ISPE® GAMP®

GAMP Good Practice Guide:

Electronic Data Archiving

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.*

ISBN 1-931879-87-7

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 confidentially 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:

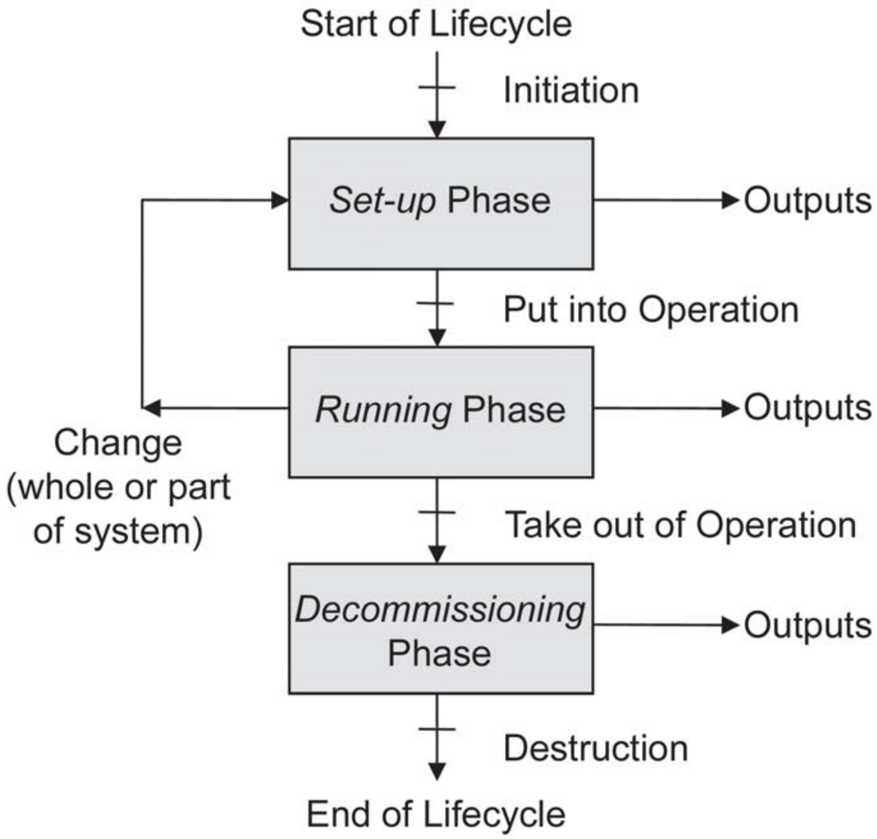
1. The Set-up Phase relates to the specification, development, testing, and verification of the system.

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.)

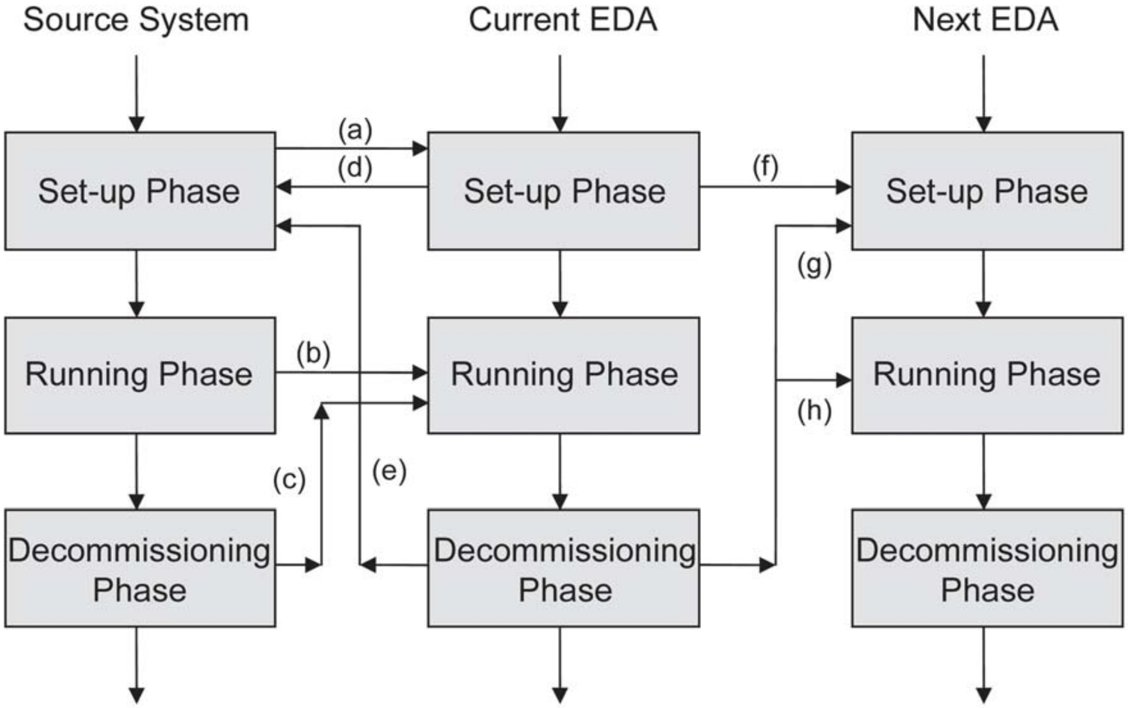
The purpose of the Archiving Strategy document is to attempt to document all necessary considerations for an EDA. It is recognized that there are unlikely to be commercial solutions for all identified issues, but by formulating them, an understanding of the demands and their implications can be gained. 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 continuingly 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:

a. 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.

b. Source system passing data to the EDA in the course of normal running.

c. 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.

d. 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.

e. 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

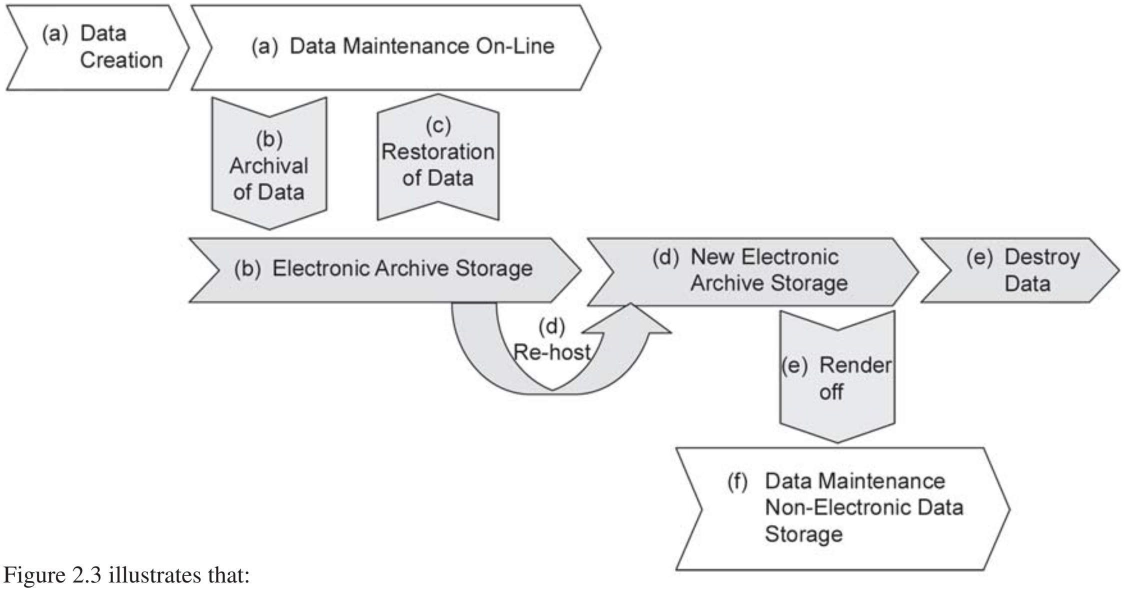

(a) Data Creation
(a) Data Maintenance On-Line
(b) Archival of Data
(c) Restoration of Data
(b) Electronic Archive Storage
(d) New Electronic Archive Storage
(e) Destroy Data
(d) Re-host
(e) Render off
(f) Data Maintenance Non-Electronic Data Storage

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

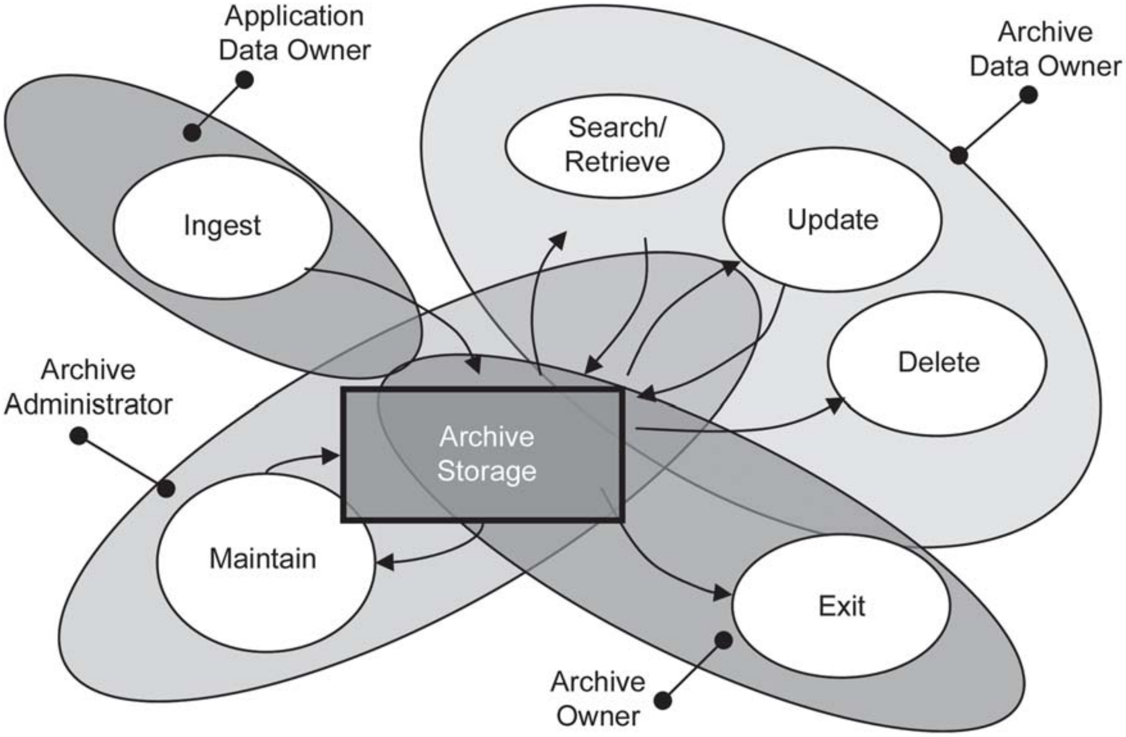


Table 2.1 defines the identified archiving processes.

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

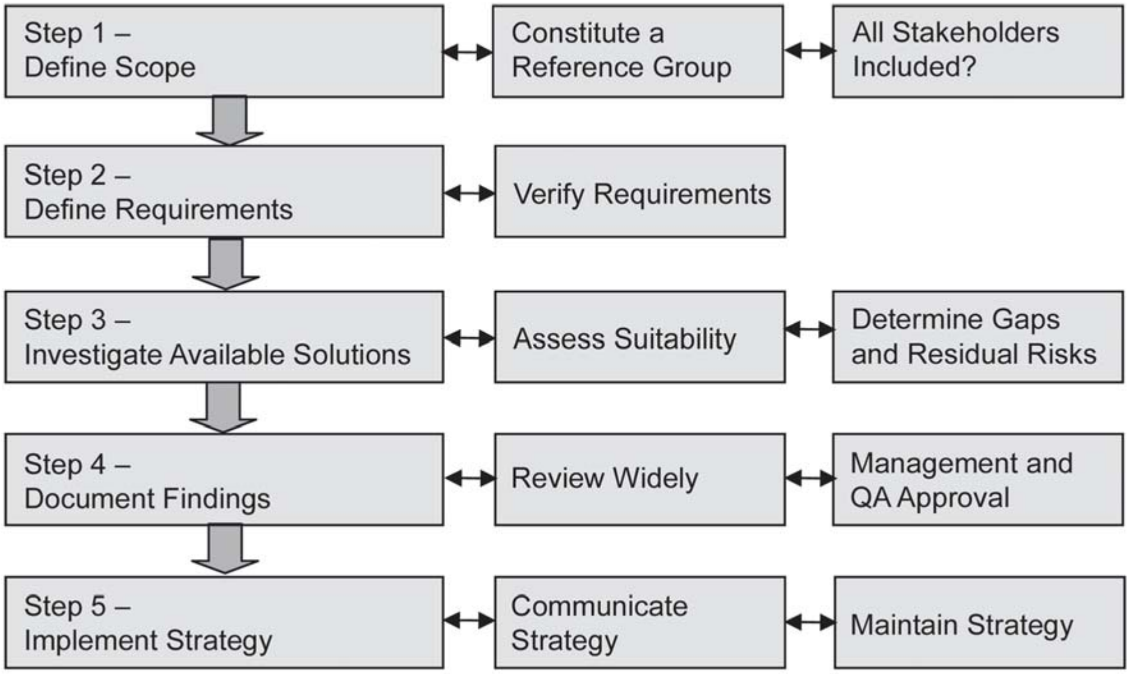
3.1 Overview

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
Constitute a Reference Group
All Stakeholders Included?
Step 2 – Define Requirements
Verify Requirements
Step 3 – Investigate Available Solutions
Assess Suitability
Determine Gaps and Residual Risks
Step 4 – Document Findings
Review Widely
Management and QA Approval
Step 5 – Implement Strategy
Communicate Strategy
Maintain 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.

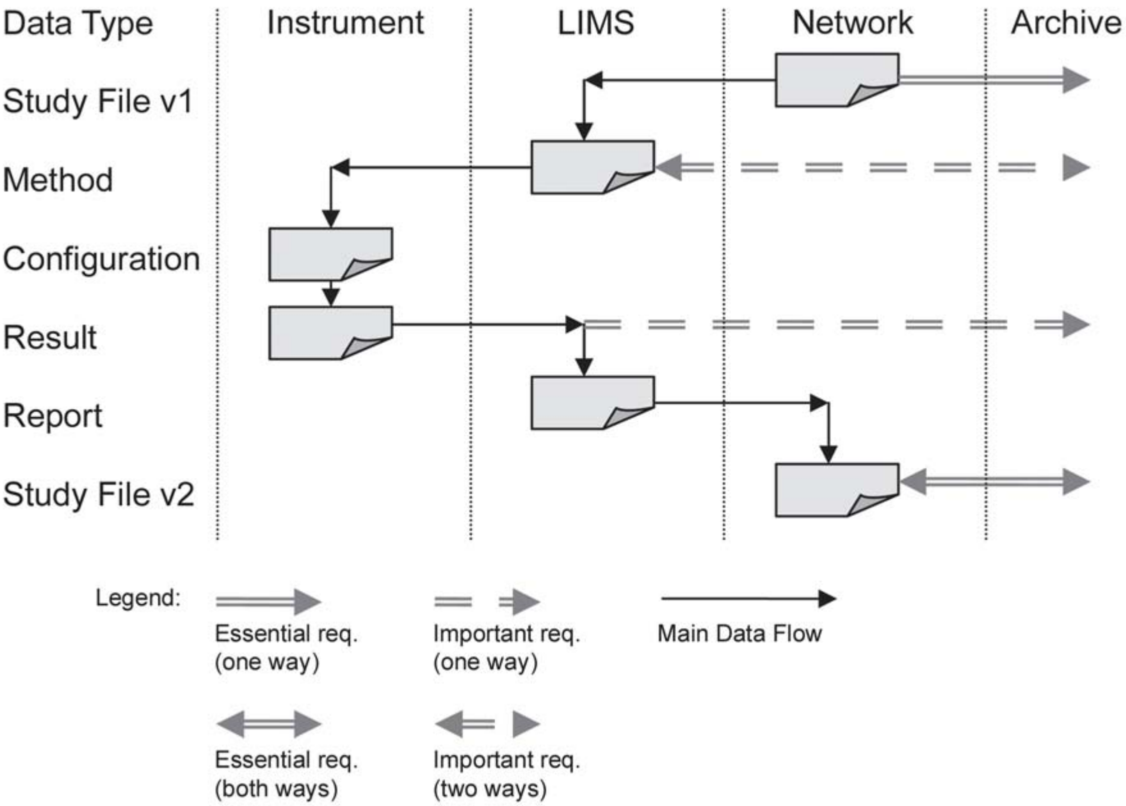
The strategy may be one that involves several transitional stages, as the envisaged technology and resources become available. When allocating roles and responsibilities, consideration should be taken of likely changes in the short and medium term, to prevent unnecessary changes in personnel due to inadequate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

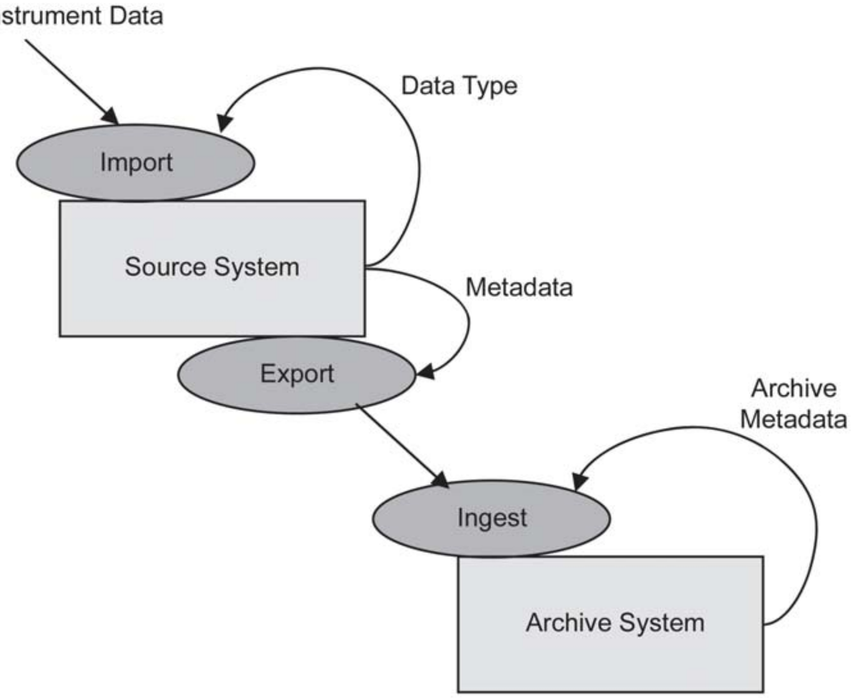
The data flow diagram can be further enhanced by adding job roles that control or approve the data. This may highlight potential change of ownership issues that need to be addressed. For example, in Figure 3.2, ownership of data may remain with the Study Director throughout the process, but also may 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 to 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 complaint 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.

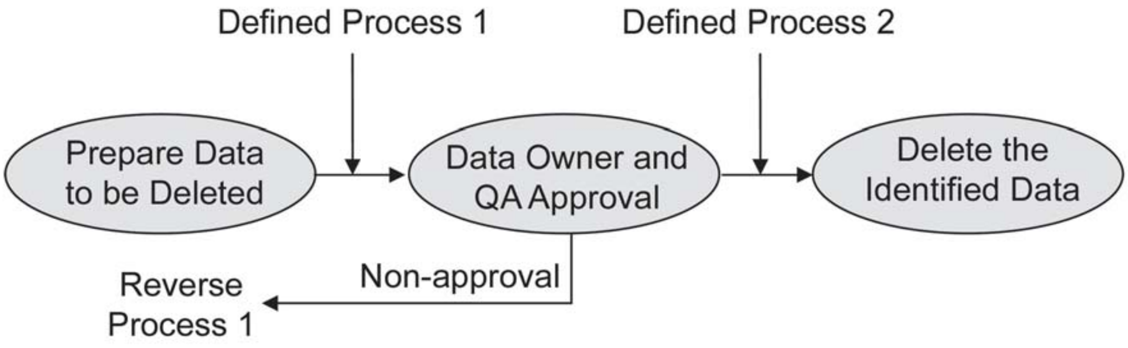
- Irrespective of whether the deletion process is automated or manual, it should be validated to include test cases that prove that only the appropriate records have been deleted. Where an organization operates a fully automated deletion process, stringent validation of the process should be carried 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


Defined Process 1
Defined Process 2
Prepare Data to be Deleted
Data Owner and QA Approval
Delete the Identified Data
Reverse Process 1
Non-approval

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.

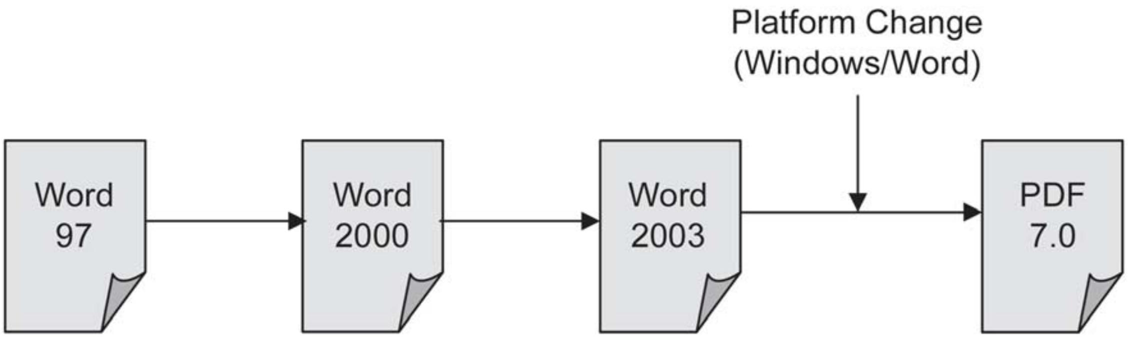
4.4 Exit Strategy

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.

Commercial database management systems make it straightforward to archive data. However, in a typical database application, the application logic is required for reprocessing of data. Where reprocessing of archived data is required this may be difficult since archiving of the application logic without resorting to the computer museum approach is problematic.

Various technologies and strategies are available; there is no single recommended approach.

Further analysis of the issues associated with databases can be found in Appendix D of this Guide. A key conclusion of this analysis is that the optimal approach is unlikely to be available for Commercial Off-The-Shelf (COTS) systems without some degree of software development.

4.6.6 Laboratory Systems

The majority of analytical instruments used in laboratories today (e.g., HPLC, GC, and UV Spectrophotometer) produce detector response profiles that require complex software to process the data and present it in a form suitable for review and interpretation.

While the probability and risk associated with the retrieval and reprocessing of the data generated by these instruments will diminish with time, the need to retrieve the data throughout the associated data retention period remains an issue. This is aggravated by the fact that many of the data formats used by the various instrument vendors to-date are proprietary, making it impossible to decode the data without the vendor’s original software.

A number of possible options exist for addressing this issue, including:

- Commercial Scientific Data Management Solutions (SDMSs) permit the automation of the capture, storage, searching, and retrieval of Electronic Records.

- Commercial SDMSs consolidate the software required to re-visualize data from multiple instruments/applications down to one application although the consolidated application is itself likely to adopt a proprietary format.

- Conversion to a technology/vendor-neutral format. Two widely recognized standards exist for HPLC and Mass Spectrometry, namely AnDI and JCAMP, although a new U.S. standard is in development that includes a new XML schema to replace these two formats. Various developments based on XML have been investigated, including the XML-based Analytical Information Mark-up Language (AnIML) standard.

The success of any standard approach is reliant upon a variety of factors, including:

- The schema definitions being publicly accessible and available for adoption by all vendors.

- The vendors themselves commit to the standards and include functionality to export/import records.

Further details of Laboratory Systems can be found in Appendix E of this Guide.

Appendix A

Example Regulatory Requirements

5 Appendix A: Example Regulatory Requirements

This Section sets out further details pertaining to the regulatory requirements. It is not intended to be a complete reference, but provide some additional information to support the statements made within the body of the document. It is split into two parts, as follows:

- Specific Regulatory Topics. This highlights a selection of key regulatory requirements and is structured according to subject area.

- Regulatory References. This provides an overview of the data archiving requirements set out in a number of key regulations and is structured according to the individual regulation.

The first part provides an illustration of the key requirements and the second part provides more regulatory details. Whilst there is some clear overlap between the two parts they are not intended to fully map on a one-to-one basis.

For easy reference, requirements have sometimes been summarized, rather included in full and where more than one regulation has quoted the same requirement only one reference is listed.

Note that this Guide mainly focuses on relevant GxP regulations. However, due to their particular relevance, the Sarbanes-Oxley Act of 2002 and EU/EEA Data Protection Directive also are referenced. There are many other regulations, such as those related to the handling of substances harmful to health, which may impose additional record retention requirements to those covered in this document.

5.1 Specific Regulatory Topics

This Section highlights the key requirements for the following topics:

- Retention periods

- Data records

- Access, security, and environment

- Maintenance/operation

- Sarbanes-Oxley Act of 2002

- EU/EEA Data Protection Directive

The following source data has been used for this Section:

Table 5.1: List of Used References

Ref #	Reference
Ref 1	US Environmental Protection Agency Directive 2185: Good Automated Laboratory Practices.
Ref 2	OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 10 GLP Consensus Document The Application of the Principles of GLP to Computerized Systems Environment Monograph No. 116 (OECD/GD(95)115).
Ref 3	GLP Advisory Leaflet No. 1 The Application of GLP Principles to Computer Systems (UK GLP Compliance Program, DoH, London 1995).
Ref 4	OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1 OECD Principles on Good Laboratory Practice (as revised in 1997) ENV/MC/CHEM(98)17.
Ref 5	21 CFR Part 11 Electronic Signatures; Electronic Records.
Ref 6	45 CFR 160 and 164 (HIPAA Implementation Regulations).
Ref 7	2003/94/EC Annex 11 Computerised Systems.
Ref 8	21 CFR 58 Good Laboratory Practice for Non-Clinical Laboratory Studies.
Ref 9	Sarbanes-Oxley Act of 2002.
Ref 10	95/46/EC Data Protection Directive.

