Long-Term Archiving of Analytical Instrument Data — Technical Guidebook

Version 1.1

February 15, 2023

Public Interest Incorporated Association
Japan Image and Information Management Association

R&D Data Archiving Committee

Table of Contents

1. Purpose

The purpose of this Technical Guidebook is to describe a technique that enables long-term archiving of analytical instrument data as specified in the Long-Term Archiving Guidance for Analytical Instrument Data guide.

Chapter 2 describes the scope of this Technical Guidebook; Chapter 3 describes the background (including regulatory requirements) that led to the creation of this Technical Guidebook. This is followed by chapters explaining the main topic of this Technical Guidebook. The main topic of this Technical Guidebook (Chapters 4-5) is as follows for its target readers.

Table 1. Composition of this Technical Guidebook

Chapter/Title	Target Reader
Chapter 2: Definition of Scope Chapter 3: Background and Regulatory Requirements	All parties (common matters such as assumptions).
Chapter 4: Operating the Package	On-site personnel (for standard package users) using and operating standard packages in accordance with Long-Term Archiving Guidance for Analytical Instrument Data.
Chapter 5: Standard Package	People who wish to know or reanalyze the technical specifications of the long-term archiving package described in Long-Term Archiving Guidance for Analytical Instrument Data (for standard package implementers).

2. Scope of Application

This Technical Guidebook provides technical considerations for the analytical instrument data required for reprocessing in the pharmaceutical industry.

3. Introduction

The purpose of Long-Term Archiving Guidance for Analytical Instrument Data is to provide a reliable, long-term, reassuring method of archiving and management with the assumption that the data may be reanalyzed. Two technical items are considered important.

The first is "assurance of reliability." To assure reliability, a "content assurance" is required to indicate that there are no changes from the time of archiving; an "identity assurance" is required to check the person responsible for archiving; a "time assurance" is required to check the archiving time; a "long-term assurance" is required to respond to long-term archiving, and an "operational assurance" is required to indicate that there are no problems in the overall operation. "Content assurance," "identity assurance," "time assurance," "long-term assurance," and "operational assurance" are components of "authenticity," and this Technical Guidebook explains the technology to maintain authenticity.

The second item is “reprocessing of data.” The possibility that the analyst performing the reprocessing is not identical to the analyst at the time of archiving, and that the analyzing environment at the time of reprocessing is not identical is a problem. The introduction of a standardized package (standard package) is required as a technical element to solve this. In this Technical Guidebook, the specification of the standard package is

explained, also taking into account the cost side.

3.1. Pharmaceutical Regulations and Guidance

Guidance for Long-Term Archiving of Analytical Devices is intended for use in the pharmaceutical industry. As regulatory information and guidance for the pharmaceutical industry, there are the Japanese ER/ES Guidelines (*1) and the U.S. Part 11 (*2), and as the standards for computerized systems and data in drug manufacturing to be complied with internationally, PIC/S GMP Guide ANNEX 11 (*3). Both have become rules and guidance for electronic records or data and electronic signatures, with commonalities in what is sought.

Table 2. Regulatory Requirements for Electronic Records and Electronic Signatures

Electronic Record	Electronic records are properly managed and operated.
Electronic Signature	The electronic signature is treated equally with handwritten signatures, linked to the electronic record and the reason for the signature, and cannot be used separately.

Regarding electronic recordings or data, predominantly management and operational methods are described. Especially, the technology of "operation" becomes important in any rule and guidance. Control of open and closed systems, explicit electronic signatures, and the linkage of electronic records to electronic signatures are required to ensure reliability.

Electronic signatures are defined as equivalent to handwritten signatures in any rule guidance. The term “electronic signature” does not refer to the technical content but is primarily the electronic sign (evidence) conferred on electronic recordings (data) in a legal sense. Technically, in addition to digital signatures using public key cryptography and public key infrastructure (PKI), there are authentication records using electronic authentication (ID, password, etc.). Electronic signatures may be identical to digital signatures in a narrow sense, but they are used in this Technical Guidebook in a broad sense that does not depend on technology. In other words, the relationship is an "electronic signature ♂ digital signature" in which both are not equal.

Assurance and electronic signatures for the control and operation of electronic records require authenticity. Techniques to assure these are discussed in Section 3.4, Outline of Assurance Techniques.

Note 1: Japanese ER/ES Guidelines: "Use of Electronic Records and Electronic Signatures in Applications for Approval or Licensing of Drugs" (PFSB/ELD Notification No. 0401022) by MHLW
https://www.pmda.go.jp/files/000158308.pdf

Content (excerpt):

3.Requirements for Use of Electronic Records

3.1. Method of Management of Electronic Records

3.1.1. Authenticity of electronic recordings (complete, accurate, reliable, clear person responsible for creating, changing, and deleting: procedures and implementation of security keeping and back-up, person identification, audit trail)

3.1.2. Readability of Electronic Recordings (That Can Be Produced in Human Readable Format)

3.1.3. Archiving of electronic recordings (keeping authenticity and readability within the archiving period: procedures and implementation of recording media management, maintenance of

archiving during media transitions)
3.2. Use of Closed Systems
3.3. Use of Open Systems
4.Requirements for the Use of Electronic Signatures

Note 2: US Part11:FDA-issued 21 CFR Part 11 "Electronic Records; Electronic Signatures
https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=11
Content (Excerpt):
Subpart B: Electronic Recordings
11.10 Controls for Closed Systems
11.30 Controls for Open Systems
11.50 Signature Manifestations
11.70 Signature/Record Linking
Sub-Part C: Electronic Signature
11.100 General Requirements
11.200 Electronic Signature Components and Controls
11.300 Controls for Identification Codes/Passwords

Note 3 PIC/S GMP Guide ANNEX 11 (Computerised Systems):PE 009-14 (Annexes)
https://picscheme.org/docview/1946
Partial Revision of "Concepts in Using the GMP-Guideline of PIC/S"
https://www.pmda.go.jp/files/000202984.pdf
Content (Excerpt):
Annex 11 Computerised Systems
Principle
General
1. Risk Management
2. Personnel
3. Suppliers and Service Providers
Project Phase
4. Validation
Operational Phase
5. Data
6. Accuracy Checks
7. Data Storage
8. Printouts
9. Audit Trails
10. Change and Configuration Management
11. Periodic Evaluation
12. Security
13. Incident Management
14. Electronic Signature
15. Batch Release
16. Business Continuity
17. Archiving

3.2. Operational Costs and Assurance Levels

The operating costs required to ensure reliability are proportional to the assurance level. In other words, to

increase the assurance level of reliability, the operation costs must also be high. However, because there are only a limited number of organizational resources in reality, limitations in operational costs are also required. Accordingly, companies need to make decisions according to their reliability assurance policy (to what extent they require reliability) and resources (such as the cost of introducing technology and the time required for operating management by staff). In this Technical Guidebook, three steps are divided into four, and explanations are provided. Referring to this, each company should consider the response. As previously described, components that assure authenticity include content assurance, identity assurance, time assurance, long-term assurance, and operational assurance, but the overall assurance level is determined by the lowest assurance level among the components. For example, even if a high assurance level of technology is used for identity assurance, the overall assurance level is reduced according to the long-term assurance level if a low assurance level of technology is used for long-term assurance. Therefore, it is recommended that each component be matched with the assurance level sought.

