document, while the signature is still valid, and records the fact by itself signing the document as having been verified (using a method that meets long-term archival requirements). This approach is based on established practice for other means of authentication that cannot themselves be archived, but can be recorded as having occurred (by a trustworthy source).

The key requirement is for the content, meaning, and identity of the signature to be preserved.

4.6.5 Databases

Databases are used extensively across all areas of regulated industry to manage electronic records and data. Systems vary in size and scope from Enterprise Resource Planning Systems (ERP), corporate accounting packages, and departmental Laboratory Information Management Systems (LIMS) to a single user laboratory data acquisition and processing system.