5.1.1 Retention Periods

The retention period, if not specified, should be sufficient to support any challenges to data integrity. [Ref 1]

Raw data, where stored, should be retrievable throughout the retention period. Provision should be made against the limited life of computer system, allowing for hardware and software changes. [Ref 2]

Note that it is sometimes necessary to retain data beyond its retention period where there is a possibility of it being required in connection with legal proceedings. [Ref 3]

In the absence of a required retention period, the final disposition of any study materials should be documented. [Ref 4]

Table 5.2 summarizes typical record retention requirements for pharmaceutical products and medical devices. Please note that the column typical product time line does not directly relate to the regulatory retention requirement, but is the estimated length of time for each particular activity in progression from R&D to product withdrawal.

Table 5.2: Typical Record Retention Requirements

Product Lifecycle Stage	Typical Product Timeline (Years)	Applicable Regulations	Need for Retaining Electronic Records	Record Retention Period	Records Required for
Discovery	2 - 10	Non-GxP (e.g., patents)	No	At least until regulatory approval obtained.	Marketing Authorization, Technology Transfer
Pre-clinical	2 - 4	GLP	Yes	At least until regulatory approval obtained plus 5 years.	Marketing Authorization, Technology Transfer
Clinical Trials	6	GCP	Yes	At least until regulatory approval obtained plus 5 years for the lifetime of the product.	Marketing Authorization, Technology Transfer; Definition of Therapeutic Application
Regulatory Review and Approval for Product Manufacturing	1 - 2	GMP GLP	Yes	For the lifetime of product.	Marketing Authorization
Launch and Marketing	> 1	GMP GDP	Yes	Minimum 1 year beyond end of shelf life of pharmaceutical product or minimum 3 years for pharmaceuticals (2 years for medical devices) after shipping, for products without defined shelf life. For at least 5 years from certification (EU) or release date (often for the lifetime of the product).	Marketing Authorization
End of Shelf Life	N/A	GMP GLP GDP	Yes	Minimum 1 year beyond end of shelf life of product. Minimum 5 years after last distribution of product.	Product Recall; Litigation
Product Discontinued	N/A	GMP GLP GDP	Yes	Minimum 2 years after discontinuation (EU). Minimum 7 years beyond discontinuation for litigation purposes.	Product Recall; Litigation

5.1.2 Data Records

Records should be protected to enable accurate and ready retrieval throughout the retention period. Readability is implied throughout retention period. [Ref 5]

Audit Trails and other Metadata required under predicate rules should be retained for at least as long as the original record and be available for review and copying. [Ref 5]

The following should be retained in the archives for the period specified by the appropriate authorities [Ref 4]:

- Study plan, raw data, samples of test and reference items, specimens, and the final report of each study
- Records of all inspections performed by the Quality Assurance Program, as well as master schedules
- Records of qualifications, training, experience, and job descriptions for personnel
- Records and reports of the maintenance and calibration of apparatus
- Validation documentation for computerized systems
- The historical file of all Standard Operating Procedures
- Environmental monitoring records

Electronic signature components are subject to the same controls as electronic records. [Ref 5]

Signature/record linking implied to be permanent throughout the retention period. [Ref 5]

Archived records containing any patient-related data should [Ref 6]:

1. Be confidential
2. Be searchable by patient identifiers
3. Store copies of or pointers to relevant authorizations
4. Be capable of amendment
5. Be capable of having records deleted
6. Support record expiry where authorizations have been given explicit expiry dates
7. Ensure that data is kept accurate and up-to-date
8. Only be kept in identifiable form for as long as is strictly necessary

Computerized system design should always provide for the retention of full audit trails to show all changes to direct computer input data, including reasons for the change, without obscuring the original data. [Ref 4]

No electronically stored data should be deleted without management authorization and relevant documentation. [Ref 2]

Supporting information such as maintenance logs and calibration records that are necessary to verify the validity of raw data or to permit reconstruction of a process or a study should be retained in the archives. [Ref 2]

Retain historical file of all versions of software with manuals, validation documentation, formulae and algorithms, and change control records. [Ref 1]

5.1.3 Access, Security, and Environment

Persons who use closed systems to create, modify, maintain, or transmit electronic records shall employ procedures and controls designed to ensure the authenticity, integrity, and when appropriate, the confidentiality of electronic records, and to ensure that the signer cannot readily repudiate the signed record as not genuine. Such procedures shall include the following:

- Limiting system access to authorized individuals [Ref 5]

- Use of authority checks to ensure only authorized individuals can use the system, electronically sign a record, access the operation or computer system input or output device, alter a record, or perform the operation at hand [Ref 5]

Archive facilities should be provided for the secure storage and retrieval of study plans, raw data, final reports, supporting material, samples of test items, and specimens. [Ref 4]

Attention should be paid to the citing of equipment in suitable conditions where extraneous factors cannot interfere with the system. [Ref 7]

Archived records should be stored in environmentally adequate storage with security against willful or accidental damage. [Ref 7]

Apparatus, including validated computerized systems, used for the generation, storage, and retrieval of data, and for controlling environmental factors relevant to the study should be suitably located and of appropriate design and adequate capacity. [Ref 4]

Only personnel authorized by management should have access to the archives. Movement of material in and out of the archives should be properly recorded. [Ref 4]

5.1.4 Maintenance/Operation

An individual shall be identified as responsible for the archives. [Ref 8]

If a test facility or an archive contracting facility goes out of business and has no legal successor, the archive should be transferred to the archives of the sponsor(s) of the study(s). [Ref 4]

Where problems with long-term access to data are envisaged or when computerized systems have to be retired, procedures for ensuring continued readability of the data should be established. This may, for example, include producing hard copy printouts or transferring the data to another system. [Ref 2]

Archives must be indexed to facilitate expedient retrieval of electronic data. [Ref 8]

It is worth noting that in some cases it may be possible to search and retrieve data in the absence of a well-defined index, as illustrated by the use of a search engine on the Internet. However, it is not possible to verify the accuracy of a search in terms of being 100% confident that it has captured all available records.

Stored data must be regularly checked for accessibility, durability, and accuracy. If changes are proposed to the computer equipment or its programs, the above-mentioned checks should be performed at a frequency appropriate to the storage medium being used. [Ref 7]

The GLP requirements for archiving data must be applied consistently to all data types. Therefore, it is important that electronic data are stored with the same levels of access control, indexing, and expedient retrieval as other types of data. [Ref 3]

SOPs must be established and followed to ensure long-term integrity and readability of data stored electronically is not compromised. [Ref 3]

Subjects for SOPs should include:

1. Responsibilities of the archivist [Ref 8]

2. The means of retrieving data after system retirement [Ref 1]

3. Backup/restore [Ref 7]

4. Computerized Systems Validation, operation, maintenance, security, change control, and backup [Ref 4]

5. Record keeping [Ref 4]

6. Reporting [Ref 4]

7. Storage [Ref 4]

8. Coding of studies [Ref 4]

9. Data collection [Ref 4]

10. Preparation of reports [Ref 4]

11. Indexing systems [Ref 4]

12. Handling of data, including the use of computerized systems [Ref 4]

Management should ensure that personnel are aware of the importance of data security, the procedures and system features that are available to provide appropriate security, and how to deal with the consequences of security breaches. Such system features could include routine surveillance of system access, the implementation of file verification routines, and exception and/or trend reporting. [Ref 2]

5.1.5 Sarbanes-Oxley Act of 2002

The Sarbanes-Oxley Act requires US listed public companies to provide more financial information than ever before and holds corporate directors and officers personally accountable for the accuracy of financial disclosures.

The Act requires internal controls to be in place within the organization to ensure accurate financial reporting, as well as an assessment of the effectiveness of these internal controls.

Although the act is directed specifically at financial reporting, it also covers:

"Whoever knowingly alters, destroys, mutilates, conceals, covers up, falsifies, or makes a false entry in a record, document, or tangible object with the intent to impede, obstruct, or influence the investigation or proper administration of any matter within the jurisdiction of any department or agency of the United States."

This could be interpreted to include protection of information held in archives that are required under predicate rules.

5.1.6 EU/EEA Data Protection Directive

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 covers the protection of individuals with regard to the processing of personal data and the free movement of such data. It is a complex and specialist area and the following Sections are only intended to give a brief overview of the most relevant Sections for archiving purposes.

This represents an overall EU requirement for data protection, mandatory on all EU member states to adopt into national law. It provides for similar obligations to HIPAA to use data only for specified purposes, and to allow Data Subject access to data to correct or have removed records, as appropriate.

The directive sets out the key principles to be applied, including the need to keep data accurate and up-to-date and the prohibition of holding data indefinitely. Data should only be kept in identifiable form for as long as is strictly necessary for declared purposes.

Although precise legal wording concerning the ability to identify varies from member state to member state, the typical definition of data being identifiable includes the ability to identify the data subject either directly from the data or from any other data in the possession of or likely to come into the possession of the data controller. This means that blinded clinical trials data, e.g., are considered identifiable while the ability to unblind it still exists.

Article 8 "Special Categories of Data" prohibits the processing of personal data revealing:

- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade-union membership
- Concerning health or sex life

The purposes for which the data are to be used should be declared and data given for one purpose may not be used for other purposes without further explicit consent. Therefore, great care should be taken over the uses to which archived data is put. It is advisable to keep copies of these consents with any archive, which relies upon them.

Article 12 "Subjects Right to Access" establishes the Data Subject's right to access data and to require correction or blocking of access, as appropriate.

Article 14 "Subjects Rights to Object" establishes a Data Subject's right to object to the processing of their data (and, by extension, require its deletion).

Taken together, Articles 12 and 14 require the ability for identifiable data in an archive to be:

- Searched on the basis of personal identification data

- Edited in such a way as to make the data anonymous

Article 18 “Obligation to Notify Regulatory Authority” establishes the obligation on Data Controllers to register the fact that they hold personal data and the purposes for which it is retained/used.

Article 25 “Transfers to Third Countries” establishes the requirements on Data Controllers to ensure that, if data is to be transferred outside of the EU, adequate legal protections exist in the target country. A very limited list of countries has been established by the EU as meeting these requirements. In particular, the status of the US is under question. Generally, legally enforceable contracts to adhere to data protection requirements are used to provide legal remedies for breaches. This is particularly significant for any organization establishing central corporate archives or who are archiving multi-national data, such as clinical trials.

The EU has published standard contractual terms which can be adopted for this purpose (EU Decision 2004/915/EC).

5.2 Regulatory References

During August/September 2003, a search was carried out for instances of archival, archive, storage and retention within the regulations listed below. A summary table was then prepared of the archiving requirements set out in these documents. It should be noted that this is not intended to be a complete list, merely a summary of some of the applicable regulations in force at the time the Guide was being prepared. The following regulations and regulatory guidance have been included:

1. 21 CFR Part 11 - Electronic Records, Electronic Signatures

2. 21 CFR Part 58 - Good Laboratory Practice for Non-Clinical Laboratory Studies

3. 21 CFR Part 210 - Current Good Manufacturing Practice in Manufacturing, Packing, or Holding of Drugs

4. 21 CFR Part 211 - Current Good Manufacturing Practice for Finished Pharmaceuticals

5. 21 CFR Part 820 - Quality System Regulation (Medical Device)

6. 45 CFR 160 & 164 - HIPAA implementation regulations

7. US Environmental Protection Agency Directive 2185 Good Automated Laboratory Practices

8. Rules and Guidance for Pharmaceutical Manufacturers and Distributors 1997 (U.K. Orange Guide)

9. Commission Directive 2003/94/EC for Good Manufacturing Practice

10. EU Guide to Good Manufacturing Practice

11. Directive 2001/83/EU on the Community Code Relating to Medicinal Products for Human Use

12. Guidelines on Good Distribution Practice of medicinal products for Human Use – 94/C 63/03

13. EUCOMED Position Paper on Good Distribution Practice

14. 95/46/EC EU Data Protection Directive

15. PIC/S Guide PI 011-1 20 August 2003 GOOD Practices for Computerized Systems in Regulated ‘GxP’ Environments

16. ICH GCP

17. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1 OECD Principles on Good Laboratory Practice (as revised in 1997) ENV/MC/CHEM(98)17

18. OECD Series on Principles of GLP and Compliance Monitoring Number 4 (Revised) Consensus Document, Quality Assurance And GLP, ENV/JM/MONO(99)20

19. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 10 GLP Consensus Document the Application of the Principles of GLP to Computerized Systems Environment Monograph No. 116 (OECD/GD(95)115)

20. Good Laboratory Practice Advisory Leaflet No. 1. The Application of GLP Principles to Computer Systems (U.K. GLP Compliance Program, DoH, London 1995)

21. EU Directive 2001/20/EC on the Implementation of Good Clinical Practice in the Conduct of Clinical Trials on Medicinal Products for Human Use