Table 3. Reliability Assurance Level for a Standard Package

Assurance Level	Summary	Description
Level 1	Assurance as Minimum-Cost Operation	Minimum operational level that is not saving on cost as much as possible.
Level 2	Balance-Based Operation	Operational level with superior cost performance.
Level 3	Highest Level of Assurance Operation	Operational level that achieves the highest assurance without considering the cost.

3.3. Standardization of Package (Standard Package)

If it is assumed that the analytical instrument data is stored for a long time and that the analytical instrument data itself is reanalyzed in the product of another company, the package format that uses only the in-house product cannot be fully addressed. Given that the verifiers differ, a common specification is required to verify authenticity by the verifiers. To solve this problem, the interoperability is maintained by determining the specifications of the standard package for which the package structure is standardized. There are three types of information required to create standard packages: Data specifications in standard packages, catalog information, and additional assurance information.

Table 4. Information Required to Create a Standard Package

Information Type	Description
Standard Package Data Specifications	The types and specifications of the files included in the standard package are not included in this Technical Guidebook. See Long-Term Archiving Guidance for Analytical Instrument Data.

Information Type	Description
Standard Package Inventory Information	The standard package adopts the ZIP format, and the META-INF/Manifest (catalog information) file is adopted as the standard package specification for the file name list and content assurance in the standard package. Basically, the inventory information becomes indispensable. The META-INF/Index (index information) of information such as time of preparation and author is also required.
Standard Package Additional Assurance Information	The META-INF/Signature file and META-INF/Timestamp file shall be adopted as standard package specifications as additional assurance information for ensuring the identity and time of each file in the standard package. This additional assurance information is voluntary.

3.4. Overview of Assurance Techniques

The Long-Term Archiving Guidance for Analytical Instrument Data guidebook proposes a standard package of long-term archiving to provide a reliable, long-term, reassuring method of archiving and management on the premise that data may be reanalyzed. Pharmaceutical regulations and guidance require authenticity, readability, and preservability to assure the reliability of standard packages for electronic recordings. In this Technical Guidebook, five components to assure authenticity are described: Content assurance, identity assurance, time assurance, long-term assurance, and operational assurance. Long-term assurance also includes preservability in the sense of preserving authenticity within the archiving period. Readability is not discussed in this Technical Guidebook.

Table 5. Components of Authenticity

Purpose	Component	Description
What	Content Assurance	Ensure that the content remains at the recorded time and is not falsified. (Related to original, accurate, complete, and consistent in ALCOA+.)
Who	Identity Assurance	The person in charge of the record shall be clarified and therefore shall not be able to repudiate its authenticity. (Related to attributable and complete in ALCOA+.)
When	Time Assurance	Assurance of recorded time (time stamp, confirmation that it exists). (Related to original, contemporaneous in ALCOA+.)
Archiving	Long-Term Assurance	Techniques and formats for ensuring and preserving long-term content assurance and identity assurance. (Related to enduring, available, and legible in ALCOA+.) Note: The Long-Term Archiving Guidance for Analytical Instrument Data guide assumes 10-30 years as a long-term archiving period, and this Technical Guide will also assume 10-30 years.
Operation	Operational Assurance	Appropriate operations should be performed as a whole, and operations should be documented and audited for assurance. A base assurance relevant to the above assurances is provided for operation.

4. Operation of Standard Package

Since operations are very important components when ensuring authenticity, the operation of standard packages is discussed mainly in this chapter. To consider operations related to standard packages, it is necessary to

understand the workflow from the first creation of a standard package to re-analyzation. In addition, we consider how to assure the authenticity of standard packages.

4.1. Standard Package Workflow

The following figure shows the workflow of the standard package.

Table 6 Standard Package Workflows

Steps	Tools Used	Description	Purpose
⓪ Measurement and Analysis	Measurement Software Analysis Software	Measure and perform the first analysis.	
① Export	HPLC Data Management System	Output the data required for analysis to the export folder.	Archiving
② Package Creation	Package Tools	Create metadata using files in the export folder and, if necessary, externally supplied information, and compile and record standard package files using ZIP. Consider setting and operation of access rights so that data is not tampered with from export to standard package creation. The export folder is a temporary area and is deleted after the package is created.	
③ Upload (Write)	Data Storage Server (WORM Media)	Upload the standard package file on the server. On the server side, record who was uploading. Consider setting and operation of access rights so that uploaded files are not tampered with. This may be stored in a non-rewritable medium.	
④ Download (Read)	Data Storage Server (WORM Media)	Search and download the standard package of data to be reanalyzed. On the server side, record who downloaded and when. In some cases, data is read from a medium that cannot be rewritten.	Reprocessing
⑤ Verification and Unpackaging	Package Tools	Verify the standard package file, check the authenticity (falsification, etc.), and output the data in the standard package to the import folder.	
⑥ Import	HPLC Data Management System	Import and record data from the import folder. Consider setting and operation of access rights so that data is not tampered with from verification decompression to import. The import folder is a temporary area and is deleted after import.	
⑦ Reprocessing	Analysis Software	Reprocessing is performed.	

