Long-term archiving guidance for analytical instrument data

Version 2.1

January 31, 2020.

Public Interest Incorporated Association

Japan Image and Information Management Association

R&D Data Archiving Committee

Table of Contents

1. Objectives 3
2. Scope of Application 3
3. Introduction 3
3.1. Overview of Analytical Instrument Data 3
3.2. Problems with Analytical Instrument Data Printed on Paper 4
3.3. Concerns on Analytical Instrument Data 5
3.3.1. Re-Processing 5
3.3.2. Long-Term Archiving 5
3.3.3. Data Integrity 7
3.3.4. Data Format from Analytical Instrument 8
4. Policy 9
4.1. Approach to Processing Results 10
4.2. Approach to Long-Term Archiving 10
4.3. Approaches to Data Integrity 11
5. Recommended Long Term Archiving Method for Analytical Instrument Data and Processing Reports 12
5.1. Procedure for Long-Term Archiving of Analytical Instrument Data 12
5.2. Method for Long-Term Archiving of Processing Reports 15
6. Future Challenges 16
6.1. Standardization of Data from Various Analytical Instruments 16
6.2. Development of Data Package Technology 16
7. Future Outlook 16
8. Glossary 17
9. Revision History 18
Appendix 1. An Example of AIA Format Data 19
Appedix 2. Detailed Description of Table 2. 23

1. Objectives

The purpose of this guidance is to present a method for long-term and reliable archiving and management of electronic data (hereinafter referred to as 'analytical instrument data') output from various analytical instrument at research facilities, laboratories, etc. in the pharmaceutical industry, assuming that the data may be re-processed.

2. Scope of Application

This guidance covers analytical instrument data handled in the pharmaceutical industry for the following purposes:

a. Data used for submission and reporting to regulatory authority
b. Data required to be archived under GxP regulations[1]
c. Data thought to need to be archived at the facilities

It is assumed that the analytical instrument data may be re-processed. If data output from analytical instruments such as a balance, pH meter is determined in its as is and not re-processed, it is not included in the subject of this guidance.

3. Introduction

With rapid changes in the IT environment, the amount of electronic records that must be archived is rapidly increasing. On the other hand, unlike paper-printed records, electronic records are feared to become unreadable in the future, such as through migrating to a new system and software upgrade.

In the case of outsourced studies, when transferring analytical instrument data between laboratories / facilities, paper-printed data can be easily transferred. But in the case of analytical instrument data itself, as there are no clear standards on how to deliver the data with authenticity[2] and how to archive the transferred analytical instrument data, the facilities are forced to respond in an optimal procedure that they considered themselves.

In the pharmaceutical industry, this problem is more serious because analytical instrument data must be archived for a long period of time in a state where the data is reliable and can re-processed (see Section 3.3).

The premises and background that led to the development of this guidance are provided below.

3.1. Overview of Analytical Instrument Data

Analytical data and processing results as well as analytical conditions, processing conditions, sample

---

[1] Any of the standards established by the regulatory authority to ensure the safety of patients and the reliability of the studies. Representative laws of the pharmaceutical industry are exemplified, but not limited to:
GLP: Good Laboratory Practice
GCP: Good Clinical Practice
GMP: Good Manufacturing Practice
GVP: Good Vigilance Practice

[2] Reference: Pharmaceutical and Food Safety Bureau Notification No. 0401022 dated April 1, 2005: Use of Electromagnetic Records and Electronic Signatures in Applications for Drug Approval or Licensing, etc. (ERES Guidelines), 3.1.1. Authenticity of Electromagnetic Records

sequences, electronic signatures, audit trails and logs of the system are stored in the analytical instrument or the computer connected to the analytical instrument. Recently, it is often stored in a data storage server, a data management system, or the like via a network.

In this guidance, the following terms are used for the analytical instrument data.

There are two storage formats of analytical instrument data, A., B. below, and the analytical instrument data includes three types of data, a., b., and c. below.

A. Original data

Electronic data output from the analytical instrument. It is also called the original record and is part of raw data under GxP regulations. It is often output in an instrument-specific format. Depending on the instrument, it may or may not include derived data and metadata other than analytical data.

B. Standard format data

Analytical data, some derived and metadata which are converted to a standard format such as AIA. The remaining derived data and metadata are missing without being converting, but by storing the missing data together with materials to supplement the missing data, they can be considered equivalent to the data before conversion.

a. Analytical data

Data obtained from analytical instruments. Chromatogram data such as high-performance liquid chromatograph (HPLC), spectrum data such as infrared spectrophotometer (IR), nuclear magnetic resonance equipment (NMR) and mass spectrometer (MS), and weighing data of balance, etc.

b. Derived data

Integrated results (e.g. peak area) and calculated results (e.g., concentration), etc. using original data.

c. Metadata

Information of units etc. supplemented to clarify the meaning of the data, the date and time of data acquisition, information identifying analytical instrument, analytical conditions, audit trails, etc. The FDA specifies metadata in its data integrity guidance.[3] In this guidance, data related to analytical instrument data and audit trails are presented separately, referring to the FDA's idea. That is, analytical conditions, processing conditions, sample sequences, and the like are denoted as analytical metadata, and the audit trail is denoted as audit trail metadata.

3.2. Problems with Analytical Instrument Data Printed on Paper

Traditionally, in the Japanese pharmaceutical industry, it has been common practice to print and store analytical instrument data and processing reports on paper. However, in order to solve the following problems and ensure data integrity, it has recently been required to store electronic data.

a. The printed data cannot be re-processed.

b. All data that are not normally printed (processing conditions, analytical conditions, etc.) and data that

[3] Source: Data Integrity and Compliance With Drug CGMP Questions and Answers Guidance for Industry (FDA, Dec. 2018)

cannot be printed (three-dimensional data, etc.) are lost, even if they are included in the electronic data.

c. Large space to store paper is required.

3.3. Concerns on Analytical Instrument Data

3.3.1. Re-Processing

It may be required in the pharmaceutical field, to re-process the analytical instrument data on several years later from the actual acquisition. One example is the case of re-processing of the related substances test. It may be confirmed how much the related substance focused on the specified lot were included in the past several lots. In this case, it may be necessary to re-process the past analytical instrument data to investigate unreported micro peaks.

At this time, even if there is properly stored analytical instrument data and an environment that can be re-processed can be prepared it cannot be guaranteed that the same result can be obtained by the differences in the algorithms, except the same manufacturer's software is used.

Further, even if the software of the same manufacturer is used, when the version is different, the same result as the original analysis result is not always obtained.

The reasons are as follows;

a. The algorithm is changed. (Including improvement of calculation precision)
b. The computer is using the binary (not the decimal), and limited digit number for the calculation. So, the last digit contains an error. The calculation results also contain error.
c. The computer (and operating system) is now moving from 32 bit to 64 bit. It may cause difference of the results.

To avoid these, it is necessary to store not only the analysis software but also the computer suite including the operating system, for the re-processing environment. But it is not realistic.

3.3.2. Long-Term Archiving

Considering the need for re-processing, it is desirable for the pharmaceutical industry to store analytical instrument data in a state that it can be re-processed for 30 years.[4] However, if hardware updates or software upgrade are performed, the processing software may not be able to accommodate new OS and analytical instrument data stored over a long period of time may not be available. In addition, when support for the database software being used ends, it may not be possible to migrate the analytical instrument data to the system of another manufacturer.

Representative measures and their respective concerns are described below.

A) Physical storage of processing computer

The relevant concerns are listed below:

a. Increase in managed systems
b. Mechanical lifetime (computer body, printer, external storage)

[4] Examples: Article 20, Issue 3, Article 30, etc. of the ministerial ordinances on GMP.

c. Maintenance expenses (including licensing expenses)
d. Storage place
e. Periodic check
f. Maintain backup of electronic data
g. Securing operators who can operate old systems

B) Virtualize and store the processing computers

The relevant concerns are listed below:

a. Increase in managed systems
b. Maintenance expenses (including licensing expenses)
c. Periodic check
d. Maintain backup of electronic data
e. Securing operators who can operate old systems
f. Measures to be taken when additional hardware is required, such as hardware keys[5], interface cards[6], etc.
g. Licensing form of the software subject to virtualization[7]
h. To be fully compatible

C) Migration from current to succeeding (or compatible) systems

The relevant concerns are listed below:

a. Restrictions on migratable systems
   It is difficult to know the constraints when introducing the system.
b. Constraints on the migratable range
   For example, audit trails are not migratable.

D) Data Migration via Standard Formats.

The relevant concerns are listed below:

a. Regardless of automatic or manual, the ability to export data and metadata is essential for a migration source system, and the ability to import them is essential for a migration destination system.
b. Loss of links to metadata or difficulty of migration of links.
c. It is difficult to keep data integrity due to the difference in data management level between the migration source system and the migration destination system. For example, no audit trail is recorded other than the data generated by the system.

In this guidance, we recommend “D) Data Migration via Standard Format”. Reasons for recommendation are discussed in Chapter 4.

[5] Hardware (Dongle) such as the USB device used for use control such as software to prevent fraudulent access and fraudulent copying.
[6] An extended card adding an input-output interface to a computer.
[7] The hardware rather than users may need licensing (e.g., Microsoft Windows DSP version, OEM version).

3.3.3. Data Integrity

On the premise of re-processing, it is important that the reliability should be maintained when the analytical instrument data is stored or the storage location is moved. Specifically, for the analytical instrument data used for application and reporting to the regulatory authority, the data integrity has come to be required. To ensure this data integrity, it is required to store not only analytical data but also derived data and metadata (analytical and audit trail) under security management while maintaining legibility, but there are difficulties in meeting the requirements, including the concerns of 3.3.1 and 3.3.2.

Excerpts of the guidance with regulatory requirements are provided below.

Regulatory Requirements for the Storage of Electronic Data.

The FDA has published the following Q&A in the guidance of "Questions and Answers on Current Good Manufacturing Practices, Good Guidance Practices, –Records and Reports."[8] The answers were extracted only from the relevant parts and arranged for easy understanding.

Q) How do the part 11 regulations and "predicate rule requirements" (in 21 CFR part 211) apply to the electronic records created by computerized laboratory systems and the associated printed chromatograms that are used in drug manufacturing and testing?

A) The printed chromatogram would not be considered an exact and complete copy of the electronic raw data used to create the chromatogram, as required by § 211.68. The chromatogram does not generally include, for example, the injection sequence, instrument method, integration method, or the audit trail, of which all were used to create the chromatogram or are associated with its validity. Therefore,

a. The printed chromatograms used in drug manufacturing and testing do not satisfy the predicate rule requirements in part 211.

b. The electronic records created by the computerized laboratory systems must be maintained under these requirements.

The FDA has also issued the following Warning Letter.[9]

a. Your firm deleted all electronic raw data supporting your HPLC testing. In addition, your firm failed to retain basic chromatographic information such as injection sequence, instrument method or integration method for the tests.

b. Your systems allowed operators to delete files. You had no access control procedure to control this practice.

[8] Source: https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm124787.htm#3

[9] Source: https://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2014/ucm404316.htm Document number 320-14-10
https://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2017/ucm546483.htm Document No. 320-17-29

FDA guidance[3] requires ALCOA for data integrity, and PIC/S guidance[10] requires CCEA as well. It is necessary to establish the ways to satisfy the requirements of the ALCOA+ prepared by integrating both of them so that it can deal with regulatory inspections (e.g., document-based compliance review) in response to the expectations of the regulatory authorities in Japan, the United States, and Europe.

ALCOA

Attributable: Signing and sealing, etc. identify the attribution and responsibility of the data.
Legible: Data is recorded concisely and clearly so that it can be easily read.
Contemporaneous: Data is recorded at the time the work is performed and without delay.
Original: Records are maintained as the original data.
Accurate: Data and records are accurate and objective.

CCEA

Complete: Records have no data unavailability and are complete.
Consistent: Records (data) are reasonable and consistent.
Enduring: Record retention is durable.
Available: Data can be retrieved as needed.

MHRA guidance[11] define "True copy" as follows in section 6.11.2:

a. *A true copy may be stored in a different electronic file format to the original record if required, but must retain the metadata and audit trails required to ensure that full meaning of the data are kept and its history can be reconstructed.*
b. *Original records and true copies must preserve the integrity of the record. True copies of original records may be retained in place of the original record (e.g. scan of paper record), if a documented system is in place to verify and record the integrity of the copy. Organisations should consider any risk associated with the destruction of original records.*
c. *It should be possible to create a true copy of electronic data, including relevant metadata, for the purpose of review, backup and archival. Accurate and complete copies for certification of the copy should include the meaning of the data (e.g. date formats, context, layout, electronic signatures and authorizations) and the full GxP audit trail. Consideration should be given to the dynamic functionality of a 'true copy' throughout the retention period (see 'archive').*

3.3.4. Data Format from Analytical Instrument

Analytical instruments generate various types of data format. It is difficult to transfer electronic data which is created and stored in one manufacturer's analytical instrument to the other manufacturer's analytical instrument without any data format conversion.

[10] Source: PIC/S Guidance Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments (PIC/S, PI041-1(Draft 3) 30 Nov. 2018)
[11] Source: 'GXP' Data Integrity Guidance and Definitions (MHRA, Revision1 March 2018).

Therefore, standard formats independent on the manufacturer have been considered. This guidance uses a chromatographic data as an example to explain the standard formats. As the standard format of chromatographic data, the following formats have been proposed, but each has merits and demerits. So, it has not been unified yet.

a. AIA (Analytical Instrument Association)
   Standard format for chromatographic data, such as HPLC, defined by the US Analytical Instrument Association-AIA. Different versions may be incompatible.
b. NetCDF (Network Common Data Form)
   The NetCDF supports a machine-independent format for representing scientific data, as developed by Unidata, part of the University Corporation for Atmospheric Research-UCAR Community Programs (UCP).
c. JCAMP (Joint Committee on Atomic and Molecular Physical Data)
   Standard format for spectral data such as NMR, IR, and MS taken over by the International Union of Pure and Applied Chemistry-IUPAC.
d. AnIML (Analytical Information Markup Language)
   XML standard formats, as defined by American Society for Testing Materials. XML format not yet finalized due to drawbacks that the data size becomes 20 times bigger than the original one and the writing speed is slow even though the data acquiring speed from analytical instrument is fast.[12]

It should be noted that neither of the above formats includes the all analytical metadata.

Chromatogram data and spectral data can be converted to standard formats, but analytical metadata such as analytical conditions and processing conditions have different formats for each analytical instrument of different manufacturer. Also, since the processing algorithm is different for each manufacturer, it is not possible to completely reproduce the data from other manufacturers even if the processing parameters can be transferred. However, in the case of quantitative analysis, equivalent quantitative results can be obtained with other manufacturers' instrument when the relative values against the reference standards are the almost same. An example of such quantitative analysis is concentration calculation using a calibration curve created using the reference standards.

In addition, metadata imported into standard formats varies from manufacturer to manufacturer. Therefore, there may be cases where data is not included in the fields specified as the standard format, or the data format related to unit, and time is different.

4. Policy

This chapter provides the concept behind the recommended methods for long-term archiving of analytical

[12] The XML format may improve the communication speed by improving the algorithm, but the modification of the algorithm is difficult because it had been developed by using machine language.

instrument data described in Chapter 5.

Table 1 provides a summary of the response policies to the concerns listed in Section 3.3.

Table 1. Response policy to address concerns (Summary)

	Concern	Response policy
1	At the time of re-processing, the processing algorithms are different for each manufacturer, so it is not possible to completely reproduce the data of other manufacturers even if the processing parameters have been migrated. (3.3.1)	Show in Section 4.1
2	At the time of re-processing, even with the same manufacturer, the same results as the original processing results may not be obtained due to differences in the software versions. (3.3.1)	Show in Section 4.1
3	When migrating data for long-term archiving, the link between original data and metadata is lost or migration of the link is difficult. (3.3.2)	Show in Section 4.2
4	When migrating data for long-term archiving, there is a limitation on the data management level in the system to which the migration is made and maintaining data integrity is difficult. For example, no audit trails are recorded other than the data generated by the migration system. (3.3.2)	Show in Section 4.3

4.1. Approach to Processing Results

As described in 3.3.1, analytical instrument data do not necessarily produce the same results as the original processing results. Even if the original data can be stored completely, it is permissible to include subtle differences depending on the available computer and software at the time of processing. Because the difference is small, it is believed that it does not substantially affect the purpose of the test in many cases.

Therefore, it is necessary not to raise the requirement level more than necessary by clarifying what and how much to seek at the time of data long-term archiving. Possible requirement levels include the following:

a. The re-processing results need to be consistent with the original processing results.
b. If re-processing is possible, differences in processing results due to algorithm changes and other reasons are tolerated.
c. It is good if it can be reviewed from the previous and another viewpoint. For example, it is necessary to enlarge the area where the presence of a related substance is of concern and confirm the peak.

4.2. Approach to Long-Term Archiving

This guidance envisages 10 to 30 years as a period in which analytical instrument data is expected to be archived in a re-processed state. For this reason, the following methods are proposed in this guidance (Fig. 1)

a. Analytical instrument data, such as original and derived data, as well as standard format data, analytical metadata including analytical conditions, processing conditions, and sample schedules, and audit trail metadata, are exported and packaged from analytical instruments (current systems).

b. This package is archived and maintained in storage, including data storage server, data management system.

c. When a re-processing or the like is required, this package is imported into a new analysis environment (successor system) for use.

As a result, it is possible that the analytical instrument data can be independent from the analytical instrument to promote fluidization, and a contract service that provides a processing environment can be realized (Reference 7).

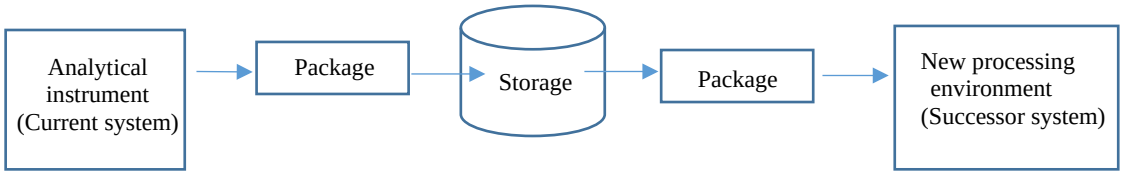


Figure 1 Packaging of Data

4.3. Approaches to Data Integrity

Data integrity is closely related to original data, derived data, standard format data, analytical metadata, audit trail metadata, data processes, OS/hardware (Fig. 2).

Exported and packaged analytical instrument data includes only analytical instrument data up to the point where the export and packaging task is performed.

In this guidance, it is recommended to export analytical instrument data on condition that the auditor confirms that data integrity is ensured under the specified conditions. The export and packaging task are performed at various times, for example, immediately after the completion of analysis and processing or when upgrading software.

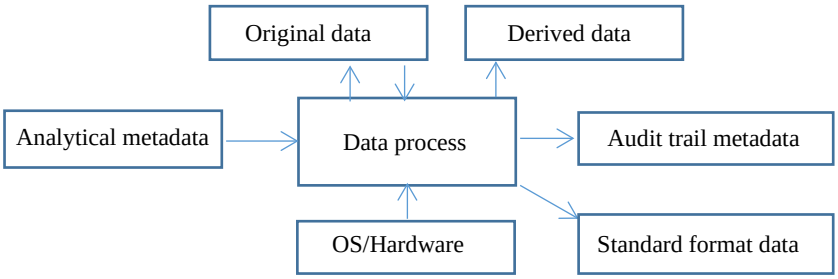


Figure 2 Elements related to data integrity

After exporting analytical instrument data, data integrity is conditionally ensured. However, if the third party (QA, etc.) guarantees the appropriateness of the process from exporting the data to importing it into a

new system, it is considered that the reliability of the data can be ensured even if the data export and import are not automated.

5. Recommended Long Term Archiving Method for Analytical Instrument Data and Processing Reports

It is the fundamental approach on the long-term archiving of analytical instrument data in this guidance that is to collect analytical instrument data necessary for ensuring reproducibility into one package and to prepare a mechanism for preventing modification of this package.

At the analytical, a processing report that summarizes the processing results in addition to the analytical instrument data is also created, and the processing report is required to be archived for a long time together with the analytical instrument data. The analytical instrument data must be kept intact for a long period of time, but the processing report may be corrected at a later date. The long-term archiving method of the analytical instrument data and the long-term archiving method of the processing report describes separately.

There are various types of analytical instrument data. In this guidance, chromatographic data was taken up as representative examples of analytical instrument data and AIA format was taken up as its standard format, and methods for long-term archiving of analytical instrument data and processing reports were examined.[13]

5.1. Procedure for Long-Term Archiving of Analytical Instrument Data

The AIA format for chromatographic data output from high-performance liquid chromatographs (HPLC) or the other instrument does not include all derived data and metadata other than chromatographic data. An example of AIA format data is shown in the appendix. On the other hand, the sample sequence is important information and it can be output as a PDF file, but it is not associated with standard format (AIA) data file. In order to clarify the relationship between these data, it is effective to archive multiple files as a unit.

In the case of HPLC measurement, samples prepared for other purposes may be analyzed together with samples for the main purpose. If the necessary data is extracted when archiving a series of data in the above situation, it is suspected to be arbitrary extraction. Therefore, all data including those measured for other purposes should be archived.

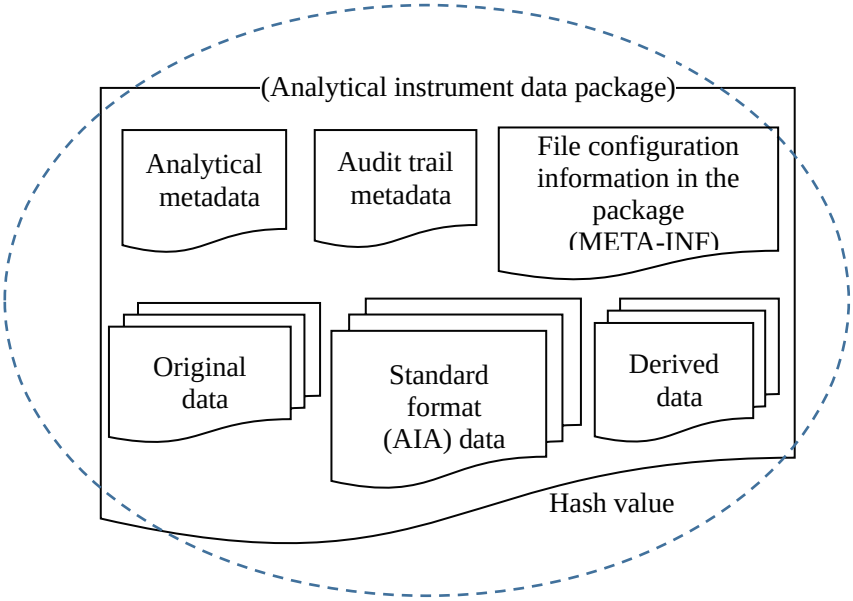
An image of the analytical instrument data package is shown in Fig. 3. The analytical instrument data package includes original data and standard format (AIA) data that can be archived for a long time while maintaining the dynamic state of the analytical data. By adding derived data, analytical metadata, audit trail metadata, etc., the authenticity of the analytical instrument data is ensured.

[13] In the HPLC of Nippon Waters Co., Ltd., Shimadzu Corporation, and Hitachi High-Tech Science Co., Ltd., it has been confirmed that waveforms can be reproduced by reading AIA format data created by other companies and handled by in-house analysis software. (Hitachi High-Tech Science Co., Ltd.'s HPLC does not have an AIA format reading function, so only writing was confirmed.)

It is recommended that the ZIP package be used as a package for these multiple files.[14]

The ZIP package stores a special directory (META-INF) describing file configuration information in the package in addition to original data, standard format (AIA) data, derived data, and metadata. This META-INF will be identified by directory. The META-INF can also be managed outside the packages. In this case, the link is made with the hash value of the package. Appropriate techniques such as time stamps or electronic signatures are used to generate hash value. When a time stamp is used, it is necessary to establish the method for assuring the effectiveness after the expiration of the certificate of the time stamp station. One method is to publish in newspapers, journals, etc., and use the National Diet Library to ensure the readability.

In the case of re-processing, instead of directly manipulating the packaged data, the data is copied, and re-processing is performed using this copy to eliminate the risk of erroneous erasure of original data or erroneous overwriting.



Dashed lines are the range of hash operations. Hash values are managed separately from the package.

Figure 3 Packaging of analytical instrument data

The requirements for data storage vary depending on the purposes of use, the status of use, the level of reliability required, and the retention period. The purposes of use include reproducibility of results, compliance with document-based compliance review, compliance with the principles of GxP. Table 2 shows an example of

[14] Other package candidates include the ZIP-based purposive package, PDF, and XML constructs. ZIP-based packages for different purposes already exist in various formats (e.g., OPC for office documents, EPUB for electronic books, ASiC (Associated Signature Container, ETSI TS 102 918) being standardized in Europe, etc.). These can also be regarded as only different META-INF definitions in the package. In the future, it will also be candidates to define a META-INF dedicated to analytical instrument data (in this instance to register a new extension corresponding to it).

analytical instrument data to be stored in the package. (Refer Appendix 2 for a detailed description)

Table 2. Example of selection of analytical instrument data for each purpose of use

Analytical instrument data		Re-processing only	Document-based compliance review	GxP regulations compliance
Original data (instrument specific format data)		○	○	○
Standard format (AIA) data		○	○	○
Derived data	Results (e.g., peak area) processed using original data etc.	－	○	○
	Calculated results (concentration) etc.	－	○	○
Analytical metadata	Sample sequence	○	○	○
	Instrument parameters	○	○	○
	Processing parameters	○	○	○
Audit trail metadata	Original data-related	－	○	○
	Processing results-related	－	－	○
	System-related	－	－	－

○: storage　－: No storage required

Instead of including the analysis metadata in each analytical instrument data package, it is possible to configure the meta-informed package as shown in Fig. 4 and add the analytical metadata from the outside.

In addition, the link information indicating the location of the analytical instrument data package that does not include analytical metadata and its hash value are stored in “Link information and their hash value of analytical instrument data” in the meta-informed package.

This makes it possible to refer to electronic data that has already been packaged for other purposes and to reduce the volume of electronic data to be stored.

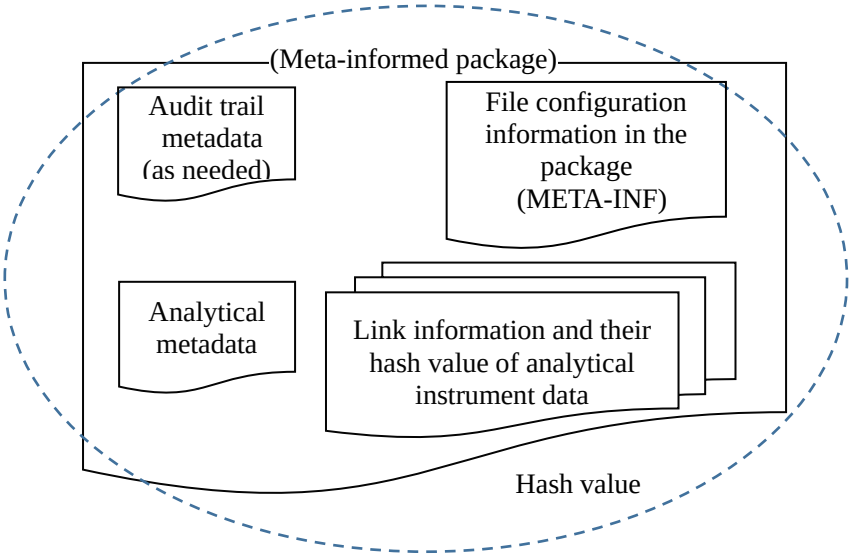


Figure 4. Packaging of analytical metadata (Meta-informed package)

5.2. Method for Long-Term Archiving of Processing Reports

The processing report may also be modified by changing the interpretation of the analytical instrument data, re-processing the data from different viewpoint later, or comparing it with other processing results. To ensure its authenticity, the processing report can be finalized, and its hash value can be recorded, after finalizing the test report and the like.

"Link information and their hash value of analytical instrument data" thereof in the processing report package stores the link information indicating the location of the analytical instrument data package and the hash value thereof. Rather than creating a file of "Link Information and their hash value of analytical instrument data", the analytical instrument data package itself may be imported (Fig. 5).

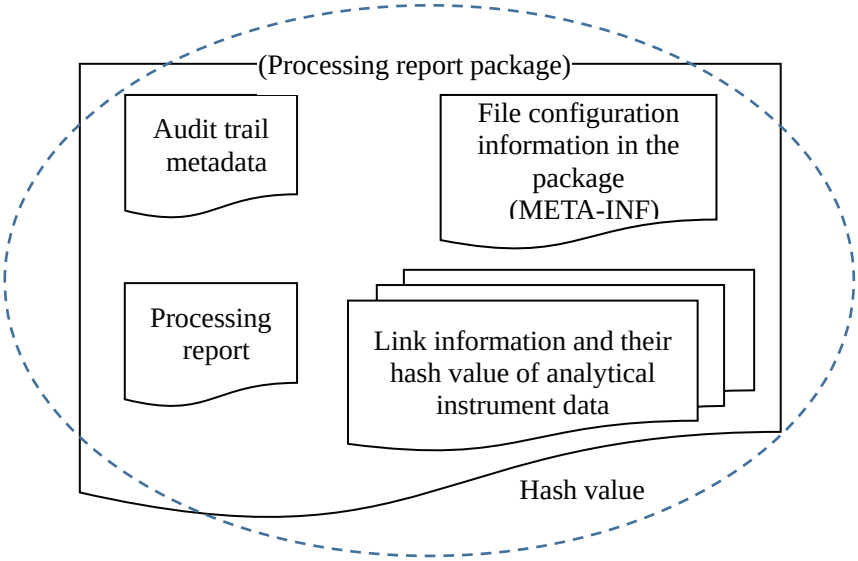


Figure 5. Packaging of processing report (processing report package)

6. Future Challenges

In order to realize the method recommended in Chapter 5, the following techniques are expected to be developed.

6.1. Standardization of Data from Various Analytical Instruments

It is desirable that the long-term archiving method of the analytical instrument data as shown in 5.1., can be applied to import and use data of analytical instruments from other manufacturers. This requires manufacturers to provide compatible data formats. At present, there are a lot of analytical instruments that don't have standard data formats including those shown in 3.3.4.

New data formats are developed along with new technologies. For these data formats, it is expected to provide model-independent software, output in a form compatible with existing instruments, disclose format definitions, etc.

6.2. Development of Data Package Technology

In order to ensure the long-term archiving method of the recommended analytical instrument data and processing reports described in Section 5, the following techniques are expected to be developed.

a. Technology for automatically packaging analytical instrument data and processing reports
b. Technology for capturing and utilizing packaged data, reports

7. Future Outlook

A highly improved accuracy of analytical instrument and its increased capacity of electronic data acquired in a single analysis have resulted in an increase in the storage capacity of electronic data. In addition, a situation has been occurred that the recording capacity of the electric data storage, such as data storage server and data management system, is occupied by the data stored but hardly utilized. As a result of these, it is becoming a problem that cost increase related in storage of electronic data in the pharmaceutical industry.

There is a need for services in the pharmaceutical industry to contract the archiving and management of electronic data, as all companies are concerned about the storage and management of electronic data (see Section 4.2). Furthermore, a service that stores and manages the processing software of all manufacturers and provides an environment in which any manufacturer's analytical instrument data can be processing would be useful to many companies in the pharmaceutical industry.

There is also a need to provide and receive analytical instrument data and to migrate data between systems, such as through company acquisitions and alliances of relevant organizations. If services are available to retain older versions of OS and processing software on virtual environments, for many companies, there is no bother problems such as migrating past analytical instrument data to new systems.

The above responses seem to be technically implementable. However, in order for such a service to be provided as a business, the number of companies that need it and the service value paid by the company need to be large enough as a business. The investigation of the feasibility, including the survey of needs is a future problem.

8. Glossary

Table 3 provides a description of the terms used in this guidance.

Table 3. Terminology Description

Term	Description
Algorithm	A set of well-defined finite rules that solves the problem by applying finite times
Importing	Transfer of data from other analytical instruments
Exporting	Transfer of data to other analytical instruments
Processing parameters	Values set for the processing method of the peak waveform, etc. It also includes baseline setting methods, denoising methods, etc.
Audit trail	A secure, computer-generated, time-stamped electronic record that allows reconstruction of the process of events associated with the creation, modification or removal of an electronic record. There are logs of original data generation, audit trails of analysis, and audit trails on the system.
QA	Quality Assurance, Confirmation work to certify quality
Sample sequence	Injection schedule
ZIP	An archival file format that handles multiple files together as a single file. Basically, file extension "zip" is used, but there are also files which extension is not zip, such as docx.
Instrument parameters	The set-points, such as operating conditions of the analytical instrument, as well as the type of detector and the type of solvent, are included.
Time-stamping	・Date and time automatically imprinted by the computer (ERES Guidelines)[15] ・Technique for providing information that can detect a change and verifying whether there has been a change since that time[16]
Data integrity	Completeness, consistency and accuracy of data
Data process	Process of processing the data
Electronic signature	Signatures that are signed electronically as equivalent to handwritten signatures or seals and are constructed electronically with a set of symbols prepared, adopted, confirmed, and approved by individuals or corporations (ERES Guidelines)[15]
File configuration information in the package	A special file describing file configuration information in the package
Hash value	Message digest generated by hash function
PDF	Portable Document Format, File-format for the exchange of documents, including texts, images, and graphics, specified as ISO 32000
Bit	Smallest unit of data handled by the computer

[15] Source: Pharmaceutical and Food Safety Bureau Notification No. 0401022 dated April 1, 2005: Use of Electromagnetic Records and Electronic Signatures in Applications for Drug Approvals or Licenses, etc. (ERES Guidelines)

[16] Time Business Forum (https://www.dekyo.or.jp/tbf/index.html) Time business lexicon

9. Revision History

Date	Version Number	Detail of revision
2020.01.31	2.1	Revised content in preparation for English translation
2019.04.18	2.0	Change the title from Guideline (draft) to Guidance Reflect the solicited comments
2018.11.28	1.0	

Appendix 1. An Example of AIA Format Data

```
Netcdf demodata {
Dimensions:
        _2_byte_string = 2 ;
        _4_byte_string = 4 ;
        _8_byte_string = 8 ;
        _12_byte_string = 12 ;
        _16_byte_string = 16 ;
        _32_byte_string = 32 ;
        _64_byte_string = 64 ;
        _128_byte_string = 128 ;
        _255_byte_string = 255 ;
        Point_number = 1201 ;
        Peak_number = 4 ;
        Error_number = 1 ;
Variables:
        Float detector_maximum_value ;
        Float detector_minimum_value ;
        Float actual_run_time_length ;
        Float actual_sampling_interval ;
        Float actual_delay_time ;
        Float ordinate_values(point_number) ;
        Ordinate_values:uniform_sampling_flag = "Y" ;
        Ordinate_values:autosampler_position = "" ;
        Float peak_retention_time(peak_number) ;
        Float peak_amount(peak_number) ;
        Float peak_start_time(peak_number) ;
        Float peak_end_time(peak_number) ;
        Float peak_width(peak_number) ;
        Float peak_area(peak_number) ;
        Float peak_area_percent(peak_number) ;
        Float peak_height(peak_number) ;
        Float peak_height_percent(peak_number) ;
        Float baseline_start_time(peak_number) ;
        Float baseline_start_value(peak_number) ;
        Float baseline_stop_time(peak_number) ;
        Float baseline_stop_value(peak_number) ;
```

Float retention_index(peak_number) ;

Float migration_time(peak_number) ;

Float peak_asymmetry(peak_number) ;

Float peak_efficiency(peak_number) ;

Float mass_on_column(peak_number) ;

Char peak_name(peak_number, _128_byte_string) ;

Char peak_amount_unit(peak_number, _128_byte_string) ;

Char peak_start_detection_code(peak_number, _128_byte_string) ;

Char peak_stop_detection_code(peak_number, _128_byte_string) ;

Char manually_reintegrated_peaks(peak_number, _128_byte_string) ;

// global attributes:

:dataset_completeness = "C1+C2" ;

:aia_template_revision = "1.0.1" ;

:netcdf_revision = "VERSION of Aug 26 2015 17:18:35 $" ;

:languages = "English" ;

:administrative_comments = "" ;

:dataset_origin = "Supplier" ;

:dataset_owner = "" ;

:dataset_date_time_stamp = "20030116164231+0900" ;

:injection_date_time_stamp = "20030116164231+0900" ;

:experiment_title = "" ;

:operator_name = "Operator" ;

:separation_experiment_type = "" ;

:company_method_name = "" ;

:company_method_id = "" ;

:pre_experiment_program_name = "" ;

:post_experiment_program_name = "" ;

:source_file_reference =

"C:¥¥Data¥¥Project1¥¥Demo_Data-001.dat" ;

:sample_id_comments = "This is comment."/;//Data File Comment

Sample_id = "1";//Sample ID

::sample_name = "STD";//Sample name

::sample_type ="Standard";//sample type

:sample_injection_volume = 10.f ;

:sample_amount = 1.f ;

:detection_method_table_name = "" ;

:detection_method_comments = "This is comment.";// Comment in Data File.

:detection_method_name =

"C:¥¥Sample¥¥LC¥¥Demo_Method.met" ;

: detector_name = "detector A";

Units of::detector_unit ="Volts";//detector

:raw_data_table_name =

"C:¥¥Data¥¥Project1¥¥Demo_Data-001.dat" ;

::retention_unit ="Seconds";//time units

:peak_processing_results_table_name = "" ;

:peak_processing_results_comments = "This is comment.";// Comment in Data File.

:peak_processing_method_name =

"C:¥¥Sample¥¥LC¥¥Demo_Method.met" ;

:peak_processing_date_time_stamp = "" ;

Data:

Detector_maximum_value = 0.01606795 ;

Detector_minimum_value = -0.002146373 ;

Actual_run_time_length = 600 ;

Actual_sampling_interval = 0.5 ;

Actual_delay_time = 0 ;

Ordinate_values = -3.786087e-006, -3.786087e-006, 6.341934e-007 ,

2.627373e-006, 4.763603e-006, 6.165505e-006, 7.691383e-006 ,

.....

5.438328e-005, 5.547047e-005, 5.688667e-005, 5.883217e-005 ,

6.048202e-005, 6.172657e-005, 6.249905e-005, 6.396294e-005 ,

6.585121e-005, 6.725788e-005, 6.734371e-005, 6.756782e-005 ,

6.769657e-005, 6.738186e-005, 6.752014e-005, 6.912708e-005 ,

7.059574e-005 ;

Peak_retention_time = 159.271, 194.451, 237.258, 278.901 ;

Peak_amount = 1, 1, 1, 1 ;

Peak_start_time = 151.5, 180.5, 229.5, 271.5 ;

Peak_end_time = 169.5, 204, 246, 287.5 ;

Peak_width = 18, 23.5, 16.5, 16 ;

Peak_area = 73509.84, 83363.7, 53565.35, 57335.54 ;

Peak_area_percent = 27.45215, 31.13206, 20.00391, 21.41188 ;

Peak_height = 13039.91, 15250.72, 11457.2, 13248.39 ;

Peak_height_percent = 24.60536, 28.77699, 21.6189, 24.99875 ;

Baseline_start_time = 151.5, 180.5, 229.5, 271.5 ;

Baseline_start_value = 4.217148e-005, 0.0005049753, 0.00154314 ,0.002390246 ;

Baseline_stop_time = 169.5, 204, 246, 287.5 ;

Baseline_stop_value = 0.0003286457, 0.001030946, 0.00189703, 0.0025459 ;

Retention_index = 0, 0, 0, 0 ;

Migration_time = 0, 0, 0, 0 ;

Peak_asymmetry = 1.193704, 1.139562, 1.168244, 1.165645 ;

Peak_efficiency = 4246.021, 6958.823, 12828.54, 20710.6 ;

Mass_on_column = 0, 0, 0, 0 ;

Peak_name =

"Methyl paraben", "Ethyl paraben", "Propyl paraben", "Butyl paraben" ;

Peak_amount_unit = "", "", "", "" ;

Peak_start_detection_code = " ", " ", " ", " " ;

Peak_stop_detection_code = " ", " ", " ", " " ;

Manually_reintegrated_peaks = "", "", "", "" ;

}

Appedix 2. Detailed Description of Table 2.

The table 2 of this guidance provides examples of selection of analytical instrument data for each purpose of use; however, in actual operation, each company should make decisions according to the situation.

The rationale for the categorization in Table 2 is provided here for reference.

Table 2. (Reshown) Example of selection of analytical instrument data for each purpose of use

Analytical instrument data		Re-processing only	Document-based compliance review	GxP regulations compliance
Original data (instrument specific format data)		○	○	○
Standard format (AIA) data		○	○	○
Derived data	Results (e.g., peak area) processed using original data etc.	－	○	○
	Calculated results (concentration) etc.	－	○	○
Analytical metadata	Sample sequence	○	○	○
	Instrument parameters	○	○	○
	Processing parameters	○	○	○
Audit trail metadata	Original data-related	－	○	○
	Processing results-related	－	－	○
	System-related	－	－	－

○: storage　－: No storage required

Re-processing only

One of the objectives of the company to store analytical instrument data for long periods of time is to enable the possibility of re-processing as needed at a later date. Requirements for re-processing were set as follows and reflected in Table 2.

- It is imperative that electronic data can be imported into the software used for re-processing. Since there is a possibility that the original data cannot be imported, its standard format data is also stored with it.
- Analytical metadata is stored as a source of sample identification and analytical conditions.
- The package is timestamped to prove that it has not been changed before re-processing.

Document-based compliance review

At the time of application for new drug approval, it is investigated whether pharmacological, pharmacokinetic, clinical pharmacokinetic, and quality studies have been collected and prepared in accordance with the “Standards

of Reliability of Application Data (Articles 43, 61, 114-22, 114-42, 137-25, or 137-42 of the Enforcement Regulations of the Pharmaceutical and Medical Devices Act)" in a document-based compliance review. It is natural to store the data from studies conducted in accordance with the above criteria, but there are differences in the long-term archiving methods and concepts among institutions. Table 2 requirements were established based on the following considerations.

- Once inspection by the internal audit department such as QA/QC is completed, the original data, derived data, analytical metadata, and audit trail metadata related to the original data are stored in the package as it is.
- In addition, standard format data is also archived in the package so that dynamic data such as waveforms can be reproduced even after long-term archiving.
- The package will be timestamped to prove no change after in-house audit.

GxP regulations compliance

Although it is imperative to store study materials in accordance with GxP regulations, there are differences in long-term archiving methods and perceptions among institutions. Table 2 requirements were established based on the following considerations.

- Once inspection by the internal audit department such as QA/QC is completed, the original data, derived data, analytical metadata, and audit trail metadata related to the original data are stored in the package as it is.
- In addition, standard format data is also archived in the package so that dynamic data such as waveforms can be reproduced even after long-term archiving.
- The package will be timestamped to prove no change after in-house audit.

In addition, audit trails of systems such as log-in and log-out were not included in the package for the following reasons:

- Although the package proposed in this guidance is made in an analytical run, the system's audit trails are managed throughout the system's use rather than in an analytical run.
- It is because the system's audit trail is used to detect inappropriate activities other than packaged data. (In GxP facilities, inappropriate operations other than data are confirmed through regular audits by internal auditing departments such as QA/QC.)

R&D Data Archiving Committee　Guidance preparation members

Akira Asagarasu	ASKA Pharmaceutical Co., Ltd.
Sayuri Uehara	Astellas Pharma Inc.
Harue Ohno	Asahi Kasei Pharma Corporation
Yasuhiro Ogawa	LSI Medience Corporation
Kozo Ogimoto	Shimadzu Corporation
Michihiro Kimura	Japan Image and Information Management Association
Hiroyoshi Kuroda	Ricoh Company, Ltd.
Kiyotaka Shiba	Otsuka Pharmaceutical Factory, Inc.
Miyo Suzuki	Seikagaku Corporation
Sachio Takeda	Asahi Kasei Pharma Corporation
Yoshiaki Hiraishi	Terumo Corporation
Rika Hiraki	Waters Corporation
Takashi Hosoi	JIP InfoBridge Co., Ltd.
Naoto Miyachi	LangEdge, Inc.
Satoshi Morino	Eisai Co., Ltd.
Akira Yamazaki	Kyowa Kirin Co., Ltd.
Yoshiaki Yamada	Hitachi High-Tech Science Corporation