22. Sarbanes-Oxley Act 2002

Table 5.3: Regulatory References

Ref #	Regulation	Section
Ref: 1	21 CFR Part 11 – Electronic Records, Electronic Signatures	Part 11.1(b)
	Summary: Scope includes records being maintained, archived, and retrieved under FDA regulation.	
Ref: 1	21 CFR Part 11 – Electronic Records, Electronic Signatures	Part 11.3(b) (6)
	Summary: Definition of electronic record encompasses records in digital form being maintained, archived, and retrieved.	
Ref: 1	21 CFR Part 11 – Electronic Records, Electronic Signatures	Part 11.10(c)
	Summary: Preamble Comments 30 and 71: Protection of records to enable their accurate and ready retrieval throughout the records retention period. Readability is implied throughout retention period by the need to retain suitable computer systems or migrate records to new systems.	
Ref: 1	21 CFR Part 11 – Electronic Records, Electronic Signatures	Part 11.10(e)
	Summary: Audit trail to be retained for at least as long as the record and to be available for FDA review and copying.	
Ref: 1	21 CFR Part 11 – Electronic Records, Electronic Signatures	Part 11.50(b)
	Summary: Electronic signature components subject to same controls as electronic records.	
Ref: 1	21 CFR Part 11 – Electronic Records, Electronic Signatures	Part 11.70
	Summary: Signature/record linking implied to be permanent throughout retention period.	
Ref: 2	21 CFR Part 58 – Good Laboratory Practice for Non-Clinical Laboratory Studies	58.33(f)
	Summary: All raw data, documentation, protocols, specimens, and final reports are transferred to the archives during or at the close of the study.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 2	21 CFR Part 58 – Good Laboratory Practice for Non-Clinical Laboratory Studies	58.51
	Summary: Space shall be provided for archives, limited to access by authorized personnel only, for the storage and retrieval of all raw data and specimens from completed studies.	
Ref: 2	21 CFR Part 58 – Good Laboratory Practice for Non-Clinical Laboratory Studies	58.190(b)
	Summary: There shall be archives for orderly storage and expedient retrieval of all raw data, documentation, protocols, specimens, and interim and final reports. Conditions of storage shall minimize deterioration of the documents or specimens in accordance with the requirements for the time period of their retention and the nature of the documents or specimens. A testing facility may contract with commercial archives to provide a repository for all material to be retained. Raw data and specimens may be retained elsewhere provided that the archives have specific reference to those other locations.	
Ref: 2	21 CFR Part 58 – Good Laboratory Practice for Non-Clinical Laboratory Studies	58.190(c)
	Summary: An individual shall be identified as responsible for the archives.	
Ref: 2	21 CFR Part 58 – Good Laboratory Practice for Non-Clinical Laboratory Studies	58.190(d)
	Summary: Only authorized personnel shall enter the archives.	
Ref: 2	21 CFR Part 58 – Good Laboratory Practice for Non-Clinical Laboratory Studies	58.190(e)
	Summary: Material retained or referred to in the archives shall be indexed to permit expedient retrieval.	
Ref: 2	21 CFR Part 58 – Good Laboratory Practice for Non-Clinical Laboratory Studies	58.195(b)
	Summary: Except as provided in paragraph (c) of this Section, documentation records, raw data, and specimens pertaining to a non-clinical laboratory study and required to be made by this part shall be retained in the archive(s) for whichever of the following periods is shortest: (1) A period of at least 2 years following the date on which an application for a research or marketing permit, in support of which the results of the non-clinical laboratory study were submitted, is approved by the Food and Drug Administration. This requirement does not apply to studies supporting Investigational New Drug applications (INDs) or applications for Investigational Device Exemptions (IDEs), records of which shall be governed by the provisions of paragraph (b)(2) of this Section. (2) A period of at least 5 years following the date on which the results of the non-clinical laboratory study are submitted to the Food and Drug Administration in support of an application for a research or marketing permit. (3) In other situations (e.g., where the non-clinical laboratory study does not result in the submission of the study in support of an application for a research or marketing permit), a period of at least 2 years following the date on which the study is completed, terminated or discontinued.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 2	21 CFR Part 58 – Good Laboratory Practice for Non-Clinical Laboratory Studies	58.195(h)
	Summary: If a facility conducting non-clinical testing goes out of business, all raw data, documentation, and other material specified in this Section shall be transferred to the archives of the sponsor of the study. The Food and Drug Administration shall be notified in writing of such a transfer.	
Ref: 3	21 CFR Part 210 – Current Good Manufacturing Practice in Manufacturing, Processing, Packing, or Holding of Drugs	N/A
	Summary: There are no references to the word archive, archival, storage, or retention in this document.	
Ref: 4	21 CFR Part 211 – Current Good Manufacturing Practice for Finished Pharmaceuticals	Subpart J, Sec 211.180 (b)
	Summary: Records shall be maintained for all components, drug product containers, closures, and labeling for at least 1 year after the expiration date or, in the case of certain OTC drug products lacking expiration dating because they meet the criteria for exemption under Sec. 211.137, 3 years after distribution of the last lot of drug product incorporating the component or using the container, closure, or labeling.	
Ref: 4	21 CFR Part 211 – Current Good Manufacturing Practice for Finished Pharmaceuticals	Subpart J, Sec 211.180 (c)
	Summary: All records required under this part, or copies of such records, shall be readily available for authorized inspection during the retention period at the establishment where the activities described in such records occurred. These records or copies thereof shall be subject to photocopying or other means of reproduction as part of such inspection. Records that can be immediately retrieved from another location by computer or other electronic means shall be considered as meeting the requirements of this paragraph.	
Ref: 4	21 CFR Part 211 – Current Good Manufacturing Practice for Finished Pharmaceuticals	Subpart J, Sec 211.180 (d)
	Summary: Records required under this part may be retained either as original records or as true copies such as photocopies, microfilm, microfiche, or other accurate reproductions of the original records. Where reduction techniques, such as microfilming, are used, suitable reader and photocopying equipment shall be readily available.	
Ref: 5	21 CFR Part 820 – Quality System Regulation (Medical Device)	Subpart M, Sec. 820.180 (b)
	Summary: General requirements: (b) Record retention period. All records required by this part shall be retained for a period of time equivalent to the design and expected life of the device, but in no case less than 2 years from the date of release for commercial distribution by the manufacturer.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 6	45 CFR 160 and 164 (HIPAA implementation regulations)	Overview
	Summary: Standards for Privacy of Individually Identifiable Health Information; Final Rule. Provides specific controls around the storage and use of identifiable healthcare data. Most production-related data is likely to fall outside of the scope of the regulations; most likely areas of impact are clinical trials, implantable medical devices sectors, and patient data at product complaint/adverse event intake. Key Relevant Provisions: Patients must give explicit permission to holding of certain data and to the uses to which it can be put. Authorizations must have an expiry date on them and cease to be valid after that date." Covered entities" must only allow agreed uses. Patients have the right to inspect and, if appropriate, correct data. Patients may withdraw authorization for use.	
Ref: 6	45 CFR 160 and 164 (HIPAA implementation regulations)	Overview
	Summary: This means that archives containing such data must be: Searchable by patient identifiers. Store copies of, or pointers to, relevant authorizations. Capable of amendment. Capable of having records deleted/redacted. Support record expiry where authorizations have been given explicit expiry dates. Note: It is not likely that a pharmaceutical company/medical device manufacturer will themselves be covered entities under HIPAA, but rather that compliance with HIPAA provisions will be a contractual requirement in order to allow the flow of information from healthcare providers/investigators.	
Ref: 6	45 CFR 160 and 164 (HIPAA implementation regulations)	160.103 and 164.514
	Summary: Defines Identifiable Information. Defines how to de-identify data. Taken together provide the criteria from whether data could be considered to ever have been identifiable protected health information, and to confirm whether sufficient work to make data anonymous has taken place.	
Ref: 6	45 CFR 160 and 164 (HIPAA implementation regulations)	160.103 and 164.514
	Summary: In the absence of an organization knowing of a way to identify individuals from a dataset, data which contains none of the following direct identifiers of the individual or of relatives, employers, or household members of the individual, can be considered anonymous: (i) Names. (ii) Postal address information below State; (first 3 digits of ZIP code allowed if designates more than 20,000 people). (iii) All elements of dates other than year for those under 90 – all 90+ in a single age category. (iv) Telephone and Fax numbers. (v) Electronic mail addresses. (vi) Social security numbers. (vii) Medical record numbers. (viii) Health plan beneficiary numbers. (ix) Account numbers. (x) Vehicle identifiers and serial numbers, including license plate numbers. (xii) Device identifiers and serial numbers. (xiii) Web Universal Resource Locators (URLs). (xiv) Internet Protocol (IP) address numbers.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 6	45 CFR 160 and 164 (HIPAA implementation regulations)	160.103 and 164.514
	Summary: (xv) Biometric identifiers, including finger and voice prints. (xvi) Full face photographic images and any comparable images. (xvii) Any other unique identifying code or number.	
Ref: 64	5 CFR 160	Section 306
	Summary: Ability of any person to make formal complaints.	
Ref: 6	45 CFR 164	Section 501
	Summary: Definition of designated record set. Designated record set means: (1) A group of records maintained by or for a covered entity that is: The medical records and billing records about individuals maintained by, or for, a covered healthcare provider. The enrollment, payment, claims adjudication, and case or medical management record systems maintained by, or for, a Health Plan. Or Used in whole, or in part, by or for the covered entity to make decisions about individuals. (2) For purposes of this paragraph, the term record means any item, collection, or grouping of information that includes protected health information and is maintained, collected, used, or disseminated by or for a covered entity.	
Ref: 6	45 CFR 164	Section 508(b) (5)
	Summary: Allows for the revocation of authorizations.	
Ref: 6	45 CFR 164	Section 514
	Summary: De-identification of data.	
Ref: 6	45 CFR 164	Section 524(1)
	Summary: Right of Access:"...an individual has a right of access to inspect and obtain a copy of protected health information about the individual in a designated record set, for as long as the protected health information is maintained in the designated record set..."	
Ref: 6	45 CFR 164	Section 526(a) (1)
	Summary: Right to amend: "An individual has the right to have a covered entity amend protected health information or a record about the individual in a designated record set for as long as the protected health information is maintained in the designated record set."	
Ref: 7	US Environmental Protection Agency Directive 2185 Good Automated Laboratory Practices – Principles and Guidance to Regulations for Ensuring Data Integrity in Automated Laboratory Operations – with Implementation Guidance	Section 8.4
	Summary: All data and modifications documented and retained. If Laboratory Raw Data (LRD) is purged from the LIMS, a verified copy of the LRD should be maintained, for at least the required retention period.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 7	US Environmental Protection Agency Directive 2185 Good Automated Laboratory Practices – Principles and Guidance to Regulations for Ensuring Data Integrity in Automated Laboratory Operations – with Implementation Guidance	Section 8.5
	Summary: Retain historical file of all versions of software with manuals, validation documentation, formulae and algorithms, and change control records. Conditions designed to be safe and secure and to adequately preserve software for required retention period. Identify means of retrieving data after system retirement.	
Ref: 7	US Environmental Protection Agency Directive 2185 Good Automated Laboratory Practices – Principles and Guidance to Regulations for Ensuring Data Integrity in Automated Laboratory Operations – with Implementation Guidance	Section 8.8
	Summary: Perform comprehensive periodic testing of LIMS performance and retain resulting documentation.	
Ref: 7	US Environmental Protection Agency Directive 2185 Good Automated Laboratory Practices – Principles and Guidance to Regulations for Ensuring Data Integrity in Automated Laboratory Operations – with Implementation Guidance	Section 8.9
	Summary: Retention of raw data, system records, and documentation to be described in SOP. Retention period specified in EPA contract or regulation, if not specified, should be sufficient to support any challenges to data integrity.	
Ref: 7	US Environmental Protection Agency Directive 2185 Good Automated Laboratory Practices – Principles and Guidance to Regulations for Ensuring Data Integrity in Automated Laboratory Operations – with Implementation Guidance	Section 8.10
	Summary: Provision of environmentally adequate storage for LIMS Raw Data and system documentation, including historical SOPs stored electronically.	
Ref: 7	US Environmental Protection Agency Directive 2185 Good Automated Laboratory Practices – Principles and Guidance to Regulations for Ensuring Data Integrity in Automated Laboratory Operations – with Implementation Guidance	Section 8.11
	Summary: Historical files of SOPs may be stored on magnetic media but the SOPs must remain available over time.	
Ref: 8	Rules and Guidance for Pharmaceutical Manufacturers and Distributors 1997 (U.K. Orange Guide)	
	Summary: The Orange Guide represents the UK's implementation of the EU Regulations. The entire guide has been reviewed for references to the following: Archiving – no references found. Storage of records – no references found. Retention of records (see below). There are numerous references to what records to retain and in what format; however, these have not been recorded.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 9	Commission Directive 2003/94/EC of 8 October 2003 laying down the principles and guidelines of good manufacturing practice in respect of medicinal products for human use and investigational medicinal products for human use	Article 9.1 Documentation
	Summary: For a medicinal product, the batch documentation shall be retained for at least one year after the expiry date of the batches to which it relates or at least five years after the certification referred to in Article 51(3) of Directive 2001/83/EC, whichever is the longer period.	
Ref: 9	Commission Directive 2003/94/EC of 8 October 2003 laying down the principles and guidelines of good manufacturing practice in respect of medicinal products for human use and investigational medicinal products for human use	Article 9.2 Documentation
	Summary: When electronic, photographic, or other data processing systems are used instead of written documents, the manufacturer shall first validate the systems by showing that the data will be appropriately stored during the anticipated period of storage. Data stored by those systems shall be made readily available in legible form and shall be provided to the competent authorities at their request. The electronically stored data shall be protected, by methods such as duplication or backup and transfer on to another storage system, against loss or damage of data, and audit trails shall be maintained.	
Ref: 10	EU Guide to Good Manufacturing Practice	Section 4.8 Documentation – General
	Summary: The records should be retained for at least one year after the expiry date of the finished product.	
Ref: 10	EU Guide to Good Manufacturing Practice	Section 4.9 Documentation – General
	Summary: Data may be recorded by electronic data processing systems, photographic, or other reliable means, but detailed procedures relating to the system in use should be available and the accuracy of the records should be checked. If documentation is handled by electronic data processing methods, only authorized persons should be able to enter or modify data in the computer and there should be a record of changes and deletions; access should be restricted by passwords or other means and the result of entry of critical data should be independently checked. Batch records electronically stored should be protected by backup transfer on magnetic tape, microfilm, paper, or other means. It is particular important that the data are readily available throughout the period of retention.	
Ref: 10	EU Guide to Good Manufacturing Practice	Section 6.8
	Summary: Any QC documentation relating to a batch record should be retained for one year after the expiry date of the batch and at least five years after the certification referred to in Directive 2001/83/EC.	
Ref: 10	EU Guide to Good Manufacturing Practice	Section 6.9
	Summary: For some kinds of data, it is recommended that records be kept in a manner permitting trend evaluation.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 10	EU Guide to Good Manufacturing Practice	Section 6.10
	Summary: In addition to the information which is part of the batch record, other original data such as laboratory notebooks and/or records should be retained and readily available.	
Ref: 10	EU Guide to Good Manufacturing Practice Annex 11 Computerized Systems	Section 10
	Summary: Authority to amend entered data should be restricted to nominated persons. Any alteration to an entry of critical data should be authorized and recorded with the reason for the change. Consideration should be given to building into the system the creation of a complete record of all entries and amendments (an audit trail).	
Ref: 10	EU Guide to Good Manufacturing Practice Annex 11 Computerized Systems	Section 11
	Summary: Alterations to a system or to a computer program should only be made in accordance with a defined procedure which should include provisions for validating, checking, approving, and implementing the change. Every significant modification should be validated.	
Ref: 10	EU Guide to Good Manufacturing Practice	Section 12
	Summary: For quality auditing purposes, it should be possible to obtain clear printed copies of electronically stored data.	
Ref: 10	EU Guide to Good Manufacturing Practice	Section 13
	Summary: Data should be secured by physical or electronic means against willful or accidental damage, in accordance with item 4.9 of the Guide. Stored data should be checked for accessibility, durability, and accuracy. If changes are proposed to the computer equipment or its programs, the above mentioned checks should be performed at a frequency appropriate to the storage medium being used.	
Ref: 10	EU Guide to Good Manufacturing Practice	Section 14
	Summary: Data should be protected by backing-up at regular intervals. Backup data should be stored as long as necessary at a separate and secure location.	
Ref: 10	EU Guide to Good Manufacturing Practice Annex 12	Section 45
	Summary: The documentation associated with the validation and commissioning of the plant should be retained for one year after the expiry date or at least five years after the release of the last product processed by the plant, whichever is the longer.	
Ref: 10	EU Guide to Good Manufacturing Practice Annex 13	Section 10
	Summary: Batch manufacture records should be retained for at least two years after completion of the clinical trial or at least two years after the formal discontinuation or in conformance with the applicable regulatory requirements.	
Ref: 10	EU Guide to Good Manufacturing Practice Annex 18	Section 6.12
	Summary: All production, control, and distribution records should be retained for at least one year after the expiry date of the batch. For APIs with retest dates, records should be retained for at least three years after the batch is completely distributed.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 10	EU Guide to Good Manufacturing Practice Annex 18	Section 6.15
	Summary: During the retention period, originals or copies of records should be readily available at the establishment where the activities described in such records occurred. Records that can be promptly retrieved from another location by electronic or other means are acceptable.	
Ref: 10	EU Guide to Good Manufacturing Practice Annex 18	Section 6.16
	Summary: Specifications, instructions, procedures, and records can be retained either as originals or as true copies such as photocopies, microfilm, microfiche, or other accurate reproductions of the original records. Where reduction techniques such as microfilming or electronic records are used, suitable retrieval equipment and a means to produce a hard copy should be readily available.	
Ref: 10	EU Guide to Good Manufacturing Practice Annex 18	Section 6.18
	Summary: If electronic signatures are used on documents, they should be authenticated and secure.	
Ref: 11	Directive 2001/83/EU on the Community code relating to medicinal products for human use – Title VII Wholesale distribution of medicinal products	Article 80 (f)
	Summary: Holders of the distribution authorization must keep the records available to the competent authorities for inspection purposes for a period of five years.	
Ref: 12	Guidelines on Good Distribution Practice (GDP) of medicinal products for human use – 94/C 63/03	Section 7
	Summary: Records should be clear and readily available. They should be retained for a period of five years at least.	
Ref: 13	EUCOMED Position Paper on Good Distribution Practice (GDP)	Section 6
	Summary: Records should be clear and readily available. They should be retained for a period of 10 years or expected lifetime of the product unless indicated otherwise by the supplier.	
Ref: 14	95/46/EC EU Data Protection Directive	Overview
	Summary: Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Overall EU requirement for data protection – mandatory on all EU member states to adopt into national law. Provides for similar obligations to HIPAA to use data only for specified purposes and to allow data subject access to data to correct or have removed as appropriate. Note: Requirement is directly upon the Data Controller to satisfy themselves that personal data has been obtained lawfully (i.e., with consent of the subject) and is properly handled thereafter.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 14	95/46/EC EU Data Protection Directive	Article 6(d) (e)
	Summary: Sets out the key principles to be applied, includes the need to keep data accurate and up to date and the prohibition of holding data indefinitely. Data must only be kept in identifiable form for as long as is strictly necessary for declared purposes. Provisions are allowed to be made on a national basis for statistical and scientific purposes, but these do not normally apply unless either the data has been made anonymous (and hence taken out of scope) or some direct relevance of maintaining identification can be shown (e.g., original data related to a condition which has a potential genetic component).	
Ref: 14	95/46/EC EU Data Protection Directive	Article 8
	Summary: In most cases, unless explicit (normally implemented as written) consent is given, prohibits the processing of personal data revealing: Racial or ethnic origin. Political opinions. Religious or philosophical beliefs. Trade-union membership. Concerning health or sex life. It is advisable to keep copies of these consents with any archive which relies upon them. Purposes must be meaningful and specific and data given for one purpose may not be used for other purposes without further explicit consent. Therefore, great care must be taken over the uses to which archived data is put.	
Ref: 14	95/46/EC EU Data Protection Directive	Article 12
	Summary: Establishes the subject's right to access data and to require correction or blocking as appropriate.	
Ref: 14	95/46/EC EU Data Protection Directive	Article 14
	Summary: Establishes a Data Subject's right to object to the processing of their data (and by extension, require its deletion). Taken together articles 12 and 14 require the ability for identifiable data in an archive to be: Searched on the basis of personal identification data. Edited or redacted in such a way as to make the data anonymous. The directive does not provide a guide as to the basis for making data anonymous; rather it requires that the identity of the individual cannot be identified from the data itself or the data in combination with any other data in the possession of the Data Controller. (i.e., blind data is identifiable if the key files to unblind it are also available). The level of anonymity established for HIPAA is likely to be acceptable. However, lower standards may be defensible if the likelihood of identification is demonstrably low.	
Ref: 14	95/46/EC EU Data Protection Directive	Article 18
	Summary: Establishes the obligation on Data Controllers to register the fact that they hold personal data and the purposes for which it is retained/used.	
Ref: 14	95/46/EC EU Data Protection Directive	Article 25
	Summary: Establishes the requirements on Data Controllers to ensure that, if data is to be transferred outside of the EU, adequate legal protections exist in the target country. A very limited list of countries has been established by the EU as meeting these requirements. In particular, the status of the US is under question. Generally, legally enforceable contracts to adhere to data protection requirements are used to provide legal remedies for breaches. This is particularly significant for any organization establishing central corporate archives or who are archiving multi-national trials.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 15	PIC/S Guide PI 011-1 20 August 2003.Good Practices for Computerized Systems in Regulated "GxP" Environments	Section 14.4 Validation Strategies and Priorities
	Summary: GxP compliance evidence is essential for the following aspects and activities related to computerized systems: Data input (capture and integrity), data filing, data-processing, networks, process control and monitoring, electronic records, archiving, retrieval, printing, access, change management, audit trails, and decisions associated with any automated GxP related activity.	
Ref: 15	PIC/S Guide PI 011-1 20 August 2003.Good Practices for Computerized Systems in Regulated "GxP" Environments	Section 19.5 Security
	Summary: The validated backup procedure, including storage facilities and media should assure data integrity. The frequency of backup is dependent on the computer system functions and the risk assessment of a loss of data. In order to guarantee the availability of stored data, backup copies should be made of such data that are required to re-construct all GxP-relevant documentation (including audit trail records).	
Ref: 15	PIC/S Guide PI 011-1 20 August 2003.Good Practices for Computerized Systems in Regulated "GxP" Environments	Section 21.1 ER/ES
	Summary: EC Directive 91/356 sets out the legal requirements for EU GMP. The GMP obligations include a requirement to maintain a system of documentation. The main requirements here being that the regulated user has validated the system by proving that the system is able to store the data for the required time, that the data is made readily available in legible form and that the data is protected against loss or damage.	
Ref: 15	PIC/S Guide PI 011-1 20 August 2003.Good Practices for Computerized Systems in Regulated "GxP" Environments	Section 21.3 ER/ES
	Summary: The central consideration here as in Directive 91/356, is that records are accurately made and protected against loss or damage or unauthorized alteration so that there is a clear and accurate audit trail throughout the manufacturing process available to the licensing authority for the appropriate time.	
Ref: 15	PIC/S Guide PI 011-1 20 August 2003.Good Practices for Computerized Systems in Regulated "GxP" Environments	Section 21.10 ER/ES
	Summary: Issues to consider where electronic records are used to retain GxP data: Procedures exist to enable the retrieval of records throughout the retention period. Archiving procedures are provided and records of use.	
Ref: 15	PIC/S Guide PI 011-1 20 August 2003.Good Practices for Computerized Systems in Regulated "GxP" Environments	Section 21.12 ER/ES
	Summary: Issues to consider when the GxP system has a provision for external access. The system has a method of ensuring that external access and inputs come only from authorized clients and that they come in the correct format, for example as encrypted, digitally-signed mail or data packets. A mechanism must exist to quarantine external inputs where security conditions are not met. The information security management arrangements need to cover the quarantine, notification, and the final sentencing of such inputs. Mechanisms are in place to ensure that all external access can be tracked. Each element of the processing stage should incorporate logging and monitoring facilities. However, inspectors may expect to see less onerous tracking for 'read only' access to a suitably secure and protected system.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 16	ICH GCP	Section 3.4
	Summary: IRB/IEC Records to be maintained for at least three years from completion of trial (SOPs, memberships, affiliations, minutes, etc.).	
Ref: 16	ICH GCP	Section 4.9.5
	Summary: Essential documents to be maintained by investigator for at least two years from last approval of a marketing application in an ICH region. Longer periods may be set in agreement with, or required by, either the sponsor or national regulators. Sponsor must explicitly permit deletion (reference to 5.5.12).	
Ref: 16	ICH GCP	Section 5.6
	Summary: The sponsor or owner of the data shall retain all sponsor related trial related essential records (as defined).	
Ref: 16	ICH GCP	Section 5.5.8
	Summary: Documents to be retained for at least two years from formal discontinuation.	
Ref: 16	ICH GCP	Section 5.5.11
	Summary: Sponsor requirements as for investigator (at least two years from last approval of a marketing application in an ICH region) plus extension to cover the requirement that there are no pending or contemplated applications.	
Ref: 16	ICH GCP	Section 5.5.12
	Summary: Sponsor must advise investigator that a trial record is no longer required.	
Ref: 16	ICH GCP	Section 8
	Summary: Detailed list (12 pages) of documents to be maintained and where (by whom) they are to be maintained/archived.	
Ref: 17	ENV/MC/CHEM (98)17 OECD Series on Principles of GLP and Compliance Monitoring Number 1.OECD Principles on Good Laboratory Practice (as revised in 1997)	Test Facility Organization and Personnel
	Summary: The OECD GLP and Compliance Monitoring No 1 has been reviewed for references to archiving and records, and the following references were found. Section 1.1 Test Facility Management's Responsibilities. At a minimum it should: (l) Ensure that an individual is identified as responsible for the management of the archive(s). (q) Establish procedures to ensure that computerized systems are suitable for their intended purpose, and are validated, operated and maintained in accordance with these Principles of Good Laboratory Practice.	
Ref: 17	ENV/MC/CHEM (98)17 OECD Series on Principles of GLP and Compliance Monitoring Number 1.OECD Principles on Good Laboratory Practice (as revised in 1997)	Section 2.1 Definitions of Terms
	Summary: Good Laboratory Practice (GLP) is a quality system concerned with the organizational process and the conditions under which non-clinical health and environmental safety studies are planned, performed, monitored, recorded, archived and reported.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 17	ENV/MC/CHEM (98)17 OECD Series on Principles of GLP and Compliance Monitoring Number 1.OECD Principles on Good Laboratory Practice (as revised in 1997)	Section 2.3 Definitions of Terms
	Summary: Terms Concerning the Non-Clinical Health and Environmental Safety Study. Raw data means all original test facility records and documentation, or verified copies thereof, which are the result of the original observations and activities in a study. Raw data also may include, for example, photographs, microfilm or microfiche copies, computer readable media, dictated observations, recorded data from automated instruments, or any other data storage medium that has been recognized as capable of providing secure storage of information for a time period as stated in Section 10, below.	
Ref: 17	ENV/MC/CHEM (98)17 OECD Series on Principles of GLP and Compliance Monitoring Number 1.OECD Principles on Good Laboratory Practice (as revised in 1997)	Section 3.4 Archive Facilities
	Summary: Archive facilities should be provided for the secure storage and retrieval of study plans, raw data, final reports, samples of test items, and specimens. Archive design and archive conditions should protect contents from untimely deterioration.	
Ref: 17	ENV/MC/CHEM (98)17 OECD Series on Principles of GLP and Compliance Monitoring Number 1.OECD Principles on Good Laboratory Practice (as revised in 1997)	Section 4.1 Apparatus, Material, and Reagents
	Summary: Apparatus, including validated computerized systems, used for the generation, storage and retrieval of data, and for controlling environmental factors relevant to the study should be suitably located and of appropriate design and adequate capacity.	
Ref: 17	ENV/MC/CHEM (98)17 OECD Series on Principles of GLP and Compliance Monitoring Number 1.OECD Principles on Good Laboratory Practice (as revised in 1997)	Section 7.4 Standard Operating Procedure
	Summary: Standard Operating Procedures should be available for, but not be limited to, the following categories of test facility activities. The details given under each heading are to be considered as illustrative examples. 1. Test and Reference Items – Receipt, identification, labeling, handling, sampling, and storage. 2. Apparatus, Materials, and Reagents a) Apparatus – Use, maintenance, cleaning, and calibration. b) Computerized Systems – Validation, operation, maintenance, security, change control, and backup. c) Materials, Reagents and Solutions – Preparation and labeling. 3. Record Keeping, Reporting, Storage, and Retrieval – Coding of studies, data collection, preparation of reports, indexing systems, handling of data, including the use of computerized systems.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 17	ENV/MC/CHEM (98)17 OECD Series on Principles of GLP and Compliance Monitoring Number 1.OECD Principles on Good Laboratory Practice (as revised in 1997)	Section 8.3 Performance of the Study
	Summary: Conduct of the Study. Data generated as a direct computer input should be identified at the time of data input by the individual(s) responsible for direct data entries. Computerized system design should always provide for the retention of full audit trails to show all changes to the data without obscuring the original data. It should be possible to associate all changes to data with the persons having made those changes, for example, by use of timed and dated (electronic) signatures. Reason for changes should be given.	
Ref: 17	ENV/MC/CHEM (98)17 OECD Series on Principles of GLP and Compliance Monitoring Number 1.OECD Principles on Good Laboratory Practice (as revised in 1997)	Section 10.1 Storage and Retention of Records and Materials
	Summary: The following should be retained in the archives for the period specified by the appropriate authorities: a) The study plan, raw data, samples of test and reference items, specimens, and the final report of each study. b) Records of all inspections performed by the Quality Assurance Program, as well as master schedules. c) Records of qualifications, training, experience, and job descriptions of personnel. d) Records and reports of the maintenance and calibration of apparatus. e) Validation documentation for computerized systems. f) The historical file of all Standard Operating Procedures. g) Environmental monitoring records. In the absence of a required retention period, the final disposition of any study materials should be documented. When samples of test and reference items and specimens are disposed of before the expiry of the required retention period for any reason, this should be justified and documented. Samples of test and reference items and specimens should be retained only as long as the quality of the preparation permits evaluation.	
Ref: 17	ENV/MC/CHEM (98)17 OECD Series on Principles of GLP and Compliance Monitoring Number 1.OECD Principles on Good Laboratory Practice (as revised in 1997).	Section 10 Storage and Retention of Records and Materials
	Summary: 10.2 Material retained in the archives should be indexed so as to facilitate orderly storage and retrieval. 10.3 Only personnel authorized by management should have access to the archives. Movement of material in and out of the archives should be properly recorded. 10.4 If a test facility or an archive contracting facility goes out of business and has no legal successor, the archive should be transferred to the archives of the sponsor(s) of the study(s).	

Table 5.3: Regulatory References (continued)

Ref: 18	OECD Series on Principles of GLP and Compliance Monitoring Number 4 (Revised) Consensus Document, Quality Assurance and GLP, ENV/JM/MONO(99)20	Section 2.2 Responsibilities of the Quality Assurance Personnel
	Summary: Section II.10.1(b) records of all inspections performed by the Quality Assurance Program, as well as master schedules, should be retained in the archives for the period specified by the appropriate authorities.	
Ref: 19	OECD Series on Principles of GLP and Compliance Monitoring Number 10 GLP Consensus Document. The Application of the Principles of GLP to Computerized Systems. Environment Monograph No. 116 (OECD/GD(95)115)	Section 5 Data
	Summary: When raw data are held electronically, it is necessary to provide for long term retention requirements for the type of data held and the expected life of computerized systems. Hardware and software system changes must provide for continued access to and retention of the raw data without integrity risks. Supporting information such as maintenance logs and calibration records that are necessary to verify the validity of raw data or to permit reconstruction of a process or a study should be retained in the archives.	
Ref: 19	OECD Series on Principles of GLP and Compliance Monitoring Number 10 GLP Consensus Document. The Application of the Principles of GLP to Computerized Systems. Environment Monograph No. 116 (OECD/GD(95)115)	Section 6 Security
	Summary: Documented security procedures should be in place for the protection of hardware, software, and data from corruption or unauthorized modification, or loss. In this context, security includes the prevention of unauthorized access or changes to the computerized system, as well as to the data held within the system. The potential for corruption of data by viruses or other agents also should be addressed. Security measures also should be taken to ensure data integrity in the event of both short-term and long- term system failure. Since maintaining data integrity is a primary objective of the GLP Principles, it is important that everyone associated with a computerized system is aware of the necessity for the above security considerations. Management should ensure that personnel are aware of the importance of data security, the procedures, and system features that are available to provide appropriate security and the consequences of security breaches. Such system features could include routine surveillance of system access, the implementation of file verification routines, and exception and/or trend reporting.	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 19	OECD Series on Principles of GLP and Compliance Monitoring Number 10 GLP Consensus Document. The Application of the Principles of GLP to Computerized Systems. Environment Monograph No. 116 (OECD/GD (95)115).	Section 9 Archives
	Summary: The GLP Principles for archiving data must be applied consistently to all data types. Therefore, it is important that electronic data are stored with the same levels of access control, indexing, and expedient retrieval as other types of data. Where electronic data from more than one study are stored on a single storage medium (e.g., disk or tape) a detailed index will be required. It may be necessary to provide facilities with specific environmental controls appropriate to ensure the integrity of the stored electronic data. If this necessitates additional archive facilities, then management should ensure that the personnel responsible for managing the archives are identified and that access is limited to authorized personnel. It also will be necessary to implement procedures to ensure that the long-term integrity of data stored electronically is not compromised. Where problems with long-term access to data are envisaged or when computerized systems have to be retired, procedures for ensuring that continued readability of the data should be established. This may, for example, include producing hard copy printouts or transferring the data to another system. No electronically stored data should be destroyed without management authorization and relevant documentation. Other data held in support of computerized systems, such as source code and development, validation, operation, maintenance, and monitoring records, should be held for at least as long as study records associated with these systems.	
Ref: 20	Good Laboratory Practice Advisory Leaflet No. 1 The Application of GLP Principles to Computer Systems (U.K. GLP Compliance Program, DoH, London, 1995).	Requirements (b)
	Summary: Check the procedures for the lodgement of raw data into the archive.	
Ref: 20	Good Laboratory Practice Advisory Leaflet No. 1 The Application of GLP Principles to Computer Systems (U.K. GLP Compliance Program, DoH, London, 1995).	Archives
	Summary: Essentially identical to Section 9 of OECD monograph 116 (OECD/GD (95) 115).	

Table 5.3: Regulatory References (continued)

Ref #	Regulation	Section
Ref: 21	EU Directive 2001/20/EC Implementation of Good Clinical Practice in the Conduct of Clinical Trials on Medicinal Products for Human Use	Article 13 Manufacture and import of investigational medicinal products
	Summary: In all cases, the qualified person must certify in a register or equivalent document that each production batch satisfies the provisions of this Article. The said register or equivalent document shall be kept up to date as operations are carried out and shall remain at the disposal of agents of the competent authority for the period specified in the provisions of the Member States concerned. This period shall in any event be not less than five years.	
Ref: 21	EU Directive 2001/20/EC Implementation of Good Clinical Practice in the Conduct of Clinical Trials on Medicinal Products for Human Use	Article 16 Notification of adverse events
	Summary: The sponsor shall keep detailed records of all adverse events which are reported to him by the investigator or investigators. These records shall be submitted to the Member States in whose territory the clinical trial is being conducted, if they so request.	
Ref: 22	Sarbanes-Oxley Act 2002	Section 404 Management Assessment of Internal Controls
	Summary: The Sarbanes-Oxley Act has been reviewed for references to archiving (none found) and storage of records (none found) and retention of records. (a) (1) State the responsibility of management for establishing and maintaining an adequate internal control structure and procedures for financial reporting. (2) Contain an assessment of the effectiveness of the internal control structure and procedures.	
Ref: 22	Sarbanes-Oxley Act 2002	Section 802 Criminal Penalties for altering documents
	Summary: Section 1519 Destruction, alteration, or falsification of records in federal investigations and bankruptcy. Whoever knowingly alters, destroys, mutilates, conceals, covers up, falsifies, or makes a false entry in a record, document, or tangible object with the intent to impede, obstruct, or influence the investigation or proper administration of any matter within the jurisdiction of any department or agency of the United States or any case filed under title 11, or in relation to or contemplation of any such matter or case, shall be fined under this title, imprisoned not more than 20 years or both. Section 1520 Destruction of Corporate Audit Records. Although the act is directed specifically at financial reporting, the above could be interpreted to include protection of information held in archives that are required under predicate rules.	

Appendix B
The OAIS Reference Model

6 Appendix B: The OAIS Reference Model

This Appendix outlines and explains a reference model for an Open Archival Information System (OAIS) as defined by the international CCSDS organization. The model has been simplified in this Guide to make it easier to understand. Some of the principles from this model are applied in the development of an archive strategy. Note that the OAIS model does not specify which processes or functions are automated or not, and makes no distinction between automated and manual processes/functions.

The model is explained through a set of data flow diagrams and bullet points that cover the environment of the OAIS, the functional model, and the packaging of information. Further details also are provided for the main processes within the model.

6.1 The OAIS Reference Model

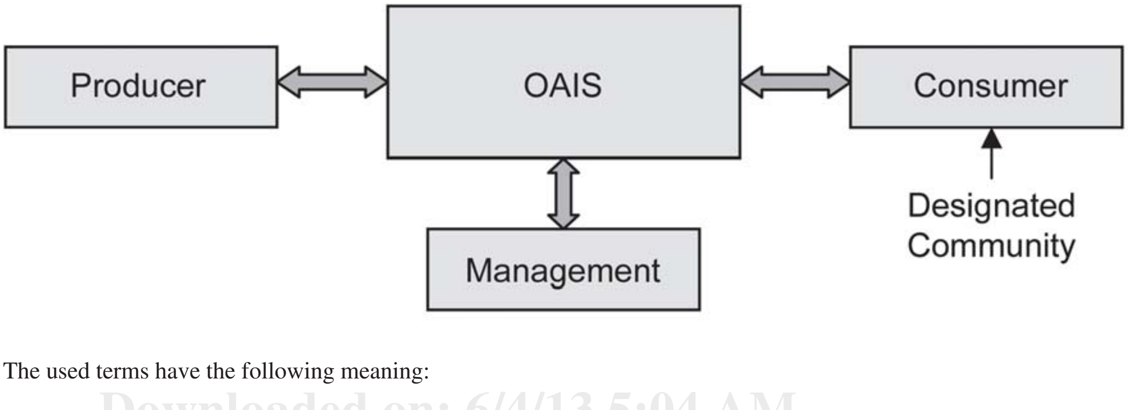
The Reference Model for an OAIS has been developed by the CCSDS. The aim of the OAIS model is to define a common framework of terms and concepts to promote future technical developments. The model is platform-independent. Although the model is a technical recommendation and not a standard in itself, it has been published as ISO standard 14721:2003. CCSDS is an international body for space agencies, but the work they have carried out on long-term preservation of digital information is of universal relevance and widely quoted in the context of EDA.

For these reasons, the OAIS model is outlined in this Guide, albeit in a simplified form. The full recommendation is some 148 pages (document reference CCSDS 650.0-B-1, January 2002) and can be downloaded from www.ccsds.org.

6.1.1 The OAIS Environment

The OAIS model defines three main parties that interact with the OAIS as shown in Figure 6.1.

Figure 6.1: OAIS Environment Model



The used terms have the following meaning:

- Producer is the organization and/or system that provide the information to be archived.

- Management is the organization and/or individual(s) that set policies for the operation and functioning of the OAIS. In this model, the day-to-day activities are captured by the Administration function within the OAIS (shown in Figure 6.2).

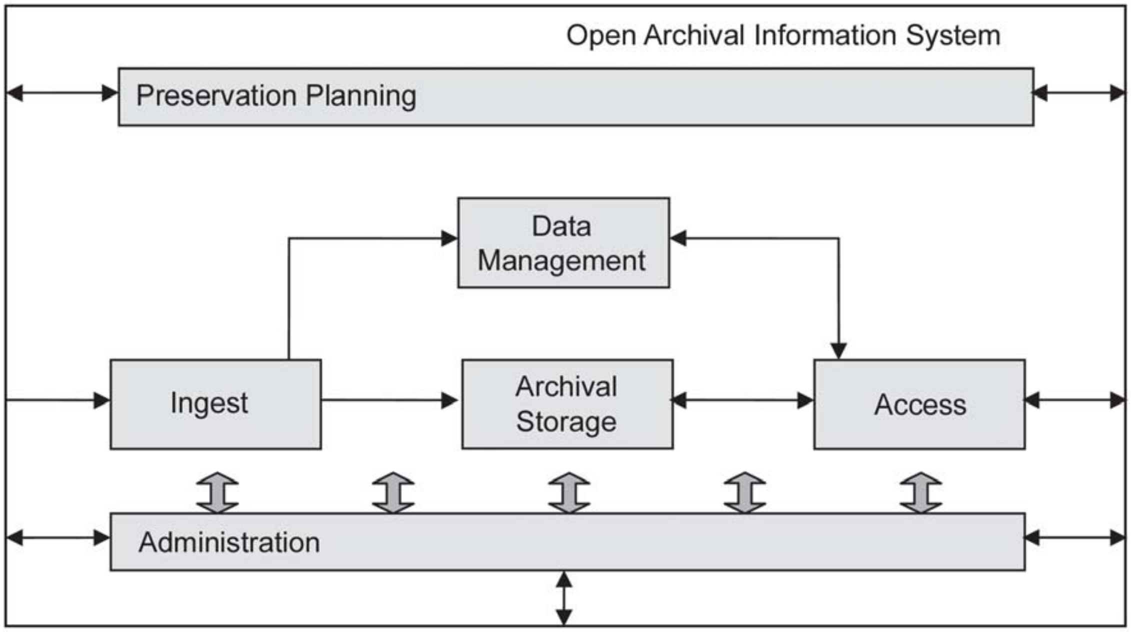
- Consumer is the organization and/or system that makes use of the archived information, i.e., search, retrieve, etc. It is conceivable that in some instances the Producer and Consumer are the same.

- Designated Community refers to the group of Consumers that are assumed to have the knowledge base to be able to understand the archived information. Depending on the type of archived information, these assumptions of background knowledge and ability to interpret information may vary. For example, for a public library the designated community is very large as the requisites on the consumer are small. But for clinical trial data, the direct opposite would be true.

6.1.2 OAIS Functional Entities

The OAIS environment model may be expanded to depict the main functional units that make up the OAIS. These have been shown simplified in Figure 6.2.

Figure 6.2: OAIS Functional Model



The six functional entities identified in Figure 6.2 have the following meaning. Each functional entity may consist of one or several processes.

- Ingest is the process whereby information is received from the Producer and prepared for storage in the archive. This includes extracting relevant reference information so the stored data can later be found, and to format the data so it is suitable for storage.

- Archival Storage are the processes whereby data is received from Ingest, for holding the data in storage, managing the storage, such as refreshing storage media, perform routine error-checking and disaster recovery, as well as passing stored data to the Access function.

- Data Management provides the function whereby the archived data can be referenced, indexed, and searched and reports can be prepared.

- Access is the process whereby the Consumer can interface with the archive, search and/or retrieve stored data, and inquire about archive properties, such as access rights, existence, and availability of information, etc.

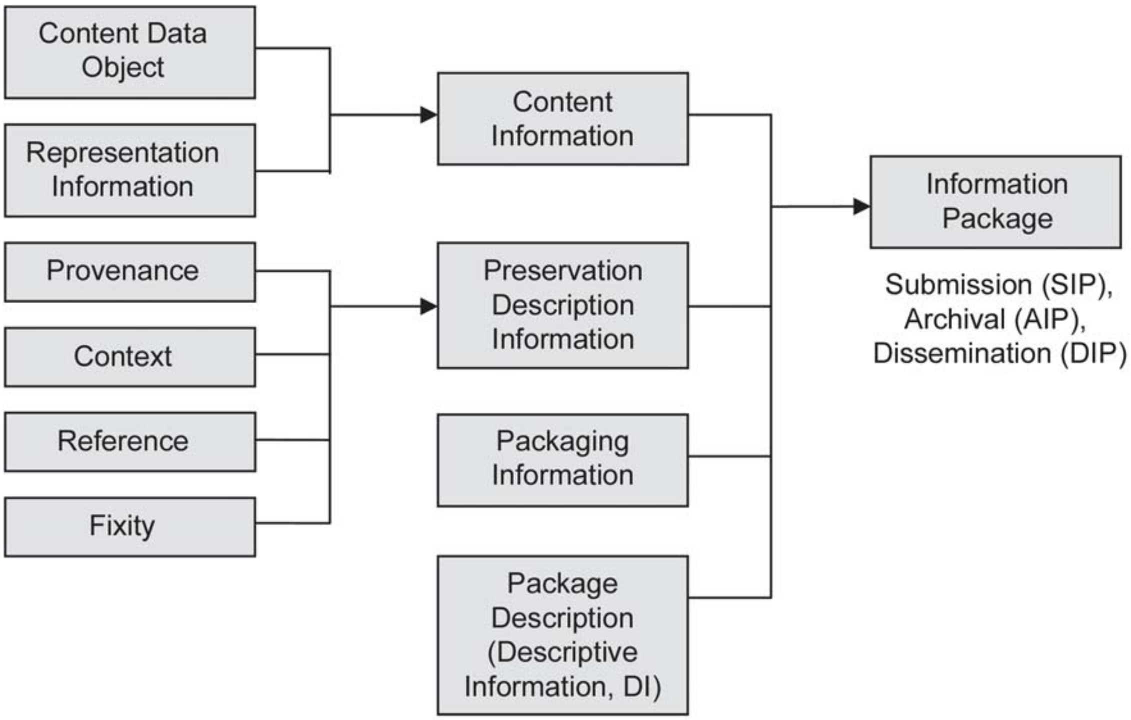
- Administration provides the services and functions for the overall operation of the archive. This includes managing and checking the various processes as well as the archive system itself, i.e., the hardware and software, associated engineering and quality tools, and functions. The Administration function interfaces with all the other five OAIS functional entities; this has been depicted by block arrows in Figure 6.2 to make the figure less cluttered. In addition, Administration interfaces with the Producer, Consumer, and Management.

- Preservation Planning monitors the overall environment for the OAIS as shown in Figure 6.1, and is the planning function for projecting and handling future changes to the platform or environment. As such, the Preservation Planning function must interface with the Producers and Consumers (and Management via the Administration function).

6.1.3 Information Packages

The OAIS recommendation describes how information is put into various packages and structures so that the data can be archived, searched, retrieved, and understood. A simplified hierarchical structure for these packages has been shown in Figure 6.3.

Figure 6.3: Package Structure


Content Data Object
Representation Information
Provenance
Context
Reference
Fixity
Content Information
Preservation Description Information
Packaging Information
Package Description (Descriptive Information, DI)
Information Package
Submission (SIP), Archival (AIP), Dissemination (DIP)

The used terms have the following meaning:

- Content Data Object is the original information that is required to be archived. For archived data, it refers back to the data provided by the Producer. Using familiar GxP terminology; this is the raw data.

- Representation Information is the necessary data and information that makes the Content Data Object understandable to the Consumer. Depending on the knowledge base of the Designated Community, the Representation Information may vary. For example, this information may be a piece of application software that is needed to be able to read the content, or it can consist of key metadata.

- Content Information is the Content Data Object together with its necessary Representation Information. Irrespective of the OAIS realization, this is the minimum information that is needed to be stored. Only retaining the Content Data Object will most likely make the data impossible to decipher; a bit like reading binary code without the rules that make up what the code structure is and what information the binary bits represent.

- Preservation Description Information is related to the long-term storage of the Content Information. It is intended to provide a clear identification of the Content Information and to capture an understanding of the environment of the Producer, i.e., a similar role to that of the orientation memorandum (see Section 3.3.3.2). Preservation Descriptive Information is made up of constituent parts as detailed in the next four bullet points.

- Provenance is akin to an audit trail, i.e., it records the source of the Content Information, ownership of the data, and what processing steps have been carried out.

- Context describes the external references to the data, such as background and reason for the data and its archival, relationship with other data (archived or stored elsewhere), etc.

- Reference is the identifier that uniquely references the data so that it cannot be accidentally interchanged with other stored data. For example, GLP data may be identified through a unique study report number, instrument used, and sample reference and date. GMP data may be identified through lot number, recipe identity, date, and machine/processing step or phase.

- Fixity is the measure by which the data is kept unadulterated. This may involve various security measures such as digital authentication to demonstrate the data has not been changed.

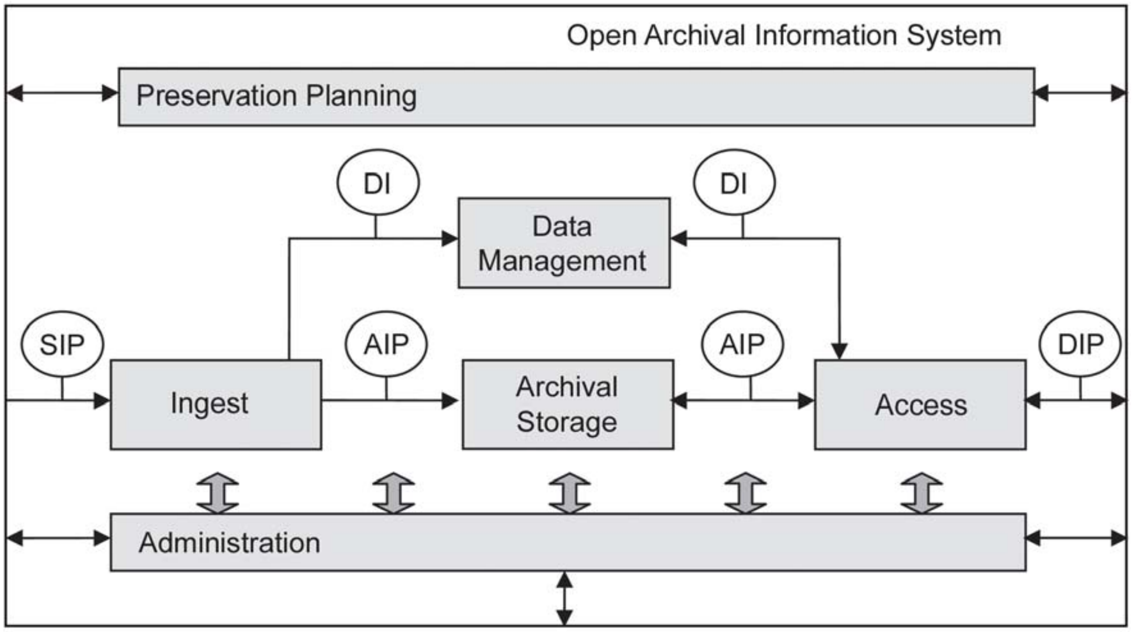
- Packaging Information holds the Content Information and Preservation Description Information together in a package. It could be something as simple as an index of the content of the Content Information and Preservation Description Information files.

- Package Description is the information needed to be able to make useful searches and retrieval of Information Packages. It could be a title, key words, executive summary, and a set of attributes or key metadata, such as type of source data, origin of data, date, etc. A generic name for the Package Description is Descriptive Information (DI).

- Information Package is the collective name for all the above described constituent parts when put together into packages. More generally, these can be referenced to as Information Objects.

The OAIS model identifies three specific Information Packages, and these have been added to Figure 6.4, which is otherwise identical to Figure 6.2.

Figure 6.4: OAIS Functional Model Showing Information Packages



The used terms have the following meaning:

- Submission Information Package (SIP) is the information that is required to transfer information from the Producer to the OAIS. Part of this package will originate from the Producer and part of the package will be imposed by and be specific to the OAIS. The Ingest process handles the SIP and makes the Descriptive Information (DI) or Package Description available to the Data Management function. Note that the DI is a constituent of the SIP or AIP and not a complete Information Package as such.

- Archival Information Package (AIP) is the information that is required to safely keep the data for long-term storage. The AIP will depend on the realization of the OAIS. There may be complex relationships between the AIP and SIP/DIP, i.e., not necessarily one-to-one or one-to-many interactions. Further, there may be interdependencies between different AIPs.

- Dissemination Information Package (DIP) is partly the mirror package to the SIP, providing the Consumer with the required data to turn this into useful information. The DIP may include only certain parts of the AIP. The Descriptive Information (DI) is required so the Consumer can relate the given information to what was asked for and is included (at least partly) in the DIP.

6.2 Overview of Processes

The main processes involved in an EDA have been outlined in Section 6.1 of this Guide as part of the OAIS model. This Section provides further descriptions of some of the processes, which are loosely based on the OAIS model. Diagrams set out in the OAIS model in this respect are fairly complex and are not deemed suitable for direct inclusion in the Guide. Instead, simplified process descriptions have been developed within this Section.

The following processes are described:

- Ingest

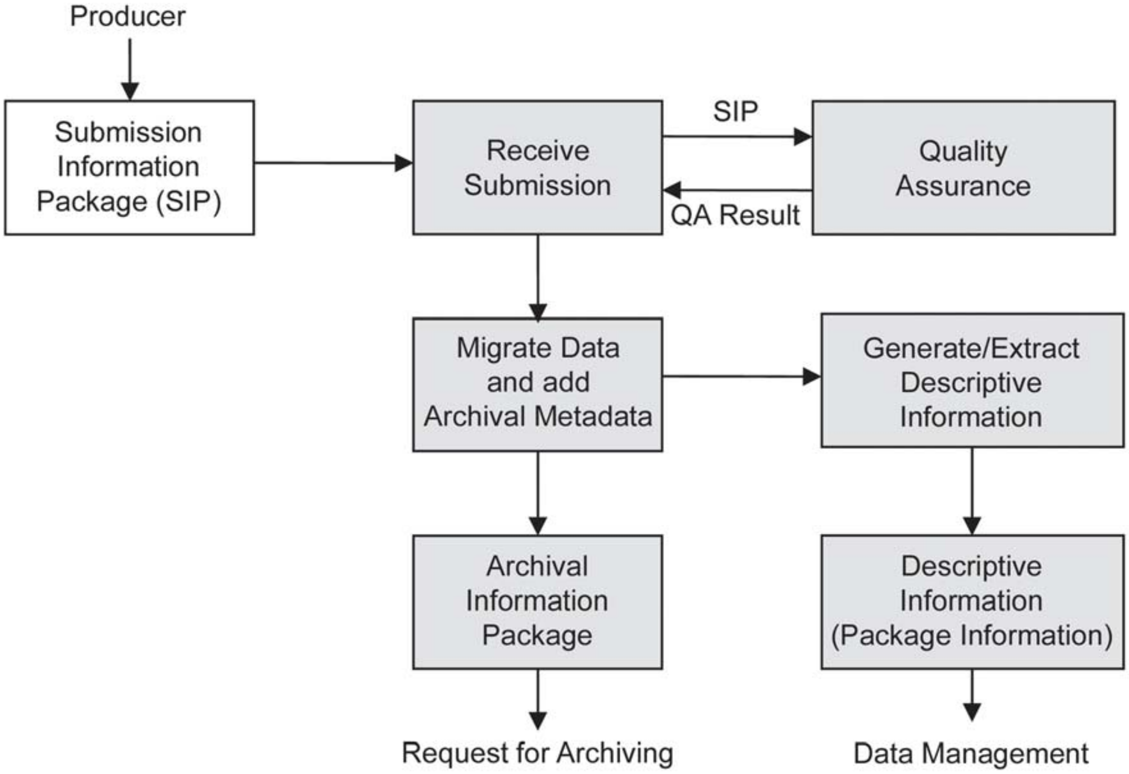
- Archival storage – receive data
- Archival storage – other functions
- Access
- Data management
- Administration
- Preservation planning

6.2.1 Ingest

A major step in the archiving process is to receive and transform one or more Submission Information Packages (SIPs) into one or more Archival Information Packages (AIPs) that conform to the archive's data formatting and documentation standards and are ready for archiving on physical storage media. The process may involve file format conversions, data representation conversions, or reorganization of the information in the SIPs. This process is called the Ingest function.

The SIP may be delivered from the Producer via electronic transfer, e.g., FTP, loaded from media submitted to the archive or simply mounted on the archive file system for access, e.g., CD-ROM.
The principle of the Ingest process is shown in Figure 6.5.

Figure 6.5: Ingest Process


Producer
Submission Information Package (SIP)
Receive Submission
SIP
QA Result
Quality Assurance
Migrate Data and add Archival Metadata
Generate/Extract Descriptive Information
Archival Information Package
Descriptive Information (Package Information)
Request for Archiving
Data Management

Since the Ingest process involves the physical transfer of the data from the Producer to the EDA staging area ready for final storage, the OAIS model includes a Quality Assurance function that this transfer has been carried out successfully.

6.2.2 Archival Storage – Receive Data

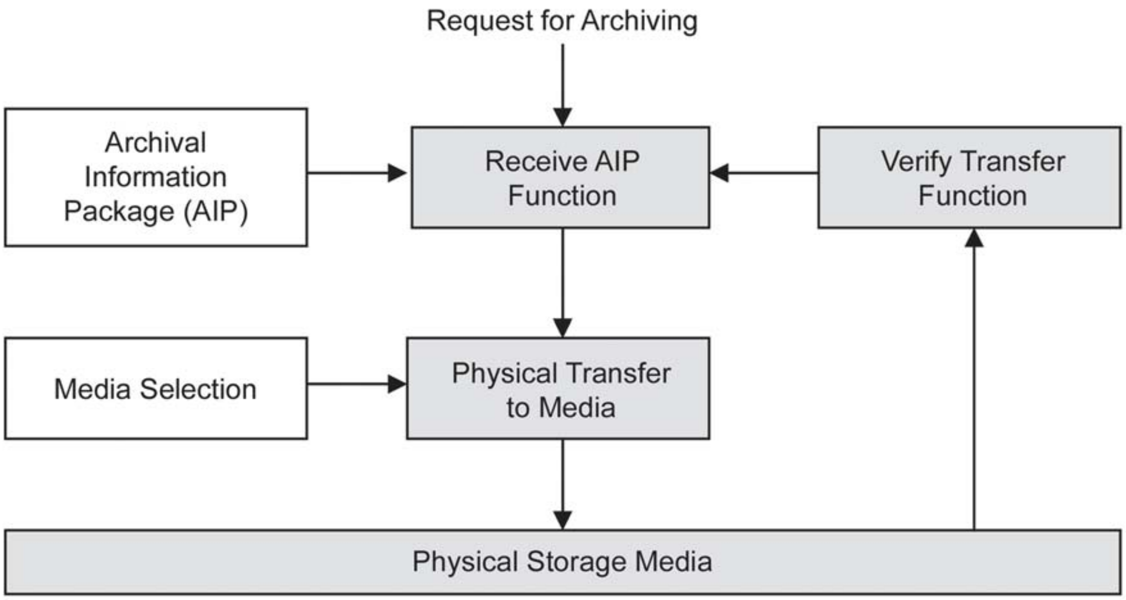
After the Ingest function follows the Adding to Archive process, which is handled by the Archival Storage function. The process mainly ensures that a failure-free archiving takes place. The Archival Storage function also manages the maintenance of the archive, i.e., that data is held securely over time and is available to the Access function and ultimately the Consumer. These latter aspects are dealt with in Section 6.2.3 of this Guide.

Before the AIP is permanently stored, the actual form of media should be selected. Examples of media forms include:

- Tape (passive system)
- CD or DVD (passive system)
- Server with hard disk (on-line system)

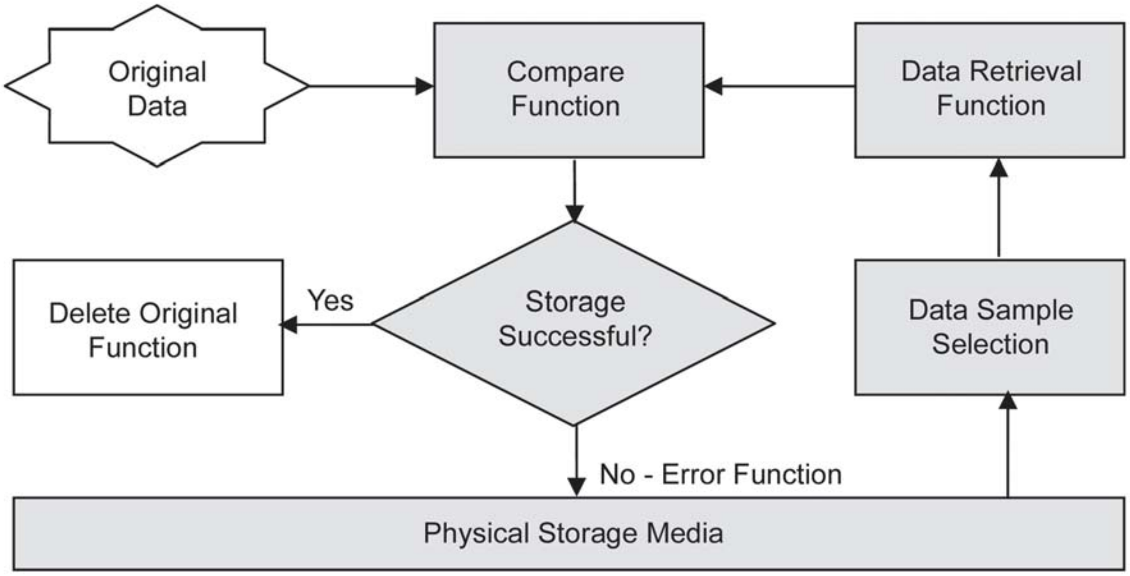
When the AIP has been created and there is a Request for Archiving (scheduled batch-job, manually initiated, etc.), a Receive Data function takes the AIP and makes the physical transfer of the data package in question.

Figure 6.6: Archival Storage – Adding to Archive



After the transfer is completed, the stored AIP should be verified. This has only been shown schematically above. Figure 6.7 provides further details.

Figure 6.7: Archival Storage – Delete Original Data



The original data can be deleted when a successful archiving has taken place - but not before. To be certain that a physical transfer of data is achieved; the archived data should be verified against the original data. Since deleting of original data is a critical operation, there should be an approval process in place that involves the owner of the source data (Application Data Owner), the owner of the archived data (Archive Data Owner), and the Quality function (QA).

6.2.3 Archival Storage – Other Functions

Apart from receiving data to be archived, the Archival Storage function is responsible for:

- Managing storage hierarchy
- Replace media
- Error-checking
- Disaster recovery
- Provide data for the access function; see Section 6.2.4 of this Guide

The Manage Storage Hierarchy function controls the placing of data on the archive media, i.e., it is the control function for the storage media. It interacts with the Receive Data function and the Error-checking function. The media is managed in accordance with the policies determined by the Administration function.

The Replace Media function handles media upgrades that are considered to be straightforward, i.e., data migration is well-defined and does not affect the Content Information and Preservation Description Information as defined in Section 6.1.3 of this Guide. It is possible that the Packaging Information is affected by the media replacement. Examples of straight forward operations include media refreshment, replication, and repackaging. More complex data migrations are handled by the Administration function.

The Error-checking function provides assurance that AIPs have not been corrupted during internal data transfers. This function is reliant upon built-in error-checking and reporting within the EDA hardware and software. In this respect, the Information Package Fixity information provides assurance against undetected corruption. Due to the often considerable size of the archive, a statistical sampling technique may be used to establish that the archive is operating error-free.

Disaster Recovery includes storing archived data in two physically separate locations so that in the case of a failure of the storage media, data can be recovered from the alternative location. For on-line storage of archived data, a backup procedure should be in place to ensure that the archived data is not lost in the event of a hard-drive failure in the on-line system. For passive storage using CD or DVD, it is recommended to make three copies of each media so that by using a voting process the correct content can be determined. This, coupled with regular verification of the media, will minimize the risk of losing the archived data.

6.2.4 Access

When data is archived on any media, it is important to be able to find the data again when needed. It is the role of the Access and Data Management functions to provide this functionality.