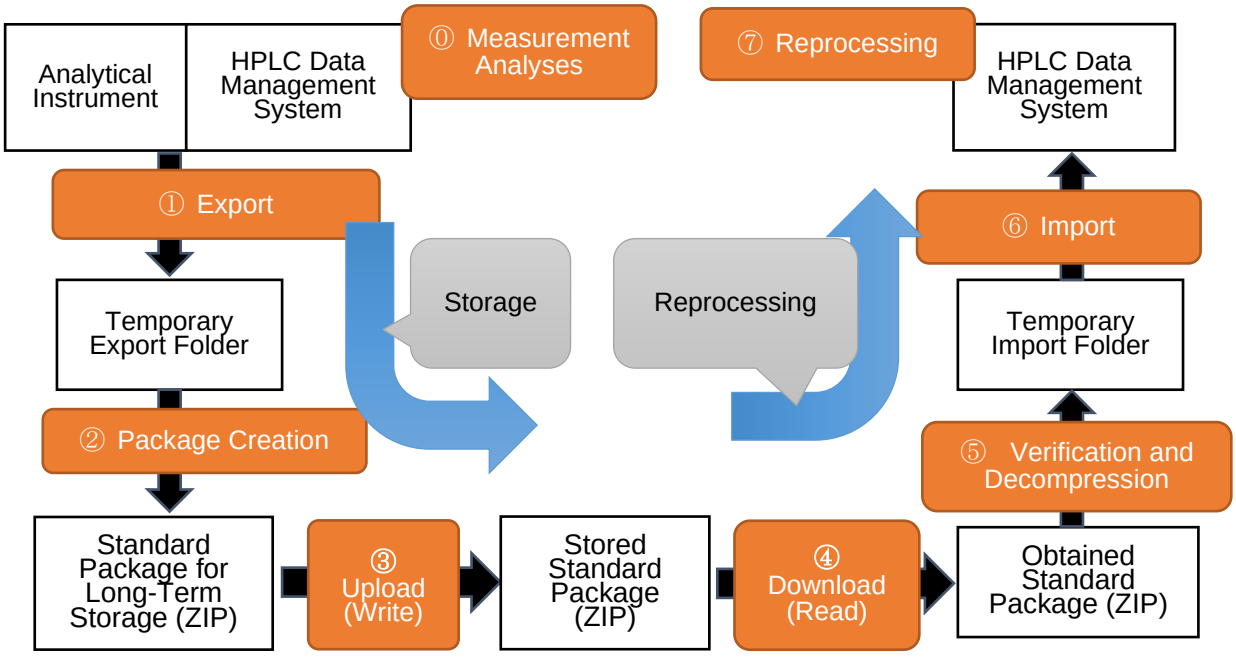

Figure 1. Flow of Analytical Instrument Data

In this case, it is assumed that the analytical instrument data will be used in the company after long-term archiving, and that the analytical instrument data will be sent to the outside company for use in two ways. In addition, two types of archiving methods are assumed: A server and an electronic recording medium. Considering this combination, the following four types of operational patterns are assumed, but not limited to:

Table 7. Assumed Operational Patterns

	During Export Data Processing Server	Storage	During Import Data Processing Server
Pattern 1	Internal Server	Internal Server	Internal Server
Pattern 2	Internal Server	Electronic Recording Media (CD, DVD, etc.) Note: Rewrite not available.	Internal Server
Pattern 3	Internal Server in the Organization A	Cloud Server	Internal Server in the Organization B
Pattern 4	Internal Server in the Organization A	Electronic Recording Media (CD, DVD, etc.) Note: Rewrite not available.	Internal Server in the Organization B

The workflow of standard packages in these four patterns is illustrated as follows.

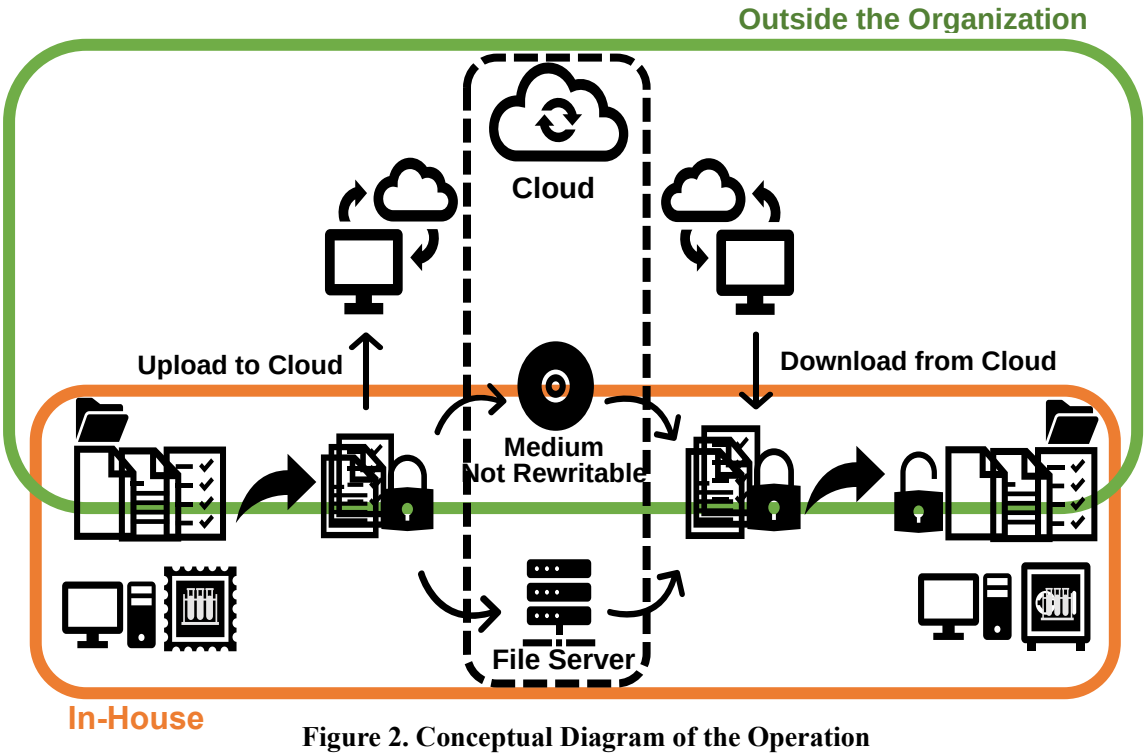

Figure 2. Conceptual Diagram of the Operation

The workflow for the operation is shown in the case of operations in a closed environment such as in-house, and in the case of operations in an open environment using the cloud.

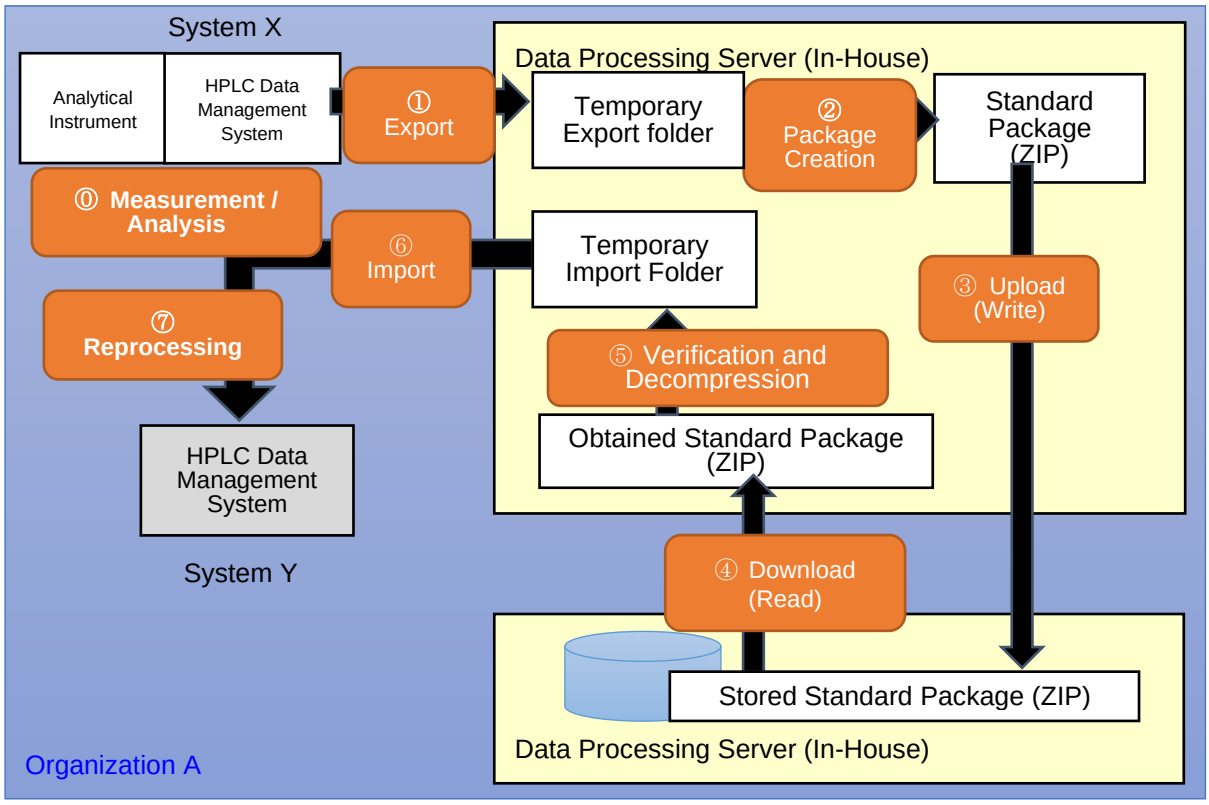

Figure 3. Flow of Analytical Instrument Data: Examples of In-House Closed Environments

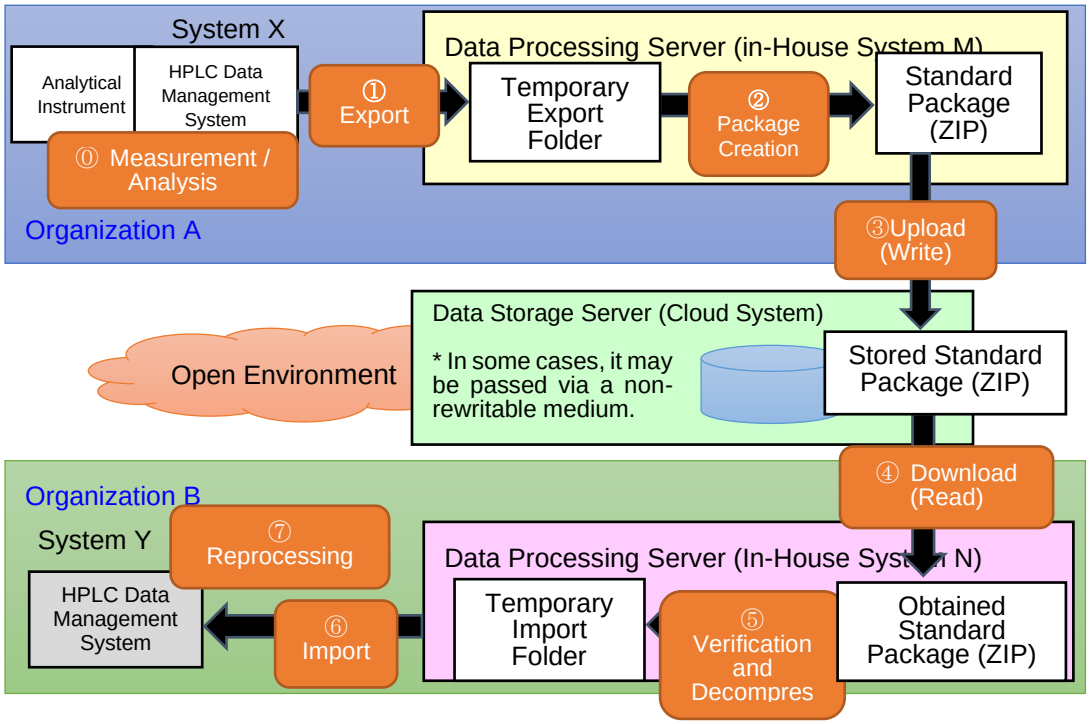

Figure 4. Flow of Analytical Instrument Data: An Example of an Open (Cloud) Environment

Details of required security requirements differ between closed and open environments. ER/ES Guidelines and the US Part 11 summarize these requirements as separate chapters. Open environments generally require a higher level of security. However, when considering the context in which the package is used, comparable management may be required in all cases. A step-by-step explanation is given below.

When using closed systems (including analytical devices) such as those in the company, content assurance and identity assurance by means of electronic signatures, etc. are required. When using an open system that is placed externally like a cloud, in addition to the requirements required for a closed system, it is also required to consider confidentiality due to coding, etc. and operational assurance to maintain authenticity in preparation and utilization procedures.

However, even if a standard package is created in a closed system, it may be provided outside the system (e.g., a package is provided to the sponsor by the Contract Research Organization (CRO). In this case, it is desirable to respond to the additional requirements required for the open system in the standard package (e.g., content assurance by digital signature and time stamp).

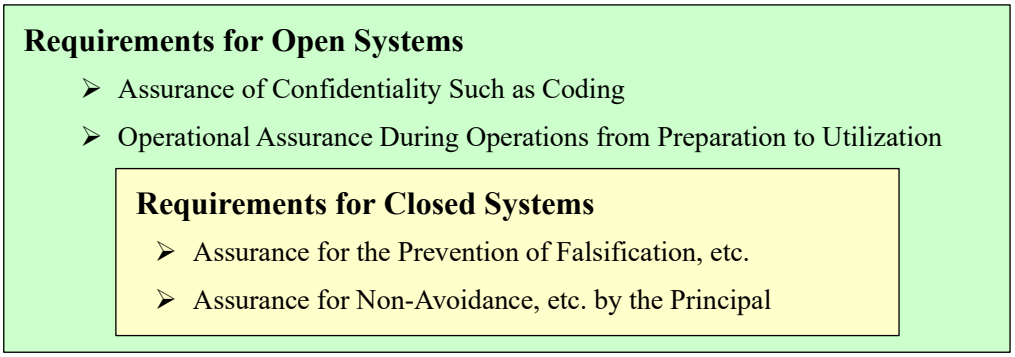

Figure 5. Requirements for Closed and Open Systems

4.2. Standard Package Authenticity Assurance Prerequisites

It is imperative that analytical instrument data before standard package storage are adequately reliable at each facility at a level appropriate to the purpose of storage. When operating in compliance with GxP, a third party (e.g., QA) must assure the appropriateness of the process from export of analytical instrument data to import into a new system under the premise that data integrity of analytical instrument data prior to packaging is ensured (see Long-Term Archiving Guidance for Analytical Instrument Data, Section 4.3). If this process can be automated, it may reduce assurance services by a third party at the time of operation by appropriately validating and ensuring the reliability of the automated process at the time of the introduction of the program involved in automation.

4.3. Examples of Assurance Methods for Standard Packages by Operational Cost

Selecting and combining each component of authenticity assurance as shown in Table 5 with the specifications of the standard package described in Section 5 should determine its operating specifications. In this section, a three-stage operation method is proposed according to the operation cost, so as to provide a reference when considering the operation specification.

4.3.1. Assurance Level 1: Assurance of Minimum Cost

The operations that should be achieved at a minimum with limited operational costs are illustrated here. Basically, external services and PKIs (certificate authorities) are not used. Although the standard package can be tampered with, it detects tampering by monitoring whether the hash value has been changed in each process and prevents tampering by operation such as not using the package when tampering is detected. Hash value calculation enables a third party to check for tampering from the time of creation.

Table 8. Case 1: Building and Operating a Data Storage Server In-House

Item	Content
Content Assurance	Perform falsification detection by hash value or operation record. Note 1: Tampering assurance must be handled by the operation of the server to be stored. Note 2: Content assurance via detection of falsification by operating records is also acceptable.
Identity Assurance	Use authentication records. Example: ID and password authentication management and audit log storage.
Time Assurance	Use system time.
Long-Term Assurance	Store in the data storage server. Store on an electronic recording medium.
Operational assurance	The operation policy is prepared, the audit log is kept as an audit trail, and the operation audit is received.
Package specifications	The content is assured by a Manifest file of hash-value computation. Reference: 5.1.1
Cost	There are virtually no operational costs as external services are not used. Resources may be required for a complementary process.

4.3.2. Assurance Level 2: Balance-Based Operation

Three types of operation methods are illustrated here, taking into consideration the balance between operational costs and assurance levels.

4.3.2.1. When Using an Electronic Document Management System

If the cost burden of introducing a vendor-provided electronic document management system is acceptable, minimal function can be realized. Introduction of the package software into the company generally requires initial and maintenance costs. On the other hand, using cloud services requires periodic operation costs. However, it has the advantage in that it is easy to introduce it because software and services that already have functional implementations can be used. Standardized packaged files may be stored in an electronic document management system, or they may be packaged in a standardized package when stored without standardization and when providing external data.

Table 9. Case 2: Use of Commercially Available Electronic Document Management Systems

Item	**Content**
Content Assurance	Conduct content assurance by detecting falsification using operating records. Note: The electronic document management system provided by the vendor is used.
Identity Assurance	Use authentication records. The authentication function of the electronic document management system provided by the vendor is used.
Time Assurance	Use system time. Note 1: When introducing an electronic document management system as a package software in the company, use the system clock of the company's server. Note 2: Use the system clock of the vendor-provided electronic document management system when using cloud services.

Item	Content
Long-Term Assurance	Keep in the electronic document management system.
Operational Assurance	The operation policy is prepared, the audit log is kept as an audit trail, and the operation audit is received.
Package Specifications	Standard packages are used to provide external data. The content is assured by the hash-valued Manifest file and the Timestamp file or the digitally signed Signature file as required. References: 5.1.1 / 5.1.2 / 5.1.3 / 5.1.4
Cost	The cost of using the electronic document management system (operation) is required.

4.3.2.2. When Using the Time Stamp Server

Time stamping can also be used to prevent tampering with standard packages. The operation cost of the time stamp server is high, but the tampering prevention is high as a standard package unit compared with the assurance level. Using an external time stamp service assures a high assurance level and eliminates the need to operate the time stamp server, but it incurs the operational costs of using the service.

Table 10 Case 3: Use a Time Stamp Server

Item	Content
Content Assurance	Time stamp tampering is prevented.
Identity Assurance	Use authentication records. Example: ID and password management and audit log storage
Time Assurance	Use a time stamp server. External time stamp services are often paid operations, so consider running a time stamp server internally to reduce operational costs.
Long-Term Assurance	Store in the data storage server. Store on an electronic recording medium.
Operational Assurance	The operation policy is prepared, the audit log is kept as an audit trail, and the operation audit is received.
Package Specifications	The content is assured by Manifest file of hash value calculation and Timestamp file of the time stamp. References: 5.1.1 / 5.1.2 / 5.1.4
Cost	The cost of building a time-stamp server or operating an external time stamp service is required. However, the use of external time stamp services will provide higher levels of assurance and would be a sufficient reason to increase costs.

4.3.2.3. Use of Digital Signatures by Private Certification Authorities

As a balance-based operation, a digital signature can be operationalized by constructing an in-house certification office (private PKI).

Table 11. Case 4: Use of Digital Signatures by a Private Certificate Authority

Item	Content
Content Assurance	Prevent falsification by digital signature.
Identity	Use certificates issued by creating an in-house Certificate Authority (Private PKI).

Assurance	Example: A certificate/private key for signing is prepared and digitally signed for each user. Although external costs are eliminated, in-house operational costs are required.
Time Assurance	Use system time.
Long-term Assurance	Store in the data storage server or on an electronic recording medium. Note: A corporate time stamp server enables long-term signing.
Operational Assurance	The operation policy is prepared, the audit log is kept as an audit trail, and the operation audit is received.
Package Specifications	Assure the content of the hash-value computation Manifest file and the digitally signed Signature file (only digitally signed signature files). References: 5.1.1 / 5.1.2 / 5.1.3
Cost	In the case of private certification authorities, the operational costs are not very high but their creation and operation require a certain level of expertise, so human resource costs are to be considered.

4.3.3. Assurance Level 3: Maximum Level Assurance (High Cost)

Using a public certificate authority or time stamp service can provide the highest level of assurance of operational costs.

Table 12. Case 5: Use of External Public Certification Authorities and Time Stamp Services

Item	Content
Contents Assurance	Prevent falsification by digital signature.
Identity Assurance	Use certificates issued using public certificate authorities (public PKIs). Note: Purchase and digitally sign certificates/confidential keys from public certificates.
Time Assurance	Use an external time stamp service (time stamp server). Note: It is the time assured by the third party and the reliability is high, but it is necessary to contract the service and the operation cost can be expensive.
Long-Term Assurance	Store in the data storage server. Store on an electronic recording medium. Long-term signatures extend the assurance period of storage in either case.
Operational Assurance	The audit log will be kept as an audit trail and the operation audit will also be conducted.
Package Specifications	Assure the content of the hash-value computation Manifest file and the digitally signed Signature file. The signature file shall be in the form of a long-term signature with a digital signature + time stamp. References: 5.1.1 / 5.1.2 / 5.1.3 Note: Long-term signatures enable assurance and verification of content assurance, identity assurance, time assurance, and long-term assurance by standard package units. For this purpose, the standard package is the most suitable form for distribution (external supply, etc.).
Cost	The operational cost of the time-stamp service is required in addition to the cost of purchase of certificates from the certification authority. In addition, the number of time stamps required to use long-term signatures also increases, so that time stamp services for multiple charges can be more costly.

5. Standard Package

This chapter provides technical explanations on interoperability in two sections: 5.1. Standard Package Specifications and 5.2. Standard Package Workflows.

5.1. Standard Package Specifications

Adopt ZIP and XML as standard package specifications for analytical instrument data. Meta-information about standard packages shall be placed in the META-INF directory. This method is also used in many existing standard packages (EPUB, etc. OOXML, ODF,). However, what is placed in META-INF varies with each standardized specification.

Table 13. Files to Add in META-INF

File name	Creation	Type	Description
Index.xml	Mandatory	Index information	Complementary information is provided.
Manifest.xml	Mandatory	Inventory Information	A file that records the reference (URI) of the target analytical device data and its hash value.
Signature.xml	Optional	Digital Signature Information	A file to assure catalog information with a digital signature. Identity assurance and prevention of falsification are possible (time assurance is also possible if time stamps are added).
Timestamp.tst	Optional	Time Stamp Information	A file to assure catalog information with a time stamp. Falsification prevention and time assurance are possible.

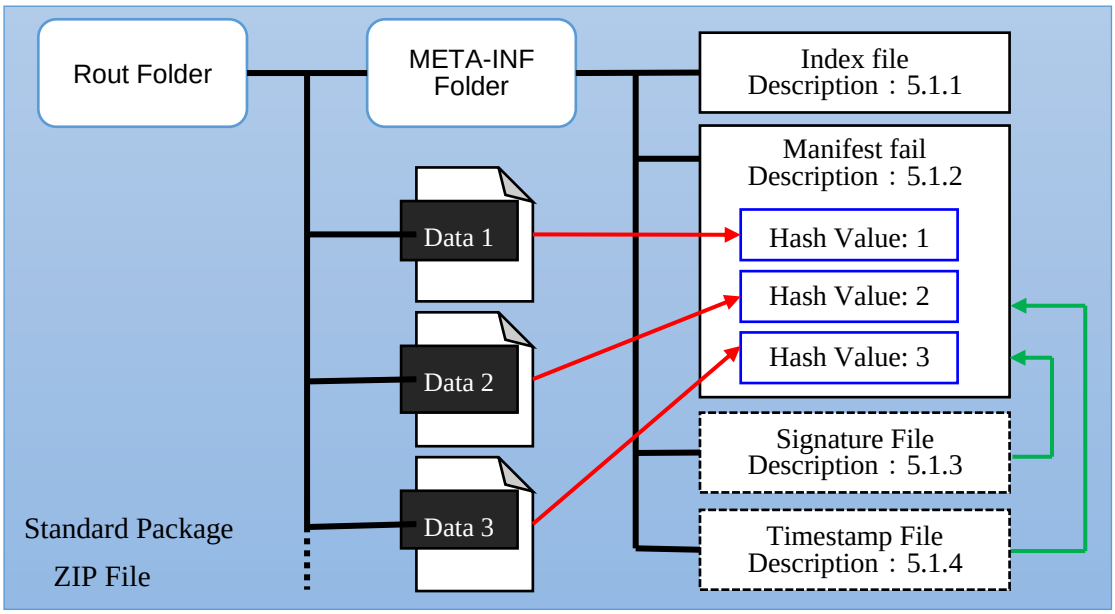

Figure 6. Examples of Structures in the Standard Package

5.1.1. META-INF/Index File (Mandatory: Indexing Information)

For this standardized package, a unique specification-META-INF/Index file-was considered. The extension of this file is set to.xml, and the full path of the file is set to META-INF/Index.xml. The Index file should contain at least the creation date and time and the author ID information, but it should also contain the name of the measured data and the search keywords to understand the content of the file. Additional information may be provided in this specification, if necessary. Since the Index file is not included in the Manifest reference, it may be updated after signing.

Table 14. XMLspec for Index

XML Tag Name	Designation	Attribute	Description
Index	Mandatory	(Id="identifier")	Index root element.
Title	Optional	---	Names of measured data.
Date	Mandatory	---	Creation date and time. The time of the system is good.
User	Mandatory	---	Creator ID (login user ID, etc.) It is best to use the user ID at the time of authentication.
Host	Optional	---	Created system name (computer name, IP address, etc.)
Keyword	Optional	---	Keywords for searching (multiple can be specified).
(Optional)	Optional	---	Any other XML tag name can be added.

Note: The format of the Index file is proprietary to this standard package.

Sample Index Files:

```
<Index>
 <Title> Sample data </ Title>
 <Date>2021-02-26T13:50:20</Date>
 <User>hanako</User>
 <Host>MyPC1</Host>
</Index>
```

5.1.2. META-INF/Manifest File (Required: Catalog Information)

The Manifest (Manifest) file, which is mandatory for standard packages, is also called catalog information. The Manifest format included in the XML Signature (the standard specification of the W3C Recommendation) shall be adopted as the standard package for measuring instrument data. For this purpose, the extension is set to xml and the full path of the file is set to META-INF/Manifest.xml. In this file, a plurality of analytical device data and related information files are specified, and the hash value of each target file is recorded. Therefore, when the target file is tampered with, the hash value in Manifest does not coincide with the recalculated hash value, so that falsification of the target file can be detected. Of course, it is not possible to tamper-protect Manifest files themselves by means of Manifest files. To prevent falsification of Manifest files, separate Signature files and Timestamp files must be protected.

Table 15. XML-Specifications for Manifest

XML tag name	Designation	Attribute	Description
Manifest	**Mandatory**	(Id="identifier")	Manifest root element.
Reference	**Mandatory**	URI="Target"	URI specification of the reference destination (multiple specifications are allowed).
DigestMethod	**Mandatory**	Algorithm = "hash method"	Algorithm specification for hash calculation.
DigestValue	**Mandatory**	---	Base64 and store hashed data.

Note 1: In the XML Signature spec, it can be omitted if it is treated as a binary file that Transforms tags are available underneath the Reference tag. If an XML file is included in the object, a Transforms may be specified.

Note 2: XML Signature Syntax and Processing Version 1.1 - W3C Recommendation 11 April 2013 https://www.w3.org/TR/xmldsig-core1/(Link)

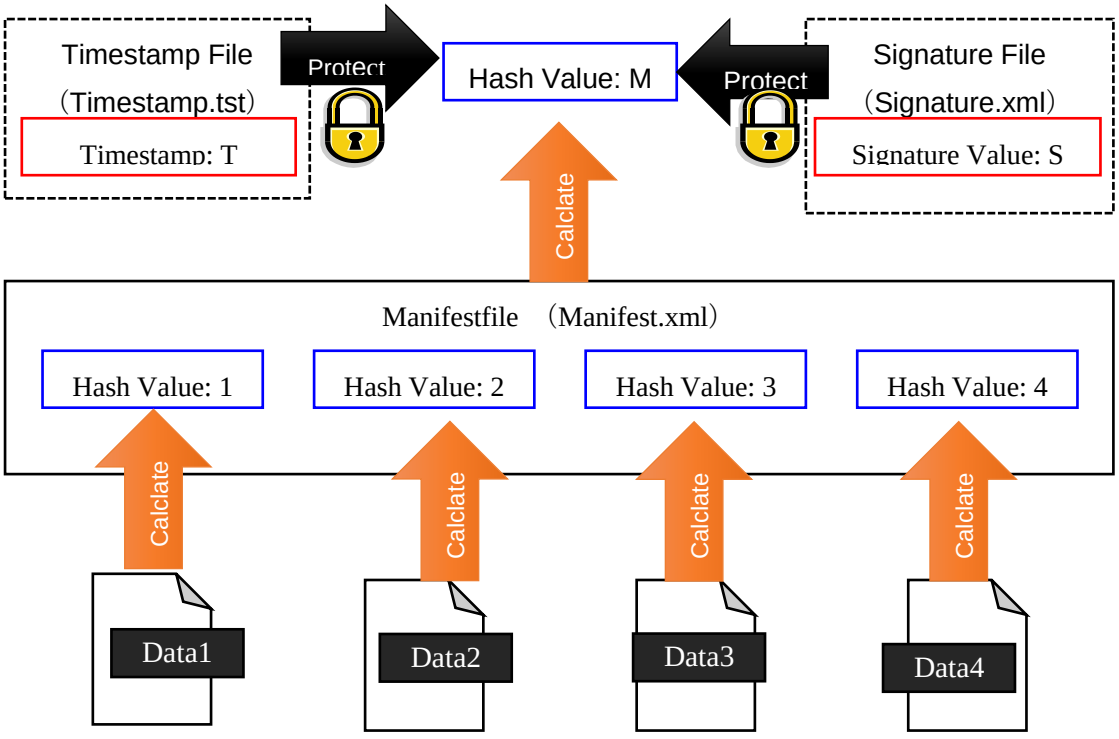

Figure 7. Relevant Diagram of Content Assurance by Each File

Sample Manifest Files:

```
<?xml version="1.0" encoding="utf-8"?>
<Manifest Id="Id-Manifest-0" xmlns="http://www.w3.org/2000/09/xmldsig#">
 <Reference URI="../Demo_Data-001.lcd">
  <DigestMethod Algorithm="http://www.w3.org/2001/04/xmlenc#sha512" />
  <DigestValue>
EA7YFkhjtpUDcCC+SwK9A1lHW+Swxh4G1dUMmLsAItyCj5y8SLhvzRcvK/j0f9+5k2R6W6N7n6WL+vUx5eTRRQ==
</DigestValue>
 </Reference>
 <Reference URI="../Result File/Demo_Data-001.pdf">
  <DigestMethod Algorithm="http://www.w3.org/2001/04/xmlenc#sha512" />
  <DigestValue>
3nJV5vSblaRvUffKPoCiDkZE1mYI4KozwWZxrEpzY84M6fi0a7LYO7fy7G5cYbgzkQ0hlJAKoc0do2xIgLFq1A==
</DigestValue>
 </Reference>
 <Reference URI="../Result File/Detector-A-Ch1.CDF">
  <DigestMethod Algorithm="http://www.w3.org/2001/04/xmlenc#sha512" />
  <DigestValue>
Azc6e4JEdkuOwmiilWKDcJl7hvBebt80RB12pqumdvDhx2CfYochudHzwJjvKUzAJyTlHBTWU7YZk6s2Udk+iQ==
</DigestValue>
 </Reference>
</Manifest>
```

5.1.3. META-INF/Signature File (Optional: Additional Identity Assurance Information)

Add a Signature file to add a digital signature to the Standards Package for the customer's assurance. In this paper, XAdES (ISO 14533-2) of long-term signing of XMLSignature is adopted. The extension is.xml, and the full META-INF/Signature of the file is.xml. It is also possible to use the XAdES-BES format for signing only, and to add a time stamp by using XAdES-T. As the signing target (SignedInfo/Reference), specify two items: META-INF/Manifest.xml and the XAdES signing target attribute (SignedProperties).

Table16 XMLspec for XAdES

XML tag name	Designation	Attribute	Description
Signature	Mandatory	(Id="identifier")	Signature root element.
---	Mandatory	---	Conform to XAdES spec.

Note: ISO/DIS 14533-2 - Part 2: profiles for XML Advanced Electronic Signatures (XAdES)
https://www.iso.org/standard/79129.html

5.1.4. META-INF/Timestamp File (Optional: Additional Time Assurance Information)

Add a Timestamp file to add a time stamp for time assurance to the standard package. Since there is no XML-based specification that gives only time stamps, a time-stamp token is employed. The hash value (MessageImplint) of the time stamp token is calculated by calculating Manifest.xml as a binary file, and the full path of the file is META-INF/Subject.tst.

Note: RFC 3161 - Time-Stamp Protocol (TSP) Laurie Start Here
https://tools.ietf.org/html/rfc3161

5.2. Procedures for Using Standard Packages

This chapter describes the technical concepts for accomplishing the standard package workflow discussed in Sections 4 and 1. Staff can also manually create or thaw standard packages based on this procedure, and this procedure can be achieved based on automated tools. Currently, the committee members provide automation tools and data storage servers to create and verify packages based on this Technical Guidebook, which is used to examine the feasibility of this workflow. The results of the trial will be provided as an appendix in the future. The process leading to the creation of the standard package is shown in Section 5.2.1 and the process for unpackaging and verification of the standard package for reprocessing is shown in Section 5.2.2.

5.2.1. Standard Package Preparation Procedure (Archiving)

The minimum requirements for creating standard packages are the indexing, catalogue information creation, and ZIP functions. In addition, the digital signature creation and time stamp acquisition functions are prepared as necessary as optional functions to assurance authenticity.

Table 17. Standard Package Preparation Procedures

Procedure	Overview	Description
1	Preliminary Preparation	Prepare all analytical instrument data for the standardization package in any folders.
2	Creation 1	Create indexing data (META-INF/Index.xml), such as the creation date and time and search keywords.
3	Creation 2	Reference all files in the folder to create a hashed catalogue (META-INF/Manifest.xml).
4	Authenticity	Optional: Create a digital signature (META-INF/Signature.xml) or time stamp (META-INF/Timestamp.tst) as a file to preserve the authenticity of the standard package.
5	Packaging	ZIP all files (including META-INF folders) in the folder.
6	Cleanup	Delete a file other than the standard package (ZIP file) that you created. Delete the temporary files prepared or created in Steps 1-4.
7	Archiving	Perform the archiving operation for the created standard package.

5.2.2. Standard Package Validation Procedure (At Reprocessing)

The ZIP decompression function and the catalog information verification function are required at a minimum to verify the standard package. In addition, the digital signature verification function and time stamp verification function are prepared as necessary.

Table 18. Standard Package Verification Procedures

Procedure	Overview	Description
1	Preparation 1	Specifies a standardized packaged file.

Procedure	Overview	Description
2	Preparation 2	ZIP extracts the standardized packaged files and extracts all the files in the standard package into the folder of your choice.
3	Verification 1	Match the existence of all files listed in the catalog (META-INF/Manifest.xml) with the hash value shown by calculating the hash value. If there is no agreement, check the content and consider the action.
4	Verification 2	If the META-INF/Signature package contains a digital signature (.xml) or time stamp (META-INF/Timestamp.tst) file, verify the content of each file. If a verification error occurs, check the content and consider the action.
5	Use	Reprocessing is performed using thawed analytical instrument data.

5.2.3. Reference: Tools for Standard Packages

An open-source tool, AdDataPackager, that supports standard packages has been released. As a creation function, the indexing, cataloguing information creation, zipping, and time stamp acquisition functions are supported, but the digital signature creation function is not supported. As verification functions, the ZIP decompression, catalog information verification, and time stamp verification functions are supported, but the digital signature verification function is not supported. The license is MPL2.0, so commercial use is also possible.

AdDataPackager Public URL

https://www.ossal.org/salproj/adpack.html

MPL2.0 (Open Source License)

https://www.mozilla.org/en-US/MPL/2.0/

Table 19. AdDataPackager Information

Name OsSAL/AdDataPackager Name：Advanced Data Packager abbreviation：ADPACK Development：OsSAL.org (Open-Source Signature and Certification Lab)
Abstract (Overview) Adpack is a tool for packing, verifying, and unpacking a package file in a ZIP format with a verification function based on hash values. The specifications of the package file conform to the package of the Long-Term Archiving of Analytical Instrument Data – Technical Guidebook Edition 1.0 published by JIIMA. You can get the Guidebook via the following link. Since the package has general-purpose specifications, it can be used for purposes other than analytical instrument data. When packing, the hash value of each file to be packed is enumerated in META-INF/Manifest.xml and then zipped and compressed. When unpacking, compare the hash value enumerated in META-INF/Manifest.xml with the hash value of each compressed file before unzipping it. META-INF/Manifest.xml follows the W3C XML Signature specification. W3C XML Signature Syntax and Processing Version 2.0 https://www.w3.org/TR/xmldsig-core2/

Functions (Main Functions)
AdPack provides two functions: packaging/Pack and unpackaging/UnPack.

* Pack Operations
    1. Automatic generation of META-INF/Index.xml
    eng) Create META-INF/Index.xml.
    2. Automatic generation of META-INF/Manifest .xml from the hash value of each target file
    eng) Create META-INF/Manifest.xml with hash values.
    3. ZIP the target file and the generated META-INF file together
    eng) Packing to ZIP file with META-INF/*.

* UnPack Operations
    1. Check hash values include files by META-INF/Manifest .xml (tamper check)
     eng) Check hash values include files by META-INF/Manifest.xml.
    2. UnZIP process
     eng) Extract to all files.

6. Glossary of Terms

Table 20. Terminology

Term	Description
Electronic Record	Legal definitions are electronic, magnetic, or other recognizable recordings that are used in a way that is not recognized by human perception and is used for information processing by electronic calculators. It generally refers to electronic data.
Electronic Signature	In the legal (broad) sense, electronic evidence to be provided to electronic records (data). In a narrow sense, it may refer to a digital signature.
Digital Signature	A type of digital signature using public key cryptography and hash functions that enables non-tampering and non-repudiation. Public Key Infrastructure (PKI) can be used to assurance identity.
Data Storage Server	A system that stores data (files) on a server. It also includes systems that are not access-controlled, such as file servers. Electronic document management systems are also included.
Electronic Document Management System	This system stores and manages electronic documents (data), and provides access control. In recent years, the development of electronic document control systems complying with the 21 CFR Part 11 and MHLW ER/ES Guidelines has promoted their use in the pharmaceutical industry.
Certification Record	This is realized by recording the results of electronic certification and can be used as a simplified electronic signature. For example, this can be realized by storing (recording) the log of the authentication results by ID and password.
Hash Value (Digest Value)	It is sometimes referred to as a message digest value, but it is consistent with a hash value in this document.
Base64	Specification for encoding binary data using only 64 printable alphanumeric characters. This has the advantage in that it can be smaller than Base16 (hex) format, which uses hexadecimal 16 alphanumeric characters, and is often used in networking, XML, etc.
Certification	To obtain confirmation of which identity (Identity) the entity (Person) in front of the terminal is linked to that was recognized by the service. Used to identify users. Specification: SP 800-63 B
Public Key Infrastructure (PKI)	Also called the public key infrastructure. A mechanism for ensuring the relationship between a public key and its owner. It is constructed by the certificate authority publishing root certificates and intermediate CA certificates and issuing digital certificates.
Certificate Authority (CA)	Organizations that issue electronic certificates (Certification). Registration, issuance, and lapse are created. You have a public public certificate authority that is published and a private certificate authority that operates on its own. There are other certificate authorities at various levels; use the certificate authority that meets your objectives and costs.
Quality Assurance (QA)	Activities to ensure that facilities and data are audited at GxP sites and that studies are conducted appropriately
Interoperability	Compatible and interoperable between various systems. Standardized specifications can be used to operationalize (use) each other.

Term	Description
Public Key Cryptography	A cryptographic method that uses a different key for encryption and decryption or signature and validation. In the signature, two keys, a signing key (also called a private key) and a public key, are used as pairs. The signing key is used for signing and the public key is used for validation.
XML Signature	XML (eXtensible Markup Language) digital signature format. Excellent readability in text format. Specification: W3C Recommendations
Long-Term Signature	A digital signature system that can deal with the expiration of digital certificates and compromise of cryptography. It is suitable for long-term archiving and utilizes a time stamp.
URI Uniform Resource Identifier	Specification to indicate the location of a file. Contains URLs (Uniform Resource Locator) that indicate the file paths and web-based locations in the system.

Revision History

Date	Version Number	Details of revision
2021.05.19	1.0	First Edition
2023.3.01	1.1	English Edition

Members of the R&D Data Arichiving Committee Technical Guidebook (abbreviated as respectful)

Akira Ishii	Ricoh Japan Co., Ltd.
Sayuri Uehara	Astellas Pharma Inc.
Akira Yamazaki	Kyowa Kirin Co., Ltd.
kiyotaka Shiba	Otsuka Pharmaceutical Factory, Inc.
Satoshi Morino	Eisai Co., Ltd.
Naoto Miyaji	Lang Edge Inc.
Harue Ono	Asahi Kasei Pharma Corporation
Kozo Ogimoto	Shimazu Access, Inc.
Miyo Suzuki	Seikagaku Corporation
Sachio Takeda	R&D Data Arichiving Committee
Yoshiaki Hiraishi	Terumo Corporation
Rika Hiraki	Waters Corporation
Junpei Miura	Business Engineering Corp..
Yoshiaki Yamada	Hitachi High-Tech Science Corporation
Michihiro Kimura	Japan Image and Information Management Association