Long-Term Archiving of Analytical Instrument Data
Data Management Guidebook
(Analytical Instrument Data and Emails)

Version 1.0

February 18, 2026

Public Interest Incorporated Association

Japan Image and Information Management Association

R&D Data Archiving Committee

Table of Contents

1. Purpose

The purpose of this guidebook is to indicate how data to be preserved should be selected when implementing long-term preservation, based on the Long-Term Archiving Guidance for Analytical Instrument Data.

2. Scope of Application

This guidebook applies to Analytical Instrument data and emails handled within the pharmaceutical industry for the following purposes:

a. Data used for regulatory submissions and reporting
b. Data required to be retained under GxP regulations[1]
c. Data deemed necessary for retention at the facility.

Note that Analytical Instrument data is assumed to be subject to potential reprocessing.

3. Introduction

The R&D Data Archiving Committee has compiled its approach for long-term archiving of Analytical Instrument data in a reprocessable format as the Long-Term Archiving Guidance for Analytical Instrument Data and has outlined specific methods in the Technical Guidebook and Operational Guidebook. In this guidebook, the data that should be preserved have been organized with respect to emails and representative analytical instruments.

4. Related Regulations

In the pharmaceutical industry, regulatory authorities require the proper retention of dynamic data through GLP and GMP regulations. These requirements are outlined in OECD GLP No. 22 for GLP and PIC/S Guidance for GMP.

4.1 GLP (OECD GLP No. 22)

No. 22 states in Section 6.14: *Data necessary to fully reconstruct the test activity should be collected and retained. Data should be retained with associated metadata, where applicable.* Section 3.1 of the same document defines that *For electronic data, some metadata is generated*

[1] A collective term for standards established by regulatory authorities to ensure patient safety and the reliability of trials. Representative regulations in the pharmaceutical industry are listed below, though this list is not exhaustive.
GLP (Good Laboratory Practice): "Guidelines for the Conduct of Non-Clinical Studies on the Safety of Pharmaceuticals"
GCP (Good Clinical Practice): "Guidelines for the Conduct of Clinical Trials of Pharmaceuticals"
GMP (Good Manufacturing Practice): "Standards for the Manufacturing and Quality Control of Pharmaceuticals"
GVP (Good Vigilance Practice): "Guidelines for Post-Marketing Surveillance of Pharmaceuticals"

as an audit trail. This implies that the metadata required to be retained under Section 6.14 includes audit trails. Furthermore, Section 6.14's Retention of Dynamic Data stipulates: *"Computerized systems that generate dynamic records must be capable of retaining the dynamic nature of the data*. However, as a response for *cases where a computerized system can no longer be maintained, such as when it becomes unsupported*, the document also states: *Data generated by electronic means may be retained in an acceptable paper or electronic format when the integrity of the raw data is justified as being maintained by static records*.

4.2 GMP（PIC/S Data Integrity Guidance）

Like GLP, the PIC/S Data Integrity (DI) Guidance also emphasizes the retention of dynamic data.

Section 7.7.3 states:

It is important that electronic records are stored in a dynamic format to allow interaction with the data. Data should be stored in a dynamic format if it is essential for data integrity or for subsequent verification.

Section 9.9 further clearly requires:

When storing data, the entire original data set and all associated metadata (including audit trails) should be stored using a secure and validated process.

In addition, Section 9.4 addresses the maintenance of legacy systems and states:

When software for legacy systems[2] *is no longer supported, consideration should be given to maintaining that software (in accordance with storage requirements, for as long as possible) to ensure continued access to the data.*

However, it also acknowledges an exception:

It is understood that migration to a file format lacking certain attributes and/or dynamic data functionality may be unavoidable to ensure continued access, but only when it is technically impossible to migrate while fully preserving the functionality of the original data.

4.3 Retention of Electronic Communications

Section 6.14 of OECD GLP No. 22, Retention of Electronic Communications, identifies electronic communications as an example of records in a dynamic state.

It states that:

Electronic communications are also an example of records in a dynamic state.

Furthermore, it requires that:

Where data is supported by electronic means of communication such as email or electronic messages (e.g., where GLP activities and responsibilities can be verified), processes must be established to ensure the retention and reconciliation of electronic communications (including

[2] The term "legacy system" here refers to the system currently in use.

verification that the records are complete and their integrity has not been compromised).
It also requires that:
Such mechanisms should be designed to maintain the attribution and integrity of the relevant electronic communications, such as by reliably identifying the sender and recipient along with the appropriate date and time. All attachments must maintain their association with the corresponding message and preserve the message chain[3].

4.4 Summary of Requirements under GLP and GMP

In summary, within the pharmaceutical industry where GLP and GMP apply, the general principle is that dynamic data generated from computerized systems should be retained in a dynamic state to the extent possible.
On that basis, consideration should be given to:

- the ability to reconstruct test activities.
- the ability to appropriately assess the impact on product quality.
- the preservation of the integrity of metadata, including audit trails.

Based on these considerations, it is necessary to determine which data should be preserved, in what format they should be retained, and how to address system updates or migrations.

5. Data Management Guide by Type

This chapter organizes which data should be preserved for each type of instrument.
When determining what data should be included in the preservation package, the intended use of the data shall also be considered.
As necessary, refer to Table 2 and Annex 2 of the Guidance on Long-Term Preservation of Instrument-Generated Data.

The examples of various analytical instruments presented in this guide were developed based on information collected from instrument users.
Depending on the manufacturer or model, instruments may generate data different from those described in this guide. In such cases, these examples should be used as a reference when selecting the data to be preserved.

Depending on the purpose of preservation, retention of audit trail metadata may be required. However, in many general analytical instruments, audit trail functionality is either not provided as a standard feature or is available only as an optional add-on. Therefore, audit trail metadata is not included in these examples.

[3] The first part of this sentence requires that attachments clearly indicate their relationship to the original communication record. The latter part requires that reply to emails and forwarded.

Section 5.3 and subsequent sections provide examples categorized as follows. Data marked with ● in the tables are recommended for preservation.

- **Primary application**s: Reference information indicating the intended use of the instrument.
- **Measurement condition**s: Examples of metadata related to measurement.
- **Instrument-generated data** (format/file type): Reference information for identifying the data initially generated by the instrument.
- **Analysis conditions** (metadata; where applicable): For instruments in which results are derived through analysis, reference information on important analytical metadata.
- **Analytical data** (format/file type; where applicable): Examples of result data output by analytical applications. In addition to generic formats, proprietary formats may also be used; this information is provided as a reference for data identification.

5.1. Email (research-related records/notifications)

In cases such as commissioned testing, important information necessary for the conduct of the test may be provided by the sponsor to the contractor via email. For this reason, retention of emails may be required.

Examples of data that should be preserved in emails and methods of preservation are as follows:

Data that should be preserved:

- Send and receive history
- Attachments
- Sender and recipient information
- Timestamps (date and time of sending and receipt)

Preservation method:

- Archive the email in its original form.
    - When archiving, searchability and security shall be ensured.
    - Ensure that the linkage to the experimental record is maintained.

5.2. HPLC (High-Performance Liquid Chromatography)

The primary applications include drug concentration analysis.

For further details, refer to the *Long-Term Archiving of Analytical Instrument – Operational Guidebook* and its Annexes (checklists).

5.3. MS (Mass Spectrometer)

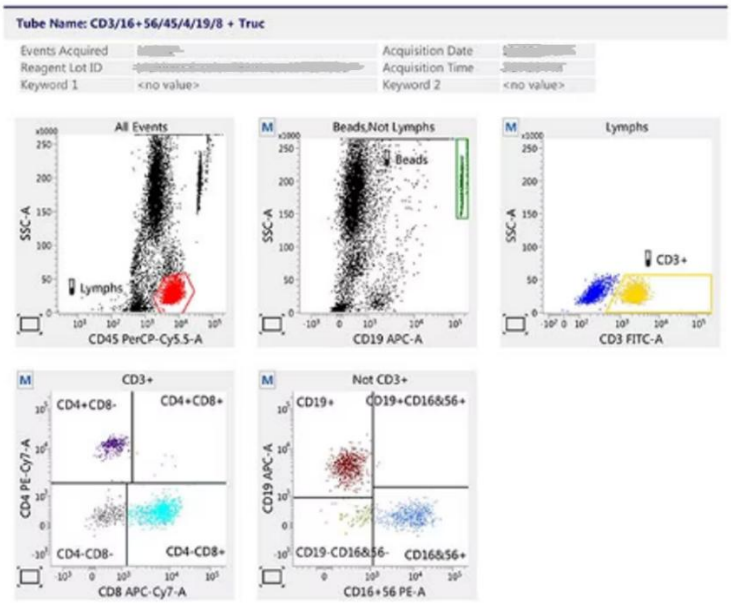
Primary Applications	While used for diverse purposes, this guidebook assumes LC/GC-MS or MS/MS systems employed for drug concentration analysis.

Measurement Conditions (metadata)	● Injection sequence information ● Sample Information (Standard/unknown sample type, concentration of standard/internal standard substance) ● Instrument conditions (MS conditions (ionization conditions, detection conditions, etc.))
Raw Instrument Data	Raw data corresponds to multiple chromatograms and is not in a universal format; within the scope of this investigation, the manufacturer's proprietary format is adopted. ● Change history data, etc. ● Manufacturer's proprietary format files ● Output files in PDF or image formats
Analysis Conditions	● Analysis Signal Channel (Target Ion) ● Waveform Analysis Conditions ● Retention Times for Target Substance and Internal Standard ● Calibration Curve for Target Substance
Analytical Data	● Quantitative Analysis Results (Custom Format, PDF, CSV)