In general, the Consumer would communicate with the archive via the Access function by means of:

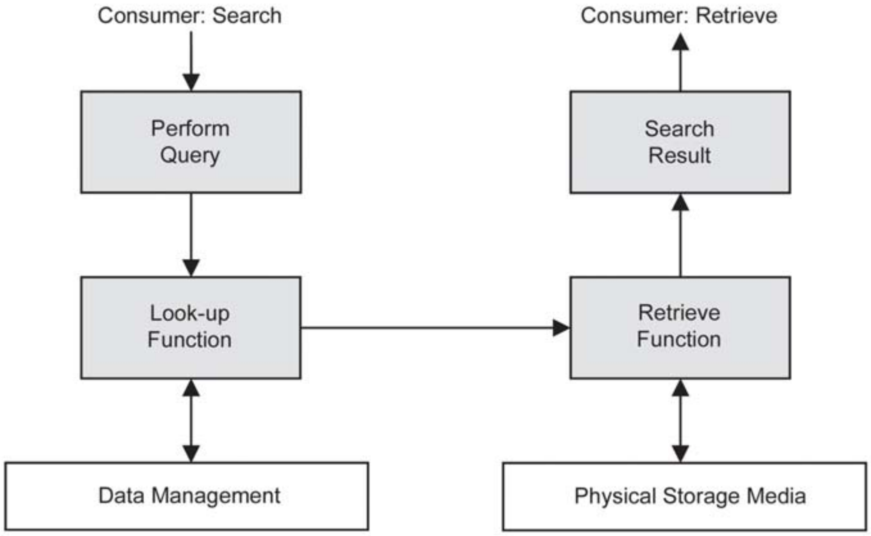
- Query Requests that are handled by Data Management and return results, but not the actual Archival Information Package (AIP). The query could be, for example, a request for availability of certain information such as study report, sterility data, SOP.

- Report Requests are similar to the query request, but may involve a combination of several query requests with the retrieved results formatted using report templates.

- Orders involve retrieving the Archived Information Package (AIP) and making this available to the Consumer in the form of a Dissemination Information Package (DIP).

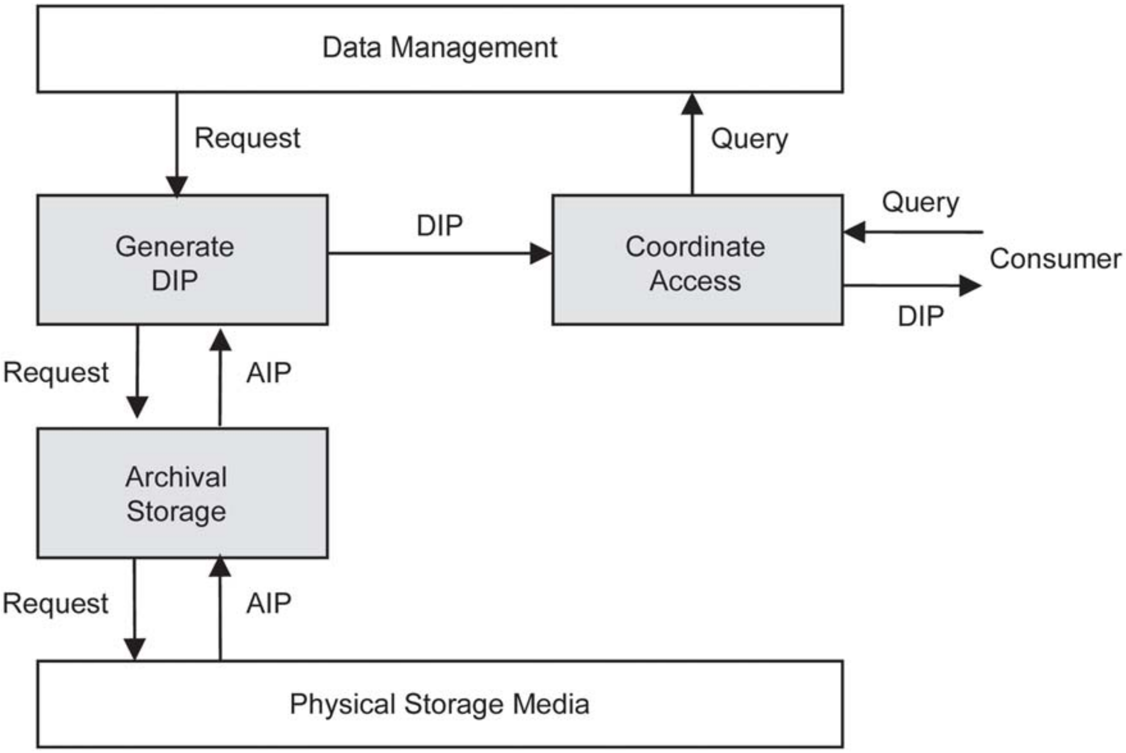
A simplified flow diagram for a generic search/retrieve process is shown in Figure 6.8.

Figure 6.8: Search/Retrieve Function



A simplified data flow diagram for an Order using the OAIS terminology is shown in Figure 6.9.

Figure 6.9: Order Request



The Co-ordinate Access function is responsible for the admission control to the archive. Admission control is important as archived data may need to be protected due to legislation and/or regulation or for commercial reasons. Examples of such data include drug master files, clinical data, patient data, audits, price information, and patents.

When data is stored on removable media, admission control involves the control of physical access to these media. It may be necessary to keep the archive in a physically closed environment to secure proper access to data.

If password protection is used on archived data, it is important to maintain proper management control of such passwords to make the data accessible to only authorized individuals both now and in the future.

6.2.5 Data Management

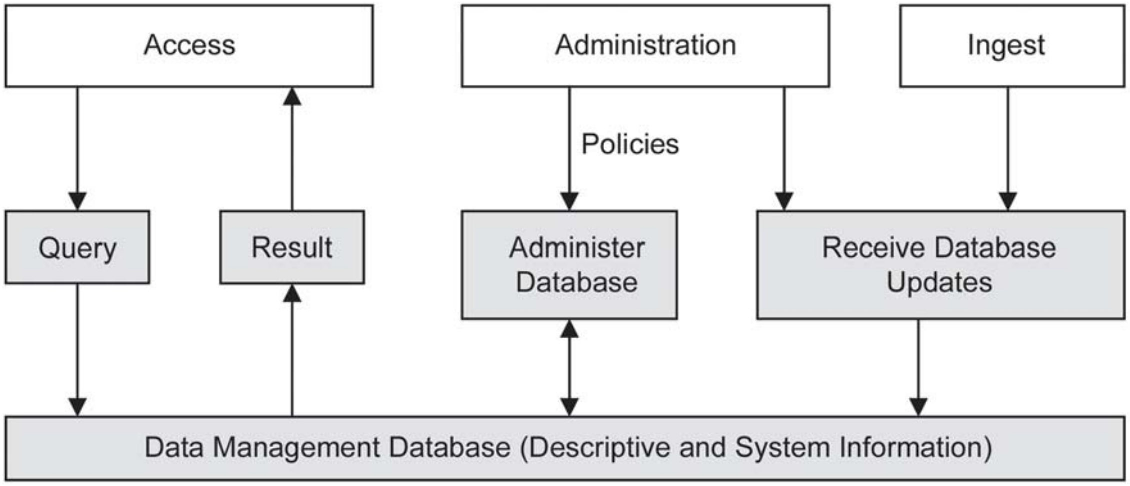
The Data Management function is responsible for the control and integrity of the Data Management database, which contains both Descriptive Information and System Information. The Descriptive Information includes metadata to support searching and retrieving the AIP from the archive. The System Information enables the archive to operate. The search and retrieval of requested archived data will be possible by using this Data Management database. The database also will explain how the archived data can be read.

The archived data should be maintained continuously. It is important to ensure that the Descriptive Information and System Information, as well as the physical media, which hold the archive data, are always valid and accessible.

The Data Management function is responsible for creating any schema or table definitions required to support defined functions; for providing the capability to create, maintain, and access customized user views of the contents of this storage and for providing internal validity checks of the content of the Data Management database.

The Data Management function is summarized in Figure 6.10.

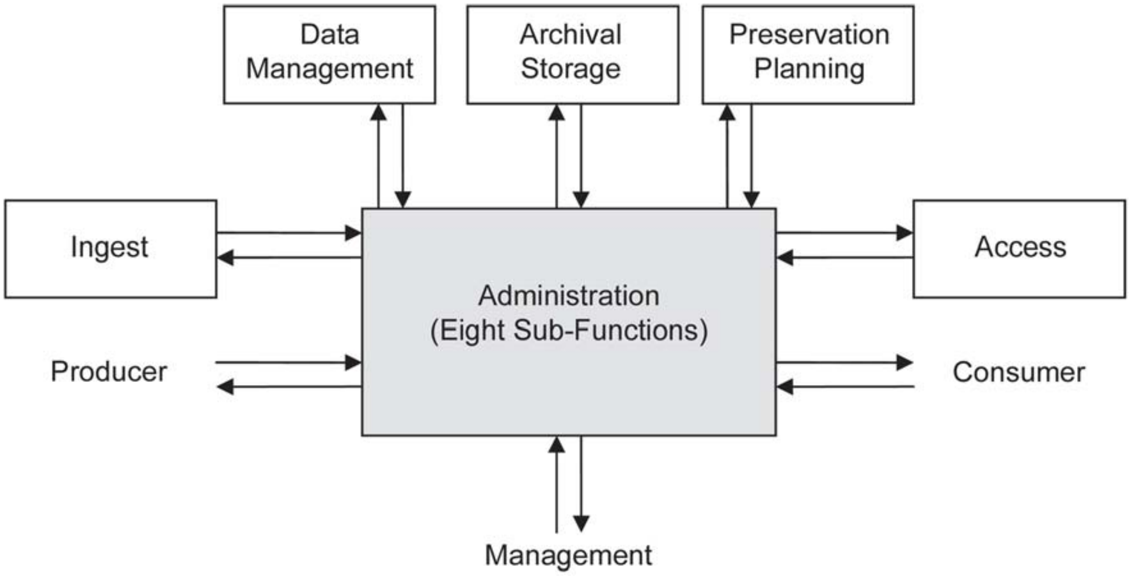
Figure 6.10: Data Management



6.2.6 Administration

The Administration function communicates with all other OAIS functional entities as well as the environment (Producer, Management, and Consumer) as shown in Figure 6.11.

Figure 6.11: Administration



Administration contains eight sub-functions, which receive and provide information as follows:

- The Negotiate Submission Agreement function agrees the SIP format with the Producer based on templates received from Preservation Planning. The function maintains a schedule of expected submissions.

- The Audit Submission function verifies that SIPs and AIPs received from Preservation Planning comply with the agreed standards. Audit reports are sent to Producer and Ingest.

- The Establish Standards and Policies function develops and maintains standards based on policies received from Management as well as recommendations from Preservation Planning. These standards cover items such as management of the archive, disaster recovery, and security. The policies and standards are communicated to Ingest, Data Management, and Archival Storage.

- The Physical Access Control function restricts access via control of door locks, etc.

- The Manage System Configuration function provides continuous monitoring on the performance of the EDA based on information received from Archival Storage and Data Management. It also controls changes to the configuration based on information from Preservation Planning.

- The Archival Information Update function is the mechanism for executing changes and as such interfaces with Access (DIP) and Ingest (SIP).

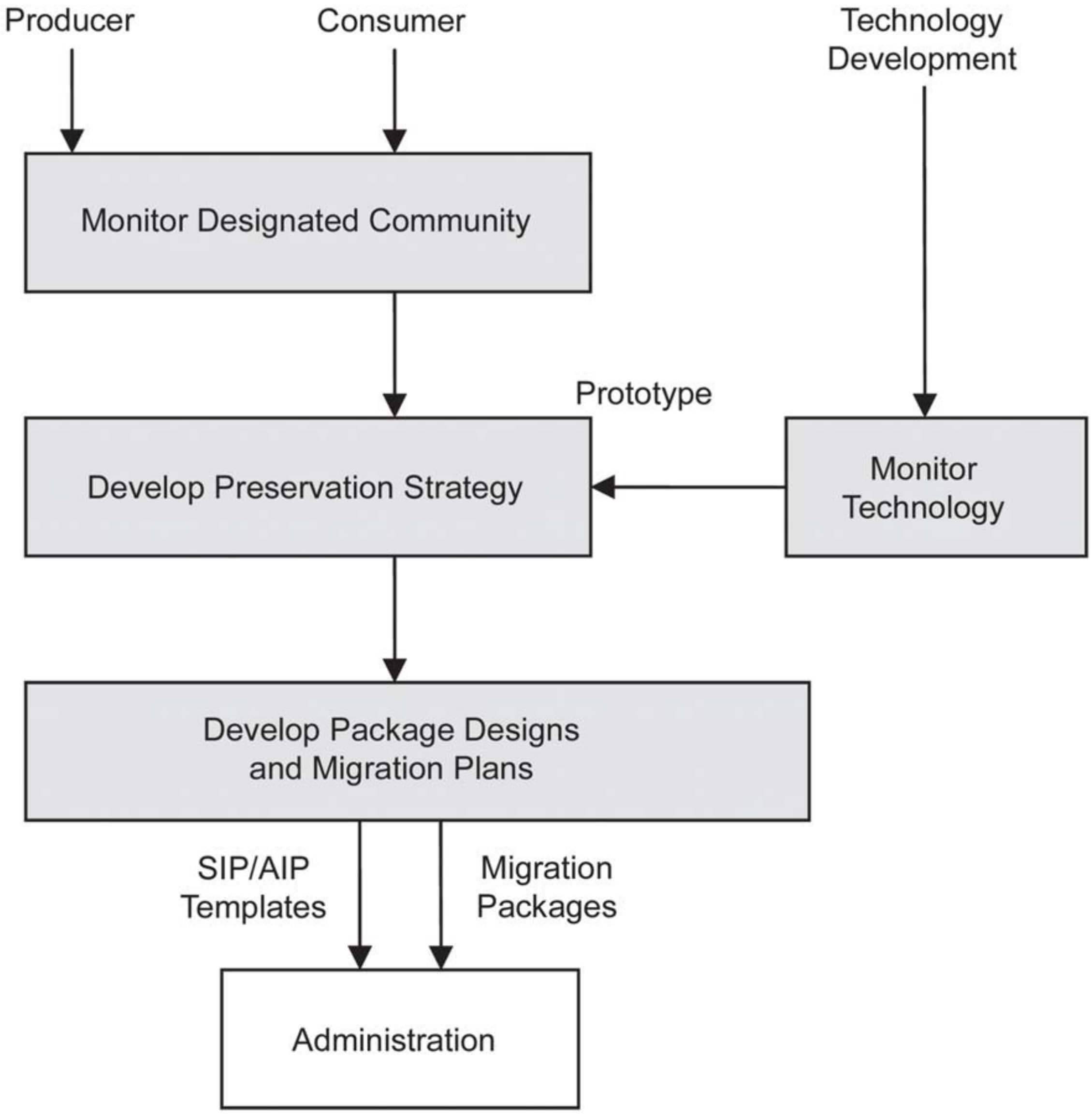
- The Activate Requests function handles scheduled and event-triggered requests from the Consumer and passes these on to Access.

- The Customer Service function handles the consumer accounts and provides billing information (based on data from Access) and collects payments from Consumer. It also passes on consumer comments to Preservation Planning.

6.2.7 Preservation Planning

Preservation Planning contains four sub-functions as depicted in Figure 6.12.

Figure 6.12: Preservation Planning



Monitor Designated Community interacts with all stake holders, soliciting information that may have an impact on the future functioning of the EDA, such as operational requirements, software and format changes, etc. These requirements are then passed on to Develop Preservation Strategy.

Monitor Technology tracks developing technologies and develops prototypes as required.

Develop Preservation Strategy uses the received information, including performance data from the operation of the EDA to develop future preservation strategies and standards that are designed to maintain the integrity and effectiveness of the EDA.

Develop Packaging Designs and Migration Plans is the final step in the link by providing Information Packages such as SIP and AIP that are consistent with approved standards. This function also provides migration plans to satisfy changed standards.

Note that the diagram and explanations of sub-functions have been simplified from the OAIS model.

Appendix C
Metadata

7 Appendix C: Metadata

This appendix contains a summary and reference list over the most commonly used standards for metadata.

7.1 Overview

Metadata gives context and meaning to data. That is, calling a data item or data structure metadata does not explain what kind of data it is; rather, calling a data structure metadata explains that it is regarded, in that context, as ancillary information pertaining to something else – usually an information resource of some kind (called an item for the purposes of this Guide).

Apparently similar pieces of metadata can fulfill different roles. For example, simple properties within metadata could be any one of:

- Properties of a particular item, e.g., a last-modified date or word count.

- Properties common to a class of items to which the item belongs, e.g., the barcode on a notebook is common to all notebooks of that type and manufacturer.

- Properties that an item has by virtue of its participation in a business process (i.e., the current state of an instance of a business process). For example, most identifiers and status codes are of this kind. This is particularly useful to bear in mind when there are apparently conflicting requirements for status or identification metadata – there is probably more than one business process involved with their respective requirements needing to be untangled.

- Properties that are assigned for the benefit of a particular search/retrieval methodology, e.g., Dewey decimal codes on copies of books in a library; Dublin Core metadata elements on electronic information resources.

- Properties that are assigned for the benefit of a particular indexing or knowledge organization methodology, for example, SNOMED CT coding of clinical records.

It is important when analyzing metadata requirements not to accept metadata as a sufficient description of any data item; its role and purpose should be made clear and verified against the functional requirements of whatever process or tool will be using the metadata.

More complex pieces of metadata are even more varied in their scope and purpose. In fact, while sometimes it is clear that an item has metadata, in other situations, a complex data structure represents from one point of view, an item with metadata, and from another point of view, a composite information item that has the smaller item as a component.

The benefits from using recognized standards in metadata go beyond interoperability with other systems using the same standard, and include free reuse of the analysis and design work that has gone into designing a metadata standard to support a common business or information management function.

There is a broad literature on metadata, including perspectives both from IT and from library and information science (a very important perspective sometimes forgotten in IT). A general discussion of metadata in the context of traditional and digital libraries can be found in Terence Smith's 1996 paper in D-Lib magazine, The Meta-Information Environment of Digital Libraries, available at: http://www.dlib.org/dlib/july96/new/07smith.html.

There is a very wide range of metadata standards, and this brief introduction will inevitably only touch on a few. The focus here is on metadata that is likely to be used in digital libraries, archives, and collections of records with a bias in the examples toward Europe and the Life Sciences industry. Unfortunately, this subject area is relatively bad for dead Web-links and lost Web sites, and some very good reference information resources (such as the European Diffuse project) have lapsed. Examples given below have been selected for relative stability and long life.

Generally, metadata standards are either standards for metadata content or standard ways of representing metadata as data (e.g., in XML or as fielded records). Some standards cover both of these, e.g., the MARC bibliographic records standard (http://www.loc.gov/marc/) specifies both content and format.

7.2 Kinds of Metadata Standards

Acknowledgement: the descriptions below are mostly summarized from descriptions on the referenced Web sites.

7.2.1 Terminologies

These provide standardized lists of terms that are then used as values in metadata. For example:

The Eurodicautom Multilingual Thesaurus http://europa.eu.int/eurodicautom/Controller
Eurodicautom is the European Commission’s multilingual term bank. It is used by translators, interpreters, terminologists, and other linguists worldwide over the Internet, where it records a daily average of 120,000 enquiries. Entries are classified into 48 subject fields (ranging from medicine to public administration). A typical entry contains the term itself and its synonyms, together with definitions, explanatory notes, references, etc. At present, the term bank contains about five and a half million entries (terms and abbreviations) subdivided into more than 800 collections. Eurodicautom covers a broad spectrum of human knowledge, but is particularly rich in technical and specialized terminology (agriculture, telecommunications, transport, legislation, finance) related to EU policy.

SNOMED CT http://www.snomed.org/snomedct/
The SNOMED CT Core terminology provides a common language that enables a consistent way of capturing, sharing, and aggregating health data across specialties and sites of care. Among the applications for SNOMED CT are electronic medical records, ICU monitoring, clinical decision support, medical research studies, clinical trials, computerized physician order entry, disease surveillance, image indexing, and consumer health information services.

A wide-ranging list of industry sector metadata initiatives was compiled by OASIS a few years ago, and is still useful. See http://xml.coverpages.org/classification.html

7.2.2 Information Management Standards

These provide standardized bases for common information management functions, such as archiving and records management. For example:

Records Management: Several national archives organizations have published metadata standards for electronic records management and there has been some work on international harmonization. An ISO standard for records management was published in 2001 (ISO 15489). The U.K. National Archives advice and standards can be seen at http://www.nationalarchives.gov.uk/electronicrecords. Their metadata standard (in their 2002 Requirements) is an integral part of a broader framework of policy and guidance and also is designed to conform to the U.K. e-Government Metadata Standard.

Dublin Core (http://dublincore.org): The Dublin Core metadata elements are a small set of elements designed for general purpose resource discovery in distributed network environments, especially the WWW. The Dublin Core Metadata Initiative (DCMI) is an organization dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems.

7.2.3 Standards for Representing Metadata

These provide standardized ways of describing metadata and representing it in various specific data interchange technologies such as XML. For example:

ISO/IEC JTC1 SC32 WG2 (http://metadata-stds.org/) is the ISO Working Group that develops international standards for metadata and related technologies, notably ISO 11179 concerning metadata registries.

Representing Metadata in XML: there is an overwhelming amount of information on this with many approaches and projects. This Web site is a good starting point to make sense of the subject: http://www.semanticweb.org/

7.3 Example of Metadata

For a single reported result obtained from a batch of samples analyzed by HPLC-UV, typical metadata would include:

- Operator's name
- Instrument/HPLC system identity, software name, and version
- Acquisition method including:
  - Instrument control (pump flow-rate, column temperature, solvent composition, solvent gradient, sample injection volume, detection wavelength)
  - Data acquisition parameters (sampling rate, start/stop events, scaling)
- Sequence (run list) including sample identity, sample type (e.g., standard, QC) matrix, source, dilution factor, and auto-sampler position
- Chromatograph trace (i.e., the digitized detector response)
- Scale (i.e., the X and Y axis)
- Processing method, including integration parameters (start/stop) and peak identification (name, retention time)
- For each method: Method name, Author, Revision, Revision History, Audit trail of changes

Appendix D
Databases

8 Appendix D: Databases

This appendix summarizes some of the key characteristics and considerations pertaining to archiving of databases.

8.1 Overview

There are a number of possible solutions and scenarios to be noted when considering the preservation of electronic data originating within databases. These include the following:

- Retaining the data within the originating production system.
- Reprocessing, readability, and information loss when data is archived and subsequently retrieved.
- Using OLAP/Data Warehousing Technology in long-term record preservation
- Other (non-OLAP) Database Strategies

Each of these is discussed in more detail below.

8.2 Data Retention in the Production System

On-line retention of data can provide an adequate means of archiving data for databases with a fairly shallow physical growth profile, provided that the application software provides adequate security functionality and maintains the integrity of the data.

In some circumstances, on-line maintenance of data can be facilitated by utilizing Hierarchical Storage Management (HSM) technology, where data storage is automatically managed between on-line and off-line media to suit the associated access frequency. However, it is difficult to apply this technology to most popular relational database systems, such as Oracle or SQL Server.

Even though disk capacities are rapidly expanding (and becoming cheaper) the exponential overhead associated with the daily maintenance (e.g., backup) of production databases suggest that indefinite retention of electronic data within a rapidly growing database is not a viable long-term solution.

Where the production environment is upgraded or replaced, it is necessary to migrate the associated data if this is to remain available on-line. The integrity and continued availability of data should be assessed via an appropriate risk assessment and/or upgrade strategy.

8.3 Reprocessing, Readability, and Information Loss

Where it is impractical to retain the data on the production system, the data should be archived to another system. The following potential issues need to be considered where such data archive is contemplated:

- When archiving a database, or part of a database, consideration should be given to the degree of reprocessing that might be required for archive retrievals. There is a balance to be made between archiving the whole database with full retrieval functionality and archiving a single report from the database. It is recommended that the amount of reprocessing required should be defined at the time of archiving.

- For reasons of consistency and efficient updating, database applications are typically normalized to exclude duplicate data. Therefore, only a single copy is stored of each static data item and aggregate information is calculated dynamically, rather than being stored.

- Databases typically contain far more information than is explicitly included in routine output. To archive data only, without application logic, is to be certain of losing some information. The Archiving Strategy should determine (based on risk) what information should be kept and what may acceptably be discarded.

- Similarly, to archive data only, without application logic, is to be certain that the ability to reprocess data will be lost. However, this may well be acceptable, based on risk. A very achievable strategy is to retrieve data from the archive into an off-line copy of the application, should reprocessing be necessary, and accept that at the point where the application is retired, the ability to reprocess will be lost.

- The ability to reprocess data is distinct from its readability, which refers to the ability of the archive to output data in human-readable form. This also may require some visualization software to be available (or developed) in addition to the raw data.

8.4 OLAP/Data Warehousing Technology

Data Warehousing or On-Line Analytical Processing (OLAP) technology has many features that could allow it to be used in long-term data preservation. OLAP techniques extract data from a company's applications or On-Line Transaction Processing (OLTP) systems and hold it in a form which is optimized for Read-Only analysis. The extraction typically uses an Extraction, Transformation, and Loading (ETL) tool which may extract a sub-set of the data, verify the data integrity, perform some summarization, and populate the OLAP tables. The extraction is typically not real-time; instead, it is intended to allow analysis of past business performance.

Use of OLAP technology in an archiving context offers some compelling advantages, but also carries some serious disadvantages. The arguments for and against are summarized below.

Advantages of an OLAP technology archive solution include:

- Commercial Off-The-Shelf technology

- Commercial tools are available which can amalgamate many proprietary data formats from disparate application systems into a single database

- Optimized for efficient retrieval from large datasets

Disadvantages of an OLAP technology archive solution include:

- Extracted data is not deleted so the typical OLAP approach is not a true archive, where the data is removed from the source system once successfully copied to the archiving system.

- The extracted records are not the original records. Nor are they necessarily certified, accurate, and complete copies although they could be validated.

- Often no ability to reprocess records

- OLAP system formats are proprietary database formats. They are not platform or technology neutral and are not ideal for long-term preservation.

- OLAP vendors currently do not perceive archiving or long-term preservation as an OLAP application.

A possible archiving solution is to use OLAP in the medium-term and a different approach for the long-term. This is not to be recommended, as such a two-stage solution would be hugely complex and long-term issues would remain. If OLAP technology is to be used for an archive, then it should be for a long-term solution, where the disadvantages above have been addressed.

In conclusion, OLAP is not currently a recommended strategy. Other (non-OLAP) strategies are considered below.

8.5 Other (Non-OLAP) Database Archiving Strategies

In all cases below, the term ‘extract’ includes deletion of the records from the active system. The extract process should meet transactional integrity criteria, i.e., Atomicity, Consistency, Isolation, and Durability (ACID) properties.

8.5.1 Archive of Whole Database

For small databases (e.g., MS Access), taking a complete copy of the database can be a viable option for routine archiving of data. The iteration of such a process will; however, result in the accumulation of multiple copies of the same (ambiguously defined) data and could raise issues with the indexing and/or cataloguing within the archive.

For larger databases (e.g., Oracle/SQL Server), it is not desirable to take an entire copy of the database to archive a selected number of records, particularly if the archiving process is conducted frequently and the previously archived copies need to be maintained.

Selective retrieval of a specific record from an entire database backup to the production system is not easy (or often technically feasible) without affecting other records already in the production system.

If the database system is to be upgraded, then the integrity and continued availability of records should be assessed via an appropriate risk assessment and/or upgrade strategy. If it is necessary to maintain previously archived database copies, then these copies should be migrated to the new database platform to ensure that data definitions and database schema are compatible with the new environment.

When the system that supports the database is finally retired, without any upgrade, it is likely that the entire data content of the database will be archived. Ideally, the application logic also will be preserved although this is extremely difficult to achieve without the computer museum approach.

8.5.2 Archive of Specific Data

The alternative to archiving the whole database is to archive specific records. This can be achieved in one of two ways, namely:

- Archiving the dynamic data alone without the static metadata
- Archiving the dynamic data together with the static metadata

The following comments apply to both approaches:

- The subtlety of the internal structure and inter-relationships of records stored within database schema is often only fully understood by the system/application vendor. Consequently, special consideration is required to ensure that a complete and accurate copy of each record and supporting metadata is obtained during the archiving process.

- The records selected for archive should be specific to the predefined business process. For example, in a GLP context, the records to be archived for a particular study (or study part) should be complete to meet predicate rule requirements, but specific to that study.

- The selection criteria (e.g., Study number or Batch identity) used to extract the database information also is likely to be key metadata used to index and catalogue the records when they are placed within the archive.

- The best solution to the issue of record selectivity when archiving records from a database is the provision of an interface within the application software. This enables the selection of all relevant records and can produce a file (or composite group of files) for archiving purposes that represent a true and accurate copy of the original electronic records, as maintained within the database. A corresponding interface to enable the reconstruction of the database schema from the extracted file(s) will subsequently enable the review and/or re-processing of the original records from within the production environment. Often, such functionality also will allow the surgical removal of the original records from the production database, which can subsequently reduce the overhead associated with the operational maintenance (e.g., backup) of the production environment.

- When considering the use of proprietary solutions, which often include the computer platform, system software and storage media, it is important to note that many such solutions bundle records together into discrete files which may be in one of numerous formats, not all of which are human readable. Where reprocessing or periodic deletion of records is not necessary this may not be an issue, but where individual records need to be deleted this can present a problem. It may be possible to overcome this by storing records of similar type and age in the same file/folder, and then deleting the entire file/folder at the appropriate time. However, this approach requires significant forethought when designing the archive structure.

- Often the exported file is in a proprietary binary file format (or has a proprietary internal structure) that is only meaningful to the application that generated it. In these circumstances, attempts to export information into a public domain format/structure can reduce the risk associated with the subsequent retrieval of information.

- If the production system is upgraded or replaced and the new input/output interface of the application maintains backward compatibility with previously archived records, then this can significantly reduce the risks and effort associated with the upgrade/replacement strategy. Under such circumstances, previously archived records may not require platform or format migration.

Further specific comments relating to the archiving of dynamic data, with and without the associated static metadata, are outlined below.

8.5.3 Archive Dynamic Data (Without Static Metadata)

The dynamic data is archived without the static metadata. This may be scheduled based on record age or triggered by record status change; the issues are similar. In order to generate meaningful information, some logic will be needed to combine the archived dynamic records with the non-archived static metadata. This could be by retrieval into an offline copy of the application in the short- or medium-term (while the application remains available).

8.5.4 Archive of Dynamic Information (With Static Metadata)

A complete record of all information required in the long-term is archived for each dynamic record (including a copy of referenced static metadata and aggregate data). This may be scheduled based on record age or triggered by record status change; the issues are similar. Suitable database output tools may pre-exist. For example, a query used to produce a summary report at the end of the record lifecycle may be adequate. However, it is more probable that a specific query or software tool is required, which extracts the information required in the long-term and outputs it in a platform-neutral format. Following a successful export only the dynamic data should be deleted.

8.5.5 Strategy Comparison

Table 8.1: Comparison of the Three Database Archiving Strategies

Consideration	Whole Database	Dynamic Data	Dynamic Information (Metadata)
Archive format can be platform-neutral (XML, PDF...)	Yes	Yes	Yes
Record processability	Maintained up to system retirement and then lost.	Lost unless retrieval into the application is possible (pre-system retirement).	Lost
Risk of strategy failure	High – The strategy may have to be implemented as an emergency at system retirement.	Medium – Separate strategy is required for static data. Risk of static data loss in the event of system failure.	Low
Monitoring and demonstrable adequacy of the strategy throughout system life	Poor	May be adequate if tools are developed to produce meaningful information from the bare dynamic data records.	Excellent
Ease of implementation (and validation) of the archive process	Very easy	Easy – Possibly use database tools rather than software development.	Hard – Custom software development may well be required for Commercial Off-The-Shelf systems.
Ease of implementation (and verification) of visualization tools needed to retrieve information from the archive	Hard	Hard	Easy
Preservation of data	Complete	Complete for dynamic data. Separate strategy is required for static data. Risk of static data loss in the event of system failure.	Data will be discarded (the database will contain data not sent to the archive). Careful export format definition is required to ensure correct partition of archived/discarded data.

Table 8.1: Comparison of the Three Database Archiving Strategies (continued)

Consideration	Whole Database	Dynamic Data	Dynamic Information (Metadata)
Preservation of aggregate data	Aggregate data will almost certainly be lost (as the application logic is lost).	Aggregate data will almost certainly be lost (as the application logic is lost).	Aggregate data can be preserved by careful export format definition.
Preservation of information	Information will almost certainly be lost (as the application logic is lost).	Information will almost certainly be lost (as the application logic is lost).	Information can be preserved by careful export format definition.
Replication of data in the archive	Low	Low	High (static data will be replicated many times).

8.5.6 Summary

The first strategy (whole database) is low cost. However, it carries a high risk that long-term record availability and readability will be compromised.

The second strategy (dynamic data) is also low cost. There is a medium risk that information will be lost (unless record visualization software is developed).

The third strategy (dynamic information) is low risk and carries a correspondingly high cost.

There is no clear distinction between the second and third strategy: a hybrid approach where some static data is included (e.g., using a database view) is possible, provided that the static data is not deleted following export.

In summary, there is no single, correct, recommended approach.

Appendix E
Laboratory Systems

9 Appendix E: Laboratory Systems

This appendix summarizes some of the key characteristics and considerations related to the archiving of laboratory system data, and in particular, standards for laboratory data.

9.1 Overview

Only the simplest types of analytical instrumentation produce electronic data in a purely numerical form, e.g., a balance or a pH meter. Most other instrument types, e.g., HPLC-UV, HPLC-MS, NMR, IR, produce multi-dimensional detector response profiles that require complex software visualization in order for scientific interpretation to be conducted. Subsequent processing and assessment of these profiles also can involve complex numerical algorithms within the software, e.g., to integrate peak areas or to perform regression analysis to calculate concentration values from a calibration curve.

Although the probability and risk associated with retrieval or reprocessing of archived analytical data is likely to reduce with time, the key objective remains that the record can be retrieved in a suitable format to meet its pre-defined business purpose throughout the retention period. Historically, with a paper-based raw data medium, regulatory (or business) expectations would be technically limited to being able to review the data, e.g., a printed chromatogram held within the archive. Electronic raw data provides new possibilities (and expectations) about the operations possible with archived data.

It is common place for instrument raw data files to be stored in a proprietary binary format which requires the parent instrument software in order to be opened and decoded.

The ease of achieving and maintaining a long-term laboratory data management strategy is inversely related to the number of different software types/versions used within the laboratory environment. An effective change management process is a key requirement in controlling the rate of changes and upgrades at the minimum level required to continue to meet scientific/business needs.

9.2 Commercial Scientific Data Management Solutions (SDMS)

In a laboratory context, there are scientific data management products that offer benefits in addition to a conventional IT backup solution. Such systems may:

- Automate the capture, storage, searching, and retrieval of electronic records from laboratory instrumentation and reduce the overhead of operating an electronic data archiving process.

- Consolidate the software required to re-visualize data from multiple instruments/applications; however, this solution is likely to adopt a proprietary format.

Such systems can decode the proprietary data file formats often used within laboratory areas and extract metadata that can supplement the index parameters used to catalogue archived records.

Commercial SDMSs are limited solutions, as they do not address the requirement to be able to re-process the data.

9.3 Conversion to a Technology/Vendor-Neutral Format

There have been several attempts at defining and implementing standard file formats for laboratory data. The two most widely recognized standards for High Pressure Liquid Chromatography (HPLC) and Mass Spectrometry (MS) data are AnDI, originally coordinated by the Analytical Instruments Association (later adopted by the ASTM) and Joint Committee on Atomic and Molecular Physics Data Exchange (JCAMP-DX). Such standards have been partially successful in facilitating a long-term data migration and management strategy, but are not universally relevant to all instrument types, nor are they adopted by all instrument vendors.

A new US standard ASTM E13 is in development (spectroscopy and chromatography) which is aimed at both instrumentation vendors and science-based companies. This includes the definition of a new XML schema to replace JCAMP/AnDI.

The use of XML does not represent an end-point to long-term data management and retrieval issues. Key decisions relating to schema definitions that describe the data structure for different analytical techniques (HPLC, LC-MS, GC, UV, etc.) are still required. The XML-based Analytical Information Mark-up Language (AnIML) standard developed primarily by NIST, but now under the guidance of ASTM is considered popular.

Key to the success of this latest approach is that:

- The schema definitions are publicly accessible and available for adoption by all vendors.
- Vendors commit to the standards and include functionality to export and import records using these formats in new versions of their laboratory software.

A true end-point will be reached only when platform and vendor-neutral tools are available that allow the data (stored in an appropriate open standard format) to be re-processed.

Appendix F
Archiving Approaches

10 Appendix F: Archiving Approaches

This Section gives an overview of some of the more commonly used current techniques and methods for retaining digital records. Techniques covered include:

- Increase of on-line storage capacity
- The mothball/time-capsule/computer museum approach
- Format/platform migration and emulation

For each technique, the potential risks and loss of information are considered. Although not a digital preservation technique, the option of transferring to non-electronic media is covered, as are archiving to commonly used file formats and storage media considerations.

This Guide makes no recommendation on which preservation technique to use. Each presented technique has benefits and disadvantages, and the selection will depend on the particular circumstances for archiving. Further considerations may be imposed where legacy systems are involved since these have generally not been designed with electronic data archiving in mind.

This Guide identifies key issues affecting choosing a commercial solution, an in-house custom solution or a modified proprietary solution, but does not make specific recommendations in this regard. The same considerations apply as for other capital projects in the life science sector although for archiving the long-term maintenance aspects should be given prime consideration.

10.1 Increase On-line Storage Capacity

In general, computer storage capacity is inexpensive and retrieval times are fast.

In the short-to-medium term, the problem of managing a continually expanding data store may be solved by purchasing more capacity and continuing to store data on the on-line system. However, this may be countered by the need to store files of increasing size. There are a number of considerations:

- Most archived records are static, revisions are rare, and retrievals are infrequent. There may be issues with system performance, even with high-speed storage devices. In this situation, it can be difficult to justify congesting disk space that supports live data with archived records.
- There is a potential for undetected changes or accidental alterations in a live system, even with background processes monitoring the integrity of files and storage media. Secure partitioning of archive data should mitigate this risk.
- It may be impracticable to separate and isolate the various user groups for live data and archived records. These user groups are not normally the same, and in particular, there is a need to demonstrate tight control of the usually restricted group of users for the archived records.
- Some regulations (e.g., GLP) require data to be stored in a (potentially separate) archive under the control of an Archivist. Using the on-line system will potentially violate this requirement and will require additional controls to be put in place.

- Frequent backups may be performed on large volumes of static records, possibly leading to overnight backup jobs spilling over into the working day. There also will be increased costs for the maintenance and management of the additional backup media although it could be argued that this cost would have been incurred for a separate archive system. Using incremental backups or mirror/ghosting may help to alleviate the time taken for backups.

- An increasing volume of records will be migrated each time the on-line system is upgraded. There is increased potential for corruption or loss, especially where the old data files need field changes/additions/deletions to match the new data structure.

- It may be important to distinguish the version of the application that was used to generate individual records. Often this can be difficult if all records (from all versions of the application) are maintained in a single pool of records.

- If the application is upgraded it also may be necessary to perform format migration (see Section 10.3 of this Guide) on old records to ensure that they do not become obsolescent while they are maintained on-line.

- Indexing of metadata for expedient retrieval may not be optimized in the live system. This means that the retrieval of static archive data may not be efficient when stored on-line.

Moving older and more static data to different hardware platforms can mitigate some of these considerations. An example of this approach might be to move files from on-line high speed disk to near on-line WORM disk or tape. This is sometimes referred to as platform migration and is an integral part of Information Lifecycle Management (ILM). It also may be referred to as Hierarchical Storage Management (HSM) or tiered storage.

10.2 Mothball/Time Capsule/Computer Museum

Theoretically, archived information can be made accessible for a potentially indefinite period by simply preserving the hardware and software hosting the records. The technique may be known as mothballing, time capsule, or the computer museum approach. It is a good way to preserve content and meaning since the data is stored with the system on which it was created and maintained.

The question arises of how long an obsolete machine can be maintained. Storage media and computer components all wear out, and due to limited availability, replacing worn out parts becomes increasingly expensive and impracticable. There also may be contractual/licensing complications and withdrawn support of legacy platforms and applications.

There is a risk of total data loss through terminal failure of the mothballed system. This situation requires a disaster recovery plan that has been proven to work and requires a replacement system or parts to be available.

There should be sufficient people in the organization with the required knowledge to access, operate, and maintain the ageing system and the obsolete platform. This knowledge can be kept current only through regular work on the system, and as time progresses, this will become an increasing challenge. For more complex and unusual maintenance tasks, it is common to rely on the system supplier. Unless engineers and technicians have the opportunity to work with the system on a regular basis, the knowledge of how to best maintain the system will be lost. User manuals and documentation describing system configuration should be maintained in safe storage.

Therefore, mothballing is recommended only as short-to-medium term solution. There should be a clear plan for the next stage that addresses record retention beyond this period.

10.3 Format Migration

Format Migration is the process of transferring data from one electronic format to a different electronic format. Replacing obsolete systems or transferring data from the on-line system to an archive will usually require a format migration and records may be migrated more than once during their life.

When up-grading an application, operating system, or hardware, the data content should migrate to the new application and platform. This migration may have implications on the format, but normally the supplier of the upgrade is able to provide details of any restrictions of use and their implications. It is recommended that such information is sought before commencing upgrades.

Format migration, for all but the simplest data structures, is complex. There are likely to be changes to both data and file structures with interactions between application programs and multiple files. There is a high risk that the process can result in loss or corruption of data. An example of this is where the original supplier has provided a poor specification of the file format, leading to incomplete conversion of all data fields. Therefore, the migration process should be properly assessed for the associated risks and validated accordingly. Following migration there should be robust testing, which can be time-consuming and expensive.

It may not be possible to migrate all data. Certain types of electronic signatures are held in encrypted forms that require recognition algorithms that are not stored as part of the data.

10.4 Archive to Standard Formats

A particular type of format migration is to convert data into a standard file format that is recognized by many applications and tools, and as such, is likely to be accessible for a long time in the future. Current popular options for EDA standard file formats include PDF and XML, which also are accepted by the FDA, MHRA, and other regulatory authorities.

Another option is to convert the data from the source system to a print file format before archiving, as these files, generally, can be captured and read by other systems. In the case of clinical study data, this also includes transfer to SAS (an industry-standard database) for long-term archival. SAS is considered to be a suitable repository for such data, provided that all data and metadata transfer and migration processes are appropriately validated.

Standards are normally established by international bodies. In other cases, standards are established through widespread use. An example of this type of standard might be a word processor document format used extensively across industry or Adobe PDF.

Whatever its origin no standard has an indefinite life and will eventually become obsolete. Program code and data generated using programming languages that have been superseded are becoming increasingly more difficult to maintain, read, and understand. The replacement of previously standard applications has implications on the ease by which data generated by these applications can be maintained. Applications change with time and the conversion of, e.g., an old MS Word or Lotus1-2-3 document to the latest version of that application is likely to introduce format changes. Changing the manufacturer of applications, e.g., Word Perfect to Ms Word is likely to introduce similar changes. This applies equally to other standard formats.

Standards may evolve over time. An example would be ASCII, which is being expanded to include UNICODE to capture additional characters. When standards evolve they may not be 100% backward compatible.

Converting to XML, PDF, or any other standard requires care and attention to detail; the conversion process should be assessed for the associated risks and validated accordingly. For example, PDF formats generated by different versions of Adobe Acrobat display variations and failure to follow good PDF rendering practice can lead to incompatible or poor quality electronic renditions.

An example is MS Word automated features such as cross-references that easily can be corrupted when upgrading the application. Likewise, XML employs schemas, which in their simplest form determine how the data in the standard XML format is rendered or presented (similar to a style sheet). These schemas are not part of the XML standard, and unless the schemas are widely agreed upon and shared, the context or meaning of data can easily become lost when viewed through an application other than that which created it. Schemas are also subject to revisions and new schemas are being developed, sometimes making the use of earlier schemas inappropriate.

A special consideration is when archiving laboratory raw data as XML or similar platform-independent formats. If re-processing of the data is required, then the full data file structure and definition of all data processing algorithms need to be known and made available in the archive. The use of specific Document Type Definitions (DTDs) allows the structure of the data to be mapped. A large number of biochemical and pharmaceutical DTDs already exist.

The industry has recognized the need for longevity of formats, and as a result, there is a new ISO standard for an archive version of PDF (PDF/A-1). Please refer to Appendix J of this Guide for further information.

Generally, standards are not controlled by the technology supplier (Adobe PDF was an exception to this.)

10.5 Emulation and Virtual Computers

This entails preserving the original data or application software and creating programs that emulate the behaviors of superseded computer systems. This enables the original application and data to be processed (emulated) on contemporary computer architectures. The acronym UVC (IBM: Universal Virtual Computer) has been given to the technique. Emulation techniques retain the original content and context of the record, while overcoming software and hardware obsolescence. Emulation may prove more cost-effective than format migration and may provide more faithful preservation of both content and behaviors.

Although emulation has been performed for a long time (e.g., MS-DOS emulation in the Windows OS), the adoption of this technology for digital preservation is still in its infancy. Work is in progress on a hybrid between migration and emulation referred to as migration on request. In this technique, the data is kept in its original format, while a tool is maintained which allows data to be migrated to a new platform at any point in the future.

Rather than create a new emulator for every previously known operating system each time that a new operating system is introduced, it is more likely that only one emulator would be developed, within which previously developed emulators would run. For example, when Microsoft introduced Windows XP, emulators were written only for Windows 2000, the previous version to XP.

Emulation can extend to the hardware. Old hardware and software platforms are, in this case, emulated in software to run on the current hardware and operating system, thus enabling the original application to be executed. For instance, it is possible to emulate an 80486 PC so that its applications can run on a Pentium 4/MS Windows XP platform. However, one consideration is that the data also may need to be migrated to new storage media, e.g., from 5¼” diskettes (for older PCs) to CD or DVD. Similar considerations relate to peripherals such as printers and their drivers.

It is difficult to demonstrate that an application running under an emulator will perform in exactly the same manner as it did originally. This problem is compounded where emulators run within emulators.

Other important considerations, which also apply to the mothball and migration methods, include:

- Contractual/licensing rights for software
- Availability of user manuals
- Staff that have the knowledge to operate and maintain the system

The method is still subject to on-going research although emulators for relatively simple platforms do exist. For example, it is possible to run Apple Macintosh programs on PCs or to run MS-DOS applications on MS Windows XP.

For more complex environments involving distributed computing or certain Client/Server architectures, this is not a recommended solution.

10.6 Transfer to Non-Electronic Media

This is not a digital preservation technique, but has been included for comparison and because transfer to non-electronic media is a common alternative to electronic data archiving.

Printing on paper or microfiche offers a means of preservation that is not affected by either application or platform obsolescence. In addition, the long-term preservation of paper and film is extremely well understood. Using prints for archive is normally acceptable to regulatory authorities where data can be rendered without loss of integrity, content, or meaning. Using non-electronic media has the advantage that this is the dominant format at time of publication and it works satisfactorily when well managed.

It is recommended that the process of transferring data to non-electronic media should be assessed for the potential risks, taking into account likely requirements for future use of the records, as well as the frequency of access and required retrieval time. These requirements will vary between different types of records and be influenced also by the performance of the archive.

Disadvantages of the approach include:

- Paper archives become very bulky and are likely to become even more so as systems become more complex. Rapid access to records can be hindered and it can be difficult to be certain that all relevant records have been retrieved.
- Paper-based systems usually involve manual elements; the cost of these and the risk of human errors should be considered.
- Retaining the context and meaning of certain records can be difficult, e.g., electronic document review and approvals. Retaining relationships between records from databases can be particularly challenging. Some records may be extremely difficult or impractical to migrate to paper, e.g., a three-dimensional scan.
- It is not usually possible to re-use the data electronically once committed to paper. This could be an issue where re-processing is required to re-assess decisions that have been made based on analytical results.
- Guaranteeing the integrity of data during printing can be both difficult and time-consuming. Where large data volumes are concerned, it may not always be possible to check each and every record and a sampling approach would need to be very carefully designed and tested.

- Depending on a company's data retention policy, it may be necessary to destroy records at the end of their defined retention period. This may not be easy if the records have been grouped on a media item with records of different age and type.

Some of the disadvantages with archiving to non-electronic media can be overcome through an electronic indexing system. This will enable fast searches and relatively rapid retrieval of records. However, the indexing system itself will face many of the same challenges as an EDA, i.e., the risk of becoming obsolete or expensive to maintain. A balance needs to be established between the complexity and sophistication of the electronic indexing system and the advantages this provides, versus the initial cost of implementation and validation and the risk to its longevity through future upgrades.

10.7 Storage Media Considerations

Archived records may be stored on a variety of storage media such as:

- Magnetic disks
- Optical disks
- Holographic media
- Magnetic tape

Each of these media may be of different types, such as high performance disks and compressed disks and different types of CDs and DVDs. The advantages and disadvantages of each media type have not been assessed, as the development and application of storage media is rapid. Selection of media is usually further restricted by what is offered by the selected system supplier.

The following general points are offered for consideration:

- All media has a limited lifetime, e.g., a study by the Library of Congress, USA, showed measurable deterioration in CDs as they age. The safe lifetime of media should be obtained from the supplier, where possible. This may be easier to achieve where well-established media is being used, rather than new technology, in which case caution should be applied until further product data becomes available.

- Media may need to be refreshed. A procedure and safe refresh frequency should be established.

- A routine process to exercise tapes should be established. This should ensure that the tape does not become too tightly wound onto the spindle, which could lead to the medium breaking.

- All media can fail. An appropriate verification procedure based on the perceived risks should be developed. Readability should be verified on a scheduled basis.

- Removable media should be securely labeled with unique identification.

- Appropriate access controls that reflect the intended use of the stored data and the maintenance requirements should be established. This includes protection against reading and copying confidential data. The controls should be verified and kept up-to-date.

- Storage conditions that are commensurate with the specification requirements for the media should be established. These include temperature, electric-magnetic fields including light, static electricity, humidity, dust, etc.

- The safety of physical storage should be established. This includes protection against theft, fire, smoke, and water.

- Protection against disasters should be established. This may mean storing multiple copies in several disparate locations, each of which will need to comply with the requirements above.

- A disaster recovery procedure should be implemented and verified. The procedure should include measures that mitigate risks from system/media failure, failures by personnel, and external factors, such as fire or defaulting suppliers.

10.8 Archiving Approaches Summary

Table 10.1: Digital Preservation Techniques Summary

Technique	Basic Characteristics	Main Strengths	Main Weaknesses
Increase on-line storage capacity	Retain aging data on existing platforms and applications by adding disk capacity.	Integrity, content, and meaning of data unaffected. Relatively inexpensive as long as system performance is not impacted.	Can impair system performance. Backup times and backup media volumes increased. Does not provide for easy separation between live and archived data and potentially does not comply with regulatory demands.
Mothball/time capsule/ computer museum	Maintain obsolete platforms and applications.	Integrity, context, and meaning of data unaffected.	Platform maintenance costs can be high. Limited life approach. Difficult to retain relevant people skills. Platform can fail with difficult disaster recovery.
Format migration	Data converted into new format and possibly structure to offset future upgrades or obsolescence.	Avoids obsolescence of platform and application. Data can still be processed.	Risk of data loss or corruption. Validation complex. Difficult to retain meaning and context. Multiple conversions may be required. Not all data can be converted.
Archive to standard formats	Conversion of data to industry standard format such as XML, rather than migrating to a new application.	Long-term accessibility to data improved. Enables wide choice of commercial EDA solutions. XML is in human readable form. Generally not controlled by the technology supplier.	Standards become obsolete. Not always possible to process data. Validation time consuming.

Table 10.1: Digital Preservation Techniques Summary (continued)

Technique	Basic Characteristics	Main Strengths	Main Weaknesses
Emulation and virtual computers	Preservation of original data and application through the operation of a simulated computer environment that sits on the latest operating system.	Context and meaning of data perfectly retained. Operating interface remains the same. Latest operating platform can be used.	Relatively immature technique and subject to on-going research and development. Difficult to validate. Dependent on external emulator development.
Transfer to non-electronic media	Migrates data to usually human readable formats on media such as paper or microfiche.	Low risk from obsolescence. Media has long life expectancy if maintained well. Data protected from changes.	Potentially high cost of storage (requires space). Takes longer to retrieve and can be more difficult. Can be difficult to delete individual records at end of life. Data can no longer be easily processed. Maintaining accurate copies at disparate geographical locations is difficult and expensive. Effectively impossible to provide remote and concurrent access.

Appendix G
Archiving Strategy Template

11 Appendix G: Archiving Strategy Template

This Section provides a template for the Archiving Strategy document. It highlights the relevant points that need to be addressed in each section of the document. For more background on the various aspects of the Archiving Strategy document, see Section 3 of this Guide.

11.1 Introduction

The purpose of the introductory Section is to summarize the content of the Archive Strategy, and to help the uninitiated reader to quickly determine the applicability of the document, without having to read the entire document. This Section also can be used to 'set the scene' and provide additional information that puts the strategy into context, e.g., by referring to applicable company and external standards that have formed the basis of or influenced the strategy, and how the strategy fits into future developments.

The authority under which the Archiving Strategy has been developed should be clearly stated. This authority will depend on the detailed scope and will determine who should review and approve the Archive Strategy.

11.2 Purpose

The purpose of the archiving strategy can be summarized as:

The purpose of the Archiving Strategy document is to guide the implementation and maintenance of Electronic Data Archival solutions that will ensure that any future changes will not have a material adverse impact on the compliant state of the archived records.

The purpose is achieved by identifying and defining the boundaries and key requirements of electronic data archiving so that these can be carefully considered and challenged. If they are seen as fit for purpose, they should be approved in accordance with End User procedures.

It is recommended that the archiving strategy is approved by QA and Management as a minimum with an IS/IT approval being highly desirable.

The Archiving Strategy document is not a detailed requirements specification. This should be considered when identifying the various boundaries and requirements, i.e., only strategically important issues and constraints need to be detailed. In realizing the strategy for a particular EDA, a Requirements Specification is envisaged (see Section 3 of this Guide).

11.3 Scope

Requirement: identify the key dimensions that define the boundaries for the Archiving Strategy document.

The scope of the Archiving Strategy should be clearly defined. Table 11.1 includes boundaries to consider.

Table 11.1: Scope Definition of Archiving Strategy

Property: Dimension	Example
Physical Location	Country, site, building, etc.
Organizational Unit	Company, group, division, department, etc.
Data Source	System, equipment, production unit, etc.
Data Owner	Functional unit, organizational layer, role or position, etc.
Data Type	Text, rich text, relational database, laboratory data, manufacturing data, etc.
Data Content	Regulated, engineering, commercial, etc.
Regulatory Requirement	GLP, GCP, GMP, GDP, EU regulated, US regulated, non-regulated, etc.

It is not suggested that all the above dimensions will apply in all circumstances; in fact, this is unlikely to be the case. But these dimensions may assist in identifying a well-defined boundary of the Archiving Strategy document. By defining the scope of the Archive Strategy, clarity is gained and unnecessary scope creep may be avoided.

11.4 Roles and Responsibilities

Requirement: identify the key archiving activities and assign who is responsible for each of these.

The various roles and responsibilities should be clearly defined. Archiving is likely to cross several operational and organizational boundaries. Given the criticality of the archiving function, having clear demarcations and definitions are key factors in assisting to achieve compliant operations.

The different roles will vary from one organization to the other, but Table 11.2 may be of assistance when defining the various responsibilities.

Table 11.2: Roles and Responsibilities

Property: Role	Sample Definition
Application Data Owner	The initial owner of the data before it is archived.
Archive Data Owner	The owner of the data once it has been archived.
Archive Administrator	The owner of the archiving operations on a daily basis.
Archive Owner	The owner of the archive system.
Technology Owner	The owner of the IS/IT infrastructure and platform for the archive.
Quality	Although in one sense everyone is responsible for the compliance and quality issues and adherence to these, Quality is overall responsible for assuring Management that procedures are adhered to. This role may be split into Regulatory, Quality, and Validation.
Management	Ultimately responsible to the regulator and authorities for the compliant operation of the EDA, including where third parties are being used. Financially responsible to the company owner.
Sponsor	The person or organization that commissions work from third parties. Responsible for work performed by such third parties.

For a global Archive Strategy, Table 11.2 may be sufficient in defining the key responsibilities. Where the strategy is applied only for a single individual archive, Table 11.2 is likely to be too imprecise, and a more detailed mapping of the archive operations against defined roles may be required.

As in all these matters, having individuals assigned to defined roles will ensure that the responsibilities are more clearly understood and acted upon. It is recommended that an individual's name is not used in the Archiving Strategy document, but where necessary, it is recommended that their job title is stated as well, in case their role changes.

Organizations tend to change fairly frequently, and the Archiving Strategy document should be updated on a regular basis to reflect the current organizational structure. It is recommended that the Archiving Strategy document is reviewed as frequently as SOPs, typically every two years. Where an organization is undergoing a major reorganization, it is recommended that the Archiving Strategy document is updated as soon as practicable after the reorganization.

11.5 Archive Content Requirements

This Section defines the requirements from an archive content perspective and focuses on compliance with operational demands for the content of the archive in a regulated environment.

11.5.1 Archive Processes

Requirement: identify and define the key archiving processes that are covered by the Archiving Strategy document. These should align with the Roles and Responsibility functions that have already been defined.

This Section should define the main processes that apply within the scope of the Archiving Strategy document. This may be achieved using a dataflow diagram that depicts the flow of archive data from the source system until eventual data deletion or export to another repository. Provided this diagram is kept reasonably simple, it may be used to help define the roles and responsibilities by overlaying the various spheres or interfaces of responsibilities. As an alternative to the diagram, a table may be used that defines the processes and the responsible person(s). An example diagram and table are shown in Section 2.6 of this Guide.

Provided this type of diagram or table is successfully created, it may be possible to combine this Section with the one defining roles and responsibilities although it is difficult to depict all activities in a single diagram format. There is a trade-off between increased detail definitions and potential loss of clarity of data and vice versa. Since this is a strategy document, definitions may be kept at a relatively high level, and further details put into, e.g., the relevant SOPs for the archive.

11.5.2 Data Definitions

Requirement: identify supported data and any limitations in use.

For each supported data type it is recommended that any limitations in use or considerations are specified. For example, not all data types may be supported to the same degree. The data types can be detailed as per the example shown in Table 11.3.

Table 11.3: Data Types (Example)

Property: Data Type	Examples
General file formats (ASCII)	CSV, XML, etc.
Written word/text (e.g., reports)	CSV and Word 97 onward are supported.
Rich media	VHS video, voice are supported.
Database	Access 2002, SQL2000 supported without restrictions. Other databases should be evaluated before use.
Laboratory data	The following proprietary laboratory data formats are supported without restrictions: (add list).
Manufacturing data	The following proprietary manufacturing data formats are supported without restrictions: (add list).

A similar table for the data content may look like this (example only):

Table 11.4: Data Content (Example)

Property: Data Content	Examples
GLP	Method file, configuration file, result file.
GCP	Consent form, dosage instruction, statistical analysis log file.
GMP	Batch record, recipe, cleaning record.
GDP	Distribution record.
Quality	Quality system, training record, competence record.
Engineering	Equipment history file, P&ID, ELD, design calculations, plant maintenance.
Business	Price list, client profile, human resource data.

The supported data definitions (type and content) should be verified, e.g., through manufacturer's documentation or by validation.

Table 11.3 and Table 11.4 are likely to be much more complex than have been shown. For example, a single laboratory instrument may be able to generate a number of data files, and it is unlikely that all of these will be able to be archived. This may be a combination of inability of the instrument to export the required file(s), and the EDA not being able to accept, archive, and interpret the data file(s).

To overcome this, it is suggested that for each entry, a sub-table is created, where applicable. The sub-table would, for each key data source, specify the supported file formats. Table 11.5 is an example table which illustrates this approach:

Table 11.5: Laboratory Data (Example)

Property: Laboratory Data	Examples
Result file	Millennium 32, AAnalyst 400.
Method file	Millennium 32, AAnalyst 400.
Configuration file	Millennium 32, AAnalyst 400.
Etc.	Etc.

These sub-tables are not intended to supplant detail design documents and are entirely optional. It could be argued that they provide too much detail to be suitable for inclusion in a strategy document, and this is particularly true for a company-wide document. Conversely, the sub-tables do clarify any short-comings in a chosen Archiving Strategy by high-lighting restrictions and limitations in archived data. They may be appropriate for inclusion in the Archiving Strategy document where this is specific to one EDA implementation. It is important that this type of detail information is defined, irrespective of where it is recorded.

11.5.3 Key Metadata

Requirement: identify the key metadata that must be associated with archive data.

There are some key attributes that need to be stored with the archived data. Required metadata may differ for the various types of data and content, but addressing the key attributes should ensure a minimum and consistent level of context information. Table 11.6 provides examples of key metadata.

Table 11.6: Key Metadata

Property: Key Metadata	Examples
Data Identity	Unique identifier of the data
Data Type	CSV
Data Content	Result file
Data Source	Instrument, system, department, etc. from which the data originates from
Data Context	Reason for content, status, relationships to other data
Data Owner	Application Data Owner and/or Archive Data Owner
Access Profile	Who can search, read, and supplement the data
Retention Period	Period of required storage before the data can be deleted

Note that each metadata item would not necessarily be applied to every archived data file and that one set of metadata may apply to several archived data files.

In addition to the above metadata, the particular Archiving Strategy or solution may impose a need for additional metadata. This metadata also should be clearly defined.

11.5.4 Deletion of Records/Data

Requirement: identify the quality safeguards for the record deletion process.

The following should be defined:

- The circumstances under which data will be deleted.
- The mechanisms by which data will be deleted and whether they will be automatic or manual or a combination of the two.
- The description of the verification processes that will be used to confirm correct record deletion.
- Measures to be implemented to ensure that data is not deleted without the required prior review and authorization from the appropriate personnel, including the Data Owner, and in the case of GxP data, a QA representative.

11.5.5 Exit Strategy

Requirement: define an 'exit strategy' for all EDAs. This is the 'fall-back position' when a record comes to its end of technological support.

How the archived records will be accessed for the duration of their associated retention period in the event that the application used to archive the records and retrieve them in the early years may no longer be available in a supported form in the longer-term, should be defined.

This definition should capture the worst case scenario and include:

- Existing application used to archive and retrieve the records in the early years
- Proposed application used to ensure longevity of the retrieval process throughout the required retention period
- Any risks associated with the proposed exit strategy
- How the data and the associated metadata will be migrated from one application to the other

11.5.6 Data Migration

Requirement: define how data and the associated metadata will be migrated from one platform to another.

Where possible, the following should be defined:

- The data and metadata that will be migrated
- The reasons for migration
- The original format of the data and metadata
- The modified format of the data and metadata
- How the migration process will be validated to confirm its accuracy and repeatability
- Who will be involved in the migration process from a review, approval, and execution perspective?

11.6 System Requirements

11.6.1 Technology

Requirement: specify the broad technology platform that will underpin the EDA and support the archive requirements.

The technology platform to be used for the EDA should be defined, including such factors as:

- Operating system
- Architecture
- Operator interface
- Supplier solutions

11.6.2 Interfaces

Requirement: identify the key data and user interfaces to/from the EDA.

Interfaces that have not been captured already in Sections 11.5.1, 11.5.2, and 11.5.3 should be described. Examples include the IS/IT infrastructure, global security policies, and maintenance provisions.

The key interfaces are those that can have a direct material impact on the functioning of the EDA, and would require significant effort to address. For instance, if there is a corporate decision to change the IS/IT infrastructure, this change may not be totally supported by the implemented technology solution of the EDA with potentially significant implications.

Stating the key interfaces clearly in the Archiving Strategy document should ensure that when considering material changes to any of the entities that have a key interface with EDA, the requirements of the EDA are taken into account. To ensure this happens, assumes that the End User has a system in place that cross-references these interdependencies between entities. It is not enough to state the key interfaces in the Archiving Strategy document, but the EDA interface requirements and assumptions need to be referenced in the applicable documentation for the identified entities that interface with the EDA.

11.6.3 Location/Environment

Requirement: specify the key location/environmental requirements for the EDA.

Key location and environmental requirements are those that, if challenged or not met, would have a direct material impact on the EDA, and would require significant effort to address.

Examples include:

- Where the scope of the EDA is for a single country using only the local language. A change in corporate structure, e.g., merging with another company, may introduce a barrier to use through the introduction of another corporate language.
- Where the scope of the EDA is expanded to another site that is not supported by the defined IS/IT infrastructure.
- Where the computer equipment puts specific requirements on the air quality and freedom from EMI emitting sources with implications on the building design that house the computers.

11.6.4 Security

Requirement: specify the key security requirements for the EDA.

Specify the following aspects of system security:

- Physical security
- System logical security – access to the computer system to start the EDA applications
- Application security – user groups and their associated access to the various aspects of the EDA application functionality and records

11.6.5 Operation

Requirement: define the significant operational demands that have not already been covered elsewhere.

The purpose of this Section is to detail any operational demands that are of significance to the rest of the organization. The operation of the EDA is closely linked to the roles and responsibilities, user groups, security, the defined processes, and the interfaces. However, these sections do not capture all key operational demands. For example, the various management processes such as operational checks and routine maintenance have been omitted from Section 11.5.1.

This Section also may be used to state critical operational parameters, such as availability, operability, accessibility, performance, and where applicable, capacity. Operation should include processes that regularly monitor the integrity of the metadata index and inspect data files for damage. Both processes should be designed into the EDA, and may utilize digital signature technology.

It is likely that regulatory and quality procedures will impose operational demands, such as definition of signature events and recording of events in an audit trail. These may be captured here or in Section 11.7.2.

Where the Archiving Strategy document covers several areas and implementations of EDA, this Section may be used to streamline and harmonize operational demands, by defining a core set of required operations. The roles and responsibilities table, together with the main archiving processes table, may act as a starting point in considering what operations should be specified in this Section.

11.6.6 Testing Environment

Requirement: define the testing methodology for proving that the EDA is able to perform the required functionality, both for the initial implementation and potential future upgrades/modifications.

The EDA is similar to a business system employing a large database and the considerations for its testing and setting to work are similar. The testing environment should be such that verification of the EDA does not impact service continuity for users. This is most likely to be an issue for upgrades of an existing EDA. The use of test data and cut-over to live operation should be carefully planned, e.g., will the cut-over be end-to-end or parallel run?

Another consideration is how to handle future upgrades and the need to verify these. Will verification be carried out on the live system and will this affect service continuity? Or will a second off-line mirror system be used for implementing and verifying changes?

11.7 Compliance Requirements

Although one of the main purposes of the whole Archiving Strategy document is to meet compliance requirements, this Section deals with any specific regulatory, quality, and validation demands and should be supported by plans and master plans, as appropriate.

11.7.1 Risk Assessment

Requirement: define the methodology that will be applied to assess the risks associated with the implementation and use of the EDA. Reference any standards, procedures, etc., used.

The risk assessment methodology described should capture the following:

- Identification of the risk hazards
- Method used to determine risk priorities based on an assessment
- Accepted residual risks that will remain after any risk reducing measures
- Rationale for the reached conclusions with regard to residual risks

11.7.2 Regulatory Requirements

Requirement: specify the applicable predicate rules for the EDA. Identify predicate electronic records and signatures.

Capture the following:

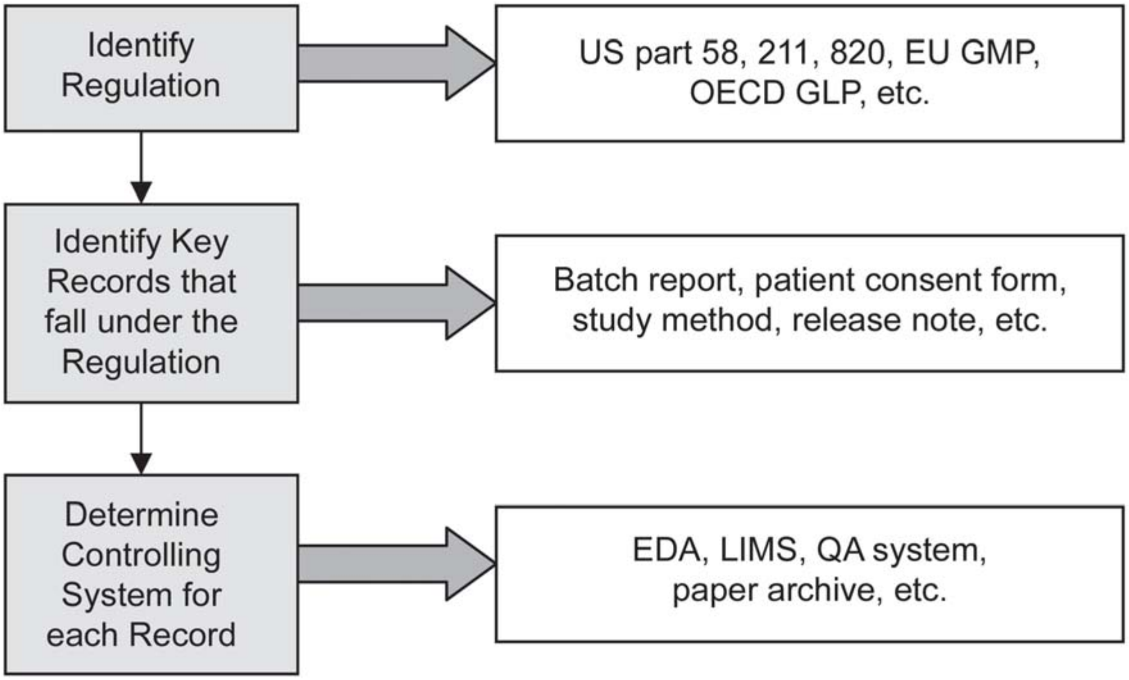
- The regulations that apply to the EDA should be stated.
- The key records that fall under the stated regulation should be identified.
- For each key record, it should be determined whether regulatory compliance is controlled by the EDA or another data repository.

The following factors should be considered:

- There may be more than single-country legislation, e.g., both USA and EU rules will apply to many End Users. Company interpretations of such regulation also should be referenced. It should be noted that such interpretations change over time and may have an implication on the operation of the EDA.
- The key records should already have been documented in the Data Definitions Section. Refer to Table 11.4.
- Regulatory requirements for archiving should be satisfied by data held in the EDA, but there may be copies of the same data held in other places for informational or management purposes. For example, a data warehouse may hold consolidated data derived from many different toxicology studies which assists the selection of new drug candidates for further development.

Figure 11.1 illustrates the suggested process for the identification of Regulatory Requirements.

Figure 11.1: Identifying Regulatory Records (Example)


Identify Regulation
US part 58, 211, 820, EU GMP, OECD GLP, etc.
Identify Key Records that fall under the Regulation
Batch report, patient consent form, study method, release note, etc.
Determine Controlling System for each Record
EDA, LIMS, QA system, paper archive, etc.

Further details of Regulatory Archiving Requirements can be found in Section 3.3.4.2 of this Guide.

11.7.3 Achieving and Maintaining Compliance

Requirement: state the model to be used for achieving and maintaining system compliance and fitness for intended use and reference procedures and guidance that apply.

The following aspects should be considered and documented within the Archiving Strategy document:

- Does the system need to be validated?
- What is the extent of validation required?
- What methodology will be followed?
- Will a standard approach be adopted?
- Where it is proposed that a guideline (such as GAMP) will be followed, which specific parts will be applied?

11.7.4 Validating the Archive Processes

Requirement: state the validation model to be used for validating the Archive Processes and reference procedures and guidance that apply.

Define how the archive process itself will be validated, covering such areas as:

- Interfaces to the EDA

- Users
- Source data
- Testing

11.7.5 Performance Verification

Requirement: define how performance of the EDA will be verified.

The following aspects of the performance verification should be defined:

- The aspects of the system functionality that will be tested
- How consistent performance of the EDA, when accessed by a large user base, will be confirmed
- Subsequent changes to and evolution of the EDA after the initial verification
- Data throughput and changes resulting from technological advances

11.7.6 Change Management and Business Continuity

Requirement: define the Change Management process that will be applied to the EDA and how Business Continuity will be assured.

The following aspects should be defined:

- How changes to the EDA will be identified, reviewed, approved, and managed through to completion, given that that the EDA will have a large user base that potentially spans many different operational and business areas.
- Those aspects of the EDA design and operation that will ensure Business Continuity in the event of a failure or disaster

11.7.7 SOPs

Requirement: define the SOPs that will be written for the EDA.

SOPs that will cover the following aspects of the EDA should be defined:

- Routine and non-routine archive processes
- System-related processes
- Administrative processes
- Quality, regulatory, and validation processes

11.7.8 Maintenance and Ongoing Evaluation

Requirement: define how the EDA will be maintained and evaluated.

The following aspects of maintenance should be defined:

- Service level agreement or similar where an external resource is employed
- A risk assessment of EDA functionality to establish required maintenance tasks, their frequency, and impact from failure
- Maintenance tasks, required record keeping, definition of acceptable results, and review and approval of results

The principles to be followed should be established, but there is not an expectation that this list is fully expanded in the Archiving Strategy document.

Appendix H
Good Practices

12 Appendix H: Good Practices

Good record management practice applies equally to electronic data archives as to its paper-based predecessor. It will help to reduce the size of the archives and preserve ownership details, links with metadata, and electronic signatures.

Good practices should consider and include:

1. Avoid scope creep when specifying an EDA. Examples include expanding the EDA into a general document management system or a knowledge-based system. Although these may be worthwhile developments, they may have a substantial cost impact on the archiving project and detract from the original project purpose.

2. When selecting the technology solution, consider the software license requirements, such as frequency and impact of necessary upgrades to prevent loss of software support.

3. Good archiving practice starts with good quality source data to facilitate archiving.

4. Archive only those records that are required by predicate rules and legislation, by business practices (including legal considerations), or by procedures. Determine whether there is a need to keep all electronic records or whether those that do not fall under these categories can be safely discarded. Decisions should be documented and justified.

5. Where the intention is to retain data, that data should not be defined as dead. There is no reason to archive dead data. Data should be retained for a reason.

6. Draft documents, where the document has been formally issued later, should not be archived. The archive is intended for long-term storage of key information. Draft documents are either superseded by issued versions of the same document or never issued. (Draft documents have the potential to obscure issued information.)

7. It may be important to understand why a document was never formally issued; the decision and reasons for this can be captured in a separate document that is archived.

8. It is not uncommon to find a record in several places. When archiving such a record, attempt to locate all copies of it, check they are compatible and archive only one record while deleting the others (following the due process for deletion). Add a link from the location of the deleted copies to the retained record.

9. When archiving a newer issue of a document, it is considered essential to record the reason for the update and what is being replaced, augmented, or deleted. This applies also to where a new document replaces one or several existing documents. Note that all previous versions would still be retained in the archive for auditing purposes.

10. Write an Orientation Memorandum. This could include the computer systems used and the processes followed, the data integrity history, explicit interpretations of the then applicable regulations, details of the persons who created the data and instigated the archiving, together with any information that might be useful for someone who, in five or 10 years' time, is trying to understand what happened. The persons who were involved with generating the original data may not be available to interpret the information.

11. A directory structure should be defined where metadata are saved in the same directory as the parent data. This will facilitate maintenance and linking of metadata with parent data.

12. Directory and filenames should be given careful consideration. Some archiving software truncates filenames. Other software changes file metadata, such as the timestamps when writing data to a CD-ROM.

13. The path name alone should not be relied upon for identifying the record. Path names may change, and are usually not sufficiently robust for record identification.

14. Avoid archiving compressed data files. These introduce another layer of conversion and the potential for data corruption, as well as bringing the issue of continued availability of the decompression tool and an operating system on which to run it. In some instances, it may not be possible to fully restore the original data, i.e., there is loss of resolution or metadata.

15. Beware loss of resolution when migrating images to formats such as JPEG; consider using TIFF instead.

16. Hyperlinks should be avoided where the URL could change without the knowledge of the Archivist.

17. Use procedures to preserve the integrity and identity of data when archiving from one set of media, such as files on hard drive to another such as CD-ROM.

18. Ensure that all storage media is uniquely identified and that all removable media is labeled with the media contents (with the system name, application name, file and folder names, versions, dates, etc.).

19. When designing an EDA, consider that the required speed of response for different system parts may not be the same. For example, a higher speed of response generally will be required for the Data Management functions, such as Search and Report, compared with the retrieval of a specific Archive Information Package.

20. Ensure that backup and disaster recovery procedures are appropriate to the system and media used. For example, the Data Management part of the EDA is likely to need more frequent backups than the archive storage, simply reflecting the amount of traffic and speed of response.

21. Archive data once the likelihood of accessing that record has reduced to a given frequency, say once a year. Alternatively, archive data at a well-defined point in the workflow, e.g., at the end of final approval of a GLP study or manufacturing batch.

22. Make sure that there is only one source of control to the archived data, using a small number of well-defined indices and procedures. There may be several archivists, but they should all be using the same methods and procedures.

23. Perform archiving and backup of data at pre-established intervals based on system and user requirements.

24. Sufficient funding should be allocated for the ongoing maintenance of the EDA.

25. Keep the Quality role separate from the Archivist role. Quality should be seen to be independent, particularly during an audit. Ensure that Quality is empowered to report directly to Management.

26. Collect relevant metrics from the operation of the archive so that its performance can be substantiated and monitored. The metrics also will enable future requirements to be better predicted and planned.

27. Verify that record integrity is being maintained, e.g., by conducting tests as part of scheduled maintenance or by regularly reviewing audit trails.

28. Make sure there is a test environment for testing software and hardware changes without the risk of corrupting data which has already been archived.

29. Handle operational difficulties through a formal event and problem reporting system.

30. Hold regular meetings with the key archive stakeholders to review archive procedures, performance, migration requirements, future requirements, etc.

31. IS/IT are important stakeholders in Archival and it is important to include IS/IT in any relevant communications and meetings. Ensure that IS/IT understands the archiving needs clearly so that the infrastructure is appropriate and can safely support the archive.

32. Consider to what level data should be deleted to become fully destructed. Many deletion processes simply remove the reference to the data without actually removing the data itself. Some data also may be retained in the audit trail.

33. Eventually EDA systems will need to be retired. An EDA that is allowed to outlive its economic life will become expensive to maintain and eventually non-maintainable. Retirement planning should be part of the design.

Appendix I

Glossary

13 Appendix I: Glossary

13.1 Definitions

These definitions have been taken from a variety of sources. Some of the definitions have been developed or modified in an effort to make them more understandable. Examples have been added as illustrations to the definitions. The list of definitions does not purport to be comprehensive, but hopefully relevant to an EDA. Some definitions apply to both the electronic and non-electronic archive, whereas others have been worded such that they only apply to the electronic archive.

Aggregate Data

Refers to cumulative information from multiple records.

Application Data Owner

The initial owner of the data before it is archived.

Appraisal

Archivists' term. The purpose of archival appraisal is to decide which records should be preserved in the long term. As the term appraisal indicates, it should reveal the value of the records for future purposes and retention decisions should be based on that value.

Archival (adjective)

The word archival is the adjective derived from archive and describes something that relates to an archive or the function of archiving, or from an archivist's perspective (e.g., archival metadata). Increasingly the word is being used on its own to refer to the activity or function of archiving.

Archival Function

The International Council on Archives' "Guide for Managing Electronic Records from an Archival Perspective" (1997) and its 2005 workbook for archivists on electronic records define the concept of archival function as follows: "The archival function is that group of related activities contributing to and necessary for accomplishing the goals of safeguarding and preserving archival [electronic] records, and ensuring that such records are accessible and understandable."

Archival Information Package

An OAIS term and abbreviated to AIP. The Information Package that is required to safely keep the record for long-term storage. The AIP is used for moving records to/from the archival storage. An AIP is constructed from one or several Submission Information Packages. One or several AIPs are used to construct the Dissemination Information Package.

Archive (noun)

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

The archiving process must retain the meaning of the data and the data within the archive must be searchable and recoverable. For archived electronic records to remain accessible and understandable over time, they may need to undergo some preservation actions over their life time.

OAIS defines archive as an organization that intends to preserve information for access and use by a designated community.

Archive (verb)

The act of placing an object into an archive.

Archive Administrator

The owner of the archiving operations on a daily basis.

Archive Data Owner

The owner of the data once it has been archived.

Archive Owner

The owner of the archive system.

Archiving

Archiving is the activity of running an archive.

Enable reliable, authentic, meaningful, and accessible records to be carried forward through time within and beyond organizational boundaries for as long as they are needed for the multiple purposes they serve. (Professor Sue McKemmish, Monash University, Australia)

Authenticity

In the context of electronic records, the quality of actually being what they purport to be. It implies that provenance can be established and the integrity of the original record is preserved.

Backup

Keeping a safe copy of information (usually done on a regular basis) so that it can be recovered in the event of loss due to human error, hardware or software failure, data corruption, theft, sabotage, or natural disaster. Backup copies need to be stored and maintained with the same degree of care as live data, as this is what the data becomes when it is restored from the backup copy.

Backup copies are often employed for short-term recovery with an archive used for long-term storage. However, an electronic archive also should have a backup. Another distinction between backup and archive is that the backup will contain copies of master records, whereas the archive is deemed to contain the master records.

Data

Webopedia (www.webopedia.com) defines data as: Distinct piece of information, the plural form of datum, but often used as both the singular and plural form of the word. In software terms, data is distinct from programs that are used for manipulating data.

Digital Signature

An electronic signature based upon cryptographic methods of originator authentication, computed by using a set of rules and a set of parameters such that the identity of the signer and the integrity of the data can be verified.

Dissemination Information Package

An OAIS term and abbreviated DIP: an Information Package that is delivered by the archive to a Consumer, i.e., someone who requests and receives information from an archive. One or several Archival Information Packages construct a DIP.

Disposition

An archivists' term: how records are handled once the retention period has been meet. Disposition usually refers to deletion, but can include transfer to another agency or archive.

Dynamic Data/Records

Live data/records that reside within an on-line data repository such as a database. The opposite to static data/records.

Electronic Record

As defined by the FDA: any combination of text, graphics, data, audio, pictorial, or other information representation in digital form that is created, modified, maintained, archived, retrieved, or distributed by a computer system.

Electronic Signature

As defined by the FDA: a computer data compilation of any symbol or series of symbols executed, adopted, or authorized by an individual to be the legally binding equivalent of the individual's handwritten signature. As well as being used in place of a handwritten signature, an electronic signature also may be used to authenticate a signed document as still having the same content as when it was signed.

Emulation

A process by which a computer imitates the actions of another computer so that the imitating system accepts the same data and executes the same computer programs as the imitated system.

Extensible Mark-up Language (XML)

Extensible Mark-up Language (XML) is a simple, very flexible text format derived from SGML and standardized (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML also is playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. See http://www.w3.org/XML/

Hybrid System/Record

A hybrid system is one that uses both non-electronic (e.g., paper or microfiche) and digital/electronic output media. Similarly, a hybrid record is a record comprising at least two components stored on different media, typically electronic and paper. An example of a hybrid record would be an electronic record that is printed and approved on paper with a handwritten wet ink signature.

Hyper Text Mark-up Language (HTML)

HTML is the common language for publishing hypertext on the World Wide Web. It is a non-proprietary format based upon SGML and can be created and processed by a wide range of tools. HTML uses tags to structure text into headings, paragraphs, lists, hypertext links, etc. See http://www.w3.org/MarkUp/

Information Lifecycle Management

Information Lifecycle Management (ILM) is a strategy for aligning an organization's IT infrastructure with the needs of the business, based on the changing value of the stored information. Through ILM, an organization achieves the most value from the information, at the lowest cost, at every point in its lifecycle.

Ingest

An OAIS term. The services and functions that accept Submission Information Packages from Producers, prepares Archival Information Packages for storage and ensures that these and their supporting Descriptive Information (DI) become established within an archive.

Information Package

An OAIS term. The content information and associated preservation description information, package information, and package description (metadata) which is needed to aid in the preservation of the content information. Information packages can be of the type Submission, Archival, and Dissemination.

Management

Ultimately responsible to the regulator and authorities for the compliant operation of the EDA, including where third parties are being used. Financially responsible to the company owner.

Metadata

ISO-15489 defines metadata as: Data describing context, content and structure of records, and their management through time.

Metadata also can be described as data associated with records required to enhance/interpret or support the record of interest to facilitate its retrieval. Or simplified: Data about data. Digital metadata normally occurs in sets of attributes that are standardized for a particular purpose so they can be searched, aggregated, etc.

Metadata is necessary to ensure a meaningful presentation or interpretation of the electronic record, or for the reconstruction of the electronic record.

Metadata comes in many forms, including the label on a disc or tape, filename, date/time stamp, sample ID and data structure/relationship. Another example of metadata is a data dictionary file which documents a relational database structure, attributes, relations, key fields, format protection, ownership, etc.

Migration

The transfer of digital information from one hardware/software configuration to another or from one generation of computer technology to a subsequent generation. The purpose of migration is to preserve the integrity of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology.

For convenience, migration can be sub-divided into the categories format migration (conversion), system migration, and media migration.

Open Archival Information System

A platform-independent model developed by the Consultative Committee for Space Data Systems (CCSDS). Its aim is to define a common framework of terms and concepts to promote future technical developments in electronic data archiving. Now adopted as published standard ISO 14721:2003.

Persistent Identifier (PI)

A maintained reliable pointer to the identity, and also possibly its location, of a digital object through time.

There are many types of PIs: The ISBN for physical books, and in some cases also electronic publications, provides a unique name, but not the location of the item. In the digital world the URI/URN identifies the resource, while the URL also locates the resource. The term Digital Object Identifier (DOI) is also used.

Personal Data

Any information relating to an identified or identifiable natural person, the data subject. An identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural, or social identity.

Predicate Rules

Within the context of 21 CFR Part 11: Requirements set forth in the Federal Food Drug and Cosmetic Act (the Act), the Public Health Service Act (PHS Act), or any other FDA regulation with the exception of 21 CFR Part 11.

Preservation

The act of maintaining information in a correct and independently understandable form over the long term. (OAIS definition). An archive is likely to require preservation actions.

Quality

Although in one sense everyone is responsible for the compliance and quality issues and adherence to these, Quality is overall responsible for assuring Management that procedures are adhered to. This role may be split into Regulatory, Quality, and Validation.

Raw Data

Any worksheets, records, memoranda, notes, or exact copies thereof that are the result of original observations, measurements, recordings, etc. of an activity, such as a study, operation, investigation, etc., and are necessary for the reconstruction and evaluation of the report of that activity.

In the event that exact transcripts of raw data have been prepared (e.g., tapes which have been transcribed verbatim, dated, and verified accurate by signature), the exact copy or exact transcript may be substituted for the original source as raw data. Raw data may include photographs, microfilm or microfiche copies, computer printouts, magnetic media, dictated observations, and recorded data from automated instruments. As such, the raw data may exist in either hard/paper copy or electronic format.

Record

In its "Guide for Managing Electronic Records from an Archival Perspective" (1997) and its 2005 workbook for archivists on electronic records, the International Council on Archives defines a record as "recorded information produced or received in the initiation, conduct or completion of an institutional or individual activity, and that comprises content, context and structure sufficient to provide evidence of the activity." Increasingly, electronic records can be distributed, compound objects.

Record Lifecycle

The stages through which a record passes during its life. Typically, these are: Creation, Active Use, Semi-Active/ Inactive Use, and Final Disposition (such as deletion).

Retention Period

The time period that a record is required to be kept available for ready inspection by a regulatory agency.

Retention Schedule

A set of instructions allocated to a class of records to determine the length of time for which they should be retained by the organization for business purposes, and the eventual fate of the records on completion of this period of time.

Sponsor

The person or organization that commissions work from third parties. Responsible for work performed by such third parties.

Standard Generalized Mark-up Language (SGML)

This is a standard for how to specify a document mark-up language or tag set. SGML is not in itself a language, but a description of how to specify one. It is a meta language. HTML and XML are examples of SGML-based languages.

Static Data/Records

Data/records, including metadata, that are not subject to change. To carry out a change of static data/records often entails code changes that can only be carried out by a privileged user, and normally only in a development/test environment as opposed to a live environment. The opposite to dynamic data/records.

Storage (noun)

The recording of an electronic record on a given medium, e.g., magnetic tape, CD, hard disk.

Storage (verb)

The process of managing media upon which information is stored.

Submission Agreement

An OAIS term. The agreement reached between an Archive and the Data Producer that specifies a data model for data submission. This data model identifies format/contents and the logical constructs used by the Data Producer and how they are represented on each media delivery or in a telecommunication session.

Submission Information Package

An OAIS term and abbreviated to SIP. An Information Package that is delivered by the Producer to an archive for use in the construction of one or several Archival Information Packages.

Technology Owner

The owner of the IS/IT infrastructure and platform for the archive.

Trustworthy Record

For a record to remain reliable, authentic, with its integrity maintained, and useable for as long as the record is needed, it is necessary to preserve its content, context, meaning, and sometimes its structure.

A trustworthy record preserves the actual content of the record itself and information about the record that relates to the context in which it was created and used. Specific contextual information will vary depending upon the business, legal, and regulatory requirements of the business activity. The record's reliability and authenticity may be impaired if its structure or arrangement are not preserved. The acronym ALCOA is used to summaries the key properties of a trustworthy record.

13.2 Acronyms and Abbreviations

ACID	Atomicity, Consistency, Isolation, and Durability
AIP	Archival Information Package
ALCOA	Attributable, Legible, Contemporaneous, Origin, and Accurate
AnDI	Analytical Data Interchange
AnIML	Analytical Information Mark-up Language
API	Active Pharmaceutical Ingredient
ASCII	American Standard Code for Information Interchange
ASTM	American Society of Testing and Materials

CAMiLEON	Creative Archiving at Michigan and Leeds Emulating the Old on the New (United Kingdom and United States of America)
CCSDS	Consultative Committee for Space Data Systems Organization
CD	Compact Disk
CFR	Code of Federal Regulations (United States of America)
CSV	Comma-Separated Variable
DCMI	Dublin Core Metadata Initiative
DI	Descriptive Information
DIP	Dissemination Information Package
DLM	Données Lisibles par Machines – Fr. (Machine-readable data) (European Union)
DoH	Department of Health (United Kingdom)
DOI	Digital Object Identifier
DOS	Disk Operating System
DTD	Document Type Definition
DVD	Digital Versatile Disk
EC	European Commission
EDA	Electronic Data Archive
EDMS	Electronic Document Management System
ELD	Engineering Line Diagram
EMEA	European Medicines Agency, previously known as European Agency for Evaluation of Medical Products or European Medicines Evaluation Agency
EMI	Electro-Magnetic Interference
EPA	Environmental Protection Agency (United States of America)
ER/ES	Electronic Records/Electronic Signatures
ERMS	Electronic Records Management System
ERP	Enterprise Resource Planning
ETL	Extraction, Transformation and Loading

EU	European Union
EUCOMED	European Medical Technology Industry Association
FDA	Food and Drug Administration (United States of America)
FS	Functional Specification
FTP	File Transfer Protocol
GAMP	Good Automated Manufacturing Practice (ISPE)
GC	Gas Chromatography
GCP	Good Clinical Practice
GDP	Good Distribution Practice
GERM	Good Electronic Record Management
GLP	Good Laboratory Practice
GMP	Good Manufacturing Practice
GxP	Collective name for GLP, GCP, GMP, and/or GDP
HIPAA	The Health Insurance Portability and Accountability Act (United States of America)
HP	Hewlett Packard
HPLC	High Performance Liquid Chromatography
HSM	Hierarchical Storage Management
HTML	Hyper Text Mark-up Language
IAAC	Information Assurance Advisory Council
ICH	International Conference on Harmonization
ICSTI	International Council for Scientific and Technical Information
ICU	Intensive Care Unit
ID	Identity
IEC	Independent Ethics Committee
IEC	International Electrotechnical Commission
ILM	Information Lifecycle Management

InterPARES	International Research on Permanent Authentic Records in Electronic Systems
IP	Internet Protocol
IR	Infra-Red
IRB	Institutional Review Board
IS	Information Systems
ISBN	International Standard Book Number
ISO	International Standards Organization
ISPE	International Society for Pharmaceutical Engineering
IT	Information Technology
JCAMP-DX	Joint Committee on Atomic and Molecular Physics – Data Exchange
LC	Liquid Chromatography
LIMS	Laboratory Information Management System
LRD	Laboratory Raw Data
MARC	Machine-Readable Cataloguing
MHRA	Medicines and Healthcare products Regulatory Agency (United Kingdom)
MS	Mass Spectrometry or Mass Spectrometer
MS	Microsoft
NARA	National Archives and Records Administration (United States of America)
NIST	National Institute of Standards and Technology (United States of America)
NMR	Nuclear Magnetic Resonance
OAIS	Open Archival Information System
OASIS	Organization for the Advancement of Structured Information Standards
OECD	Organization for Economic Cooperation and Development
OLAP	On-Line Analytical Processing
OLTP	On-Line Transaction Processing
OQ	Operation Qualification

OS	Operating System
PC	Personal Computer
PDF	Portable Document Format
PI	Persistent Identifier
PIC/S	Pharmaceutical Inspection Convention and Pharmaceutical Inspection Cooperation Scheme
P&ID	Piping and Instrumentation Diagram or Process and Instrumentation Diagram
PQ	Performance Qualification
QA	Quality Assurance
QC	Quality Control
R&D	Research and Development
ROM	Read Only Memory
RS	Requirements Specification
SDMS	Scientific Data Management System
SGML	Standard Generalized Mark-up Language
SIG	Special Interest Group
SIP	Submission Information Package
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms (College of American Pathologists)
SOP	Standard Operating Procedure
SOX	Sarbanes-Oxley Act of 2002 (United States of America)
SQL	Structured Query Language
TR	Technical Report
URI	Uniform Resource identifier
URL	Uniform Resource Locator or Universal Resource Locator
URN	Uniform Resource Name
UV	Ultra-Violet
UVC	Universal Virtual Computer (IBM)

VERS	Victorian Electronic Records Strategy (Australia)
VHS	Video Home System
WORM	Write Once Read Many
WWW	World-Wide Web
XML	Extensible Mark-up Language

Appendix J
References

14 Appendix J: References

This Section contains a list of source material. The SIG has made use of some of this information during its work on the Guide. The list is not exhaustive, as the number of relevant publications is huge and continues to grow, reflecting the importance and currency of this topic. For each reference, a Web link is given and a brief comment is made on its applicability to the subject of EDA and this Guide.

Although all given links were working at the time of publication, no guarantee can be made that they will continue to do so. However, it should be possible to locate the referenced organizations and documents.

1. GAMP Guide for Validation of Automated Systems; Good Automated Manufacturing Practice GAMP 4, December 2001. Published by ISPE, see www.ispe.org. This is the most widely used guide for validation of computer systems, and the model presented in the GAMP Guide can generally be applied to an EDA.

2. Good Practice and Compliance for Electronic Records and Signatures, Part 1 – Good Electronic Records Management (GERM), September 2002. Published by ISPE and PDA, see www.ispe.org and www.pda.org. This contains default information of what constitutes GERM.

3. GAMP Good Practice Guide: A Risk-Based Approach to Compliant Electronic Records and Signatures, February 2005. Published by ISPE, see www.ispe.org. This guide introduces a risk-based concept for addressing compliance issues related to the use of electronic records and signatures.

4. ISO 19005-1 Document Management - Electronic Document File Format for Long-term Preservation - Part 1: Use of PDF 1.4 (PDF/A-1), October 2005. Published by ISO, see www.iso.org. This is the platform independent standard of the archive version of Adobe Acrobat PDF.

5. ISO 15489 Information and Documentation – Records Management. Published by ISO, see www.iso.org. This standard comes in two parts: Part 1 covers the general principles of best records management practices, and part 2 is an implementation guide, including methodology, overview of processes, and bibliography.

6. ISO/TR 15801 Electronic Imaging – Information Stored Electronically – Recommendations for the Trustworthiness and Reliability, 2004. Published by ISO, see www.iso.org. This technical report is an extension of the ISO 15489 standard, and contains recommendations for information management systems with regard to trustworthiness, reliability, authenticity, and integrity of stored information. Although the report deals with specific items such as scanning of images and image processing, much of it is generally applicable to an EDA.

7. CCSDS 650.0-B-1 Reference Model for an Open Archival Information System (OAIS) Blue Book, January 2002. Published by CCSDS, see www.ccsds.org. This adopted standard provides a platform independent model for an EDA. The model is described in Section 6.

8. The World Wide Web Consortium can be accessed at www.w3.org. This international organization develops guidance and recommendations for web standards that are non-proprietary. The Web site is a useful source of Web-related information.

9. Risk Management of Digital Information: A File Format Investigation, June 2000. Published by the Council of Library and Information Resources, see www.clir.org/pubs/reports/pub93/contents.html. This is a learned study conducted at Cornell University of in particular file formats and migration issues, and is a useful reference work on these subjects. The Council of Library and Information Resources have published a number of works on strategies and tools for the digital library, see www.clir.org/pubs/reports/strategies.html.

10. The Creative Archiving at Michigan and Leeds Emulating the Old on the New (CAMiLEON) project is a joint UK/USA university undertaking for the development and evaluation of a range of technical strategies for the long term preservation of digital materials. CAMiLEON can be found at www.si.umich.edu/CAMILEON. CAMiLEON has published reports on software longevity, emulation, and migration.

11. DLM Forum is an EU body for the development of guidance on all aspects associated with archiving of digital material. Although focused on public administrations and national archives, the forum contains representatives from across industry and the research community. In French, DLM stands for Données Lisibles par Machines (or Machine-readable data). The forum can be found at: http://ec.europa.eu/transparency/archival_policy/index_en.htm. In particular, the Guidelines on Best Practices for Using Electronic Information, published 1997, can be found at: http://europa.eu.int/ISPO/dlm/documents/guidelines.html.

12. The Long-term Preservation of Authentic Electronic Records, 2004. Published by the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project, see www.interpares.org/book/index.htm. InterPARES is a knowledge-based organization for addressing the concerns in long-term preservation of electronically stored records. The output from the InterPARES 1 research project was published in the above referenced document, which addresses the conceptual requirements for authenticity, methods of selection and preservation, and an intellectual framework for the development of policies and strategies. The InterPARES 2 research project aims to develop theory and methods capable of ensuring the reliability, accuracy, and authenticity of electronic records.

13. Records Management Guidance for Agencies Implementing Electronic Signatures Technologies, October 2000. Published by National Archives and Records Administration (NARA) USA, see www.archives.gov/records-mgmt/policy/electronic-signature-technology.html. This fairly brief document also covers the record lifecycle, what constitutes a trustworthy record and key records management issues.

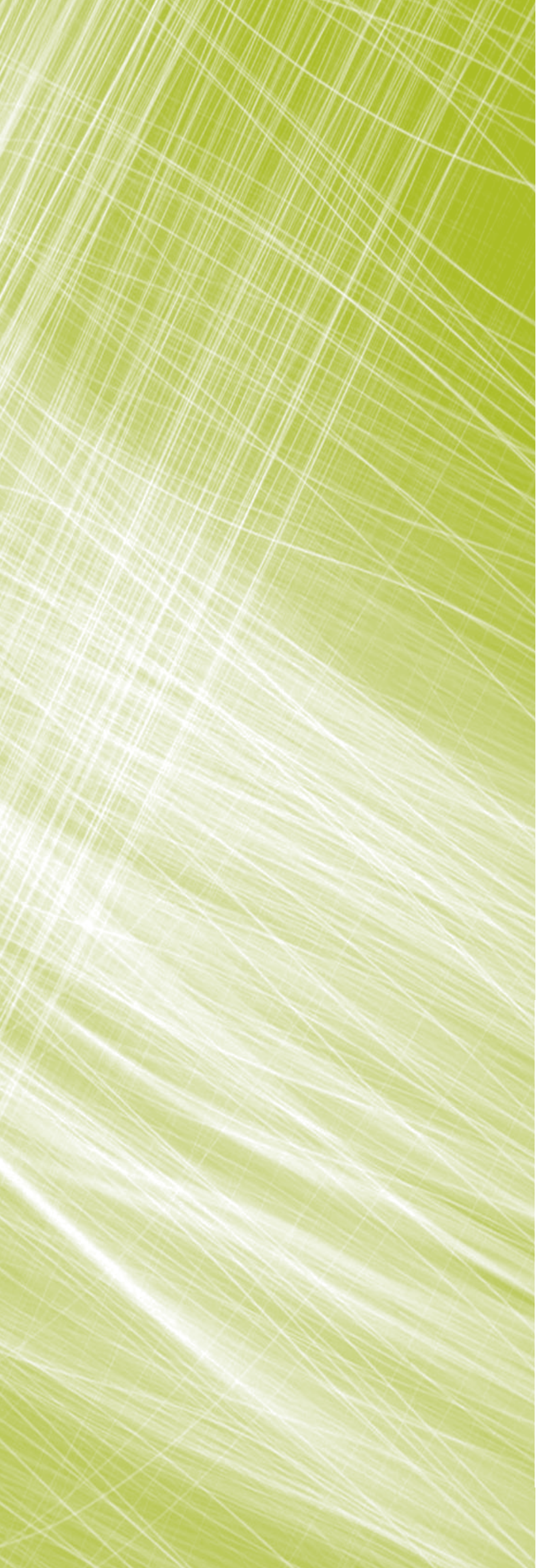
14. Requirements for Electronic Records Management Systems (ERMS), 2002. Published by Public Records Office, The National Archives, U.K. see www.nationalarchives.gov.uk/electronicrecords/function.htm. This is a comprehensive generic requirements specification. The Public Records Office has other useful information, such as Management, Appraisal and Preservation of Electronic Records, 1999, see www.nationalarchives.gov.uk/electronicrecords/advice/guidelines.htm.

15. Digital Electronic Archiving: The state of the Art and the State of the Practice, April 1999. Published by International Council for Scientific and Technical Information (ICSTI), see www.icsti.org. This comprehensive approach adopts an information lifecycle approach. The study has been summarized in an article in D-Lib magazine, see www.dlib.org/dlib/january00/01hodge.html.

16. Victorian Electronic Records Strategy (VERS) – Final Report, 1998. Published by the Public Record Office Victoria, see www.prov.vic.gov.au/vers/pdf/final.pdf. This is a comprehensive report on a strategy based on standardization of electronic record format using XML or PDF that encapsulates content, context, and authenticity. This standardization extends to the metadata. The VERS site at www.prov.vic.gov.au/vers contains a wealth of useful information related to EDA.

17. Directors and Corporate Advisors' Guide to Digital Investigations and Evidence, September 2005. Published by Information Association Advisory Council (IAAC) U.K., see www.iaac.org.uk. IAAC is a cross-industry society that includes government bodies, industry, and research-led organizations for the promotion of a secure information society. Although primarily U.K. based, IAAC has members from several other countries. The referenced report discusses the risks and issues involved in providing and retaining electronic evidence that can be used in a court of law. It highlights that the financial penalties for not addressing these risks and issues can be extremely severe. The Web site is a useful source for security issues and their implications on complying with the law.

18. Refer also to the list of regulatory references detailed in Appendix A of this Guide.



ENGINEERING
PHARMACEUTICAL
INNOVATION

ISPE Headquarters
3109 W. Dr. Martin Luther King Jr. Blvd., Suite 250
Tampa, Florida 33607 USA
Tel: +1-813-960-2105, Fax: +1-813-264-2816

ISPE Asia Pacific Office
73 Bukit Timah Road, #04-01 Rex House, Singapore 229832
Tel: +65-6496-5502, Fax: +65-6336-6449

ISPE China Office
Suite 2302, Wise Logic International Center
No. 66 North Shan Xi Road, Shanghai, China 200041
Tel +86-21-5116-0265, Fax +86-21-5116-0260

ISPE European Office
Avenue de Tervueren, 300, B-1150 Brussels, Belgium
Tel: +32-2-743-4422, Fax: +32-2-743-1550

www.ISPE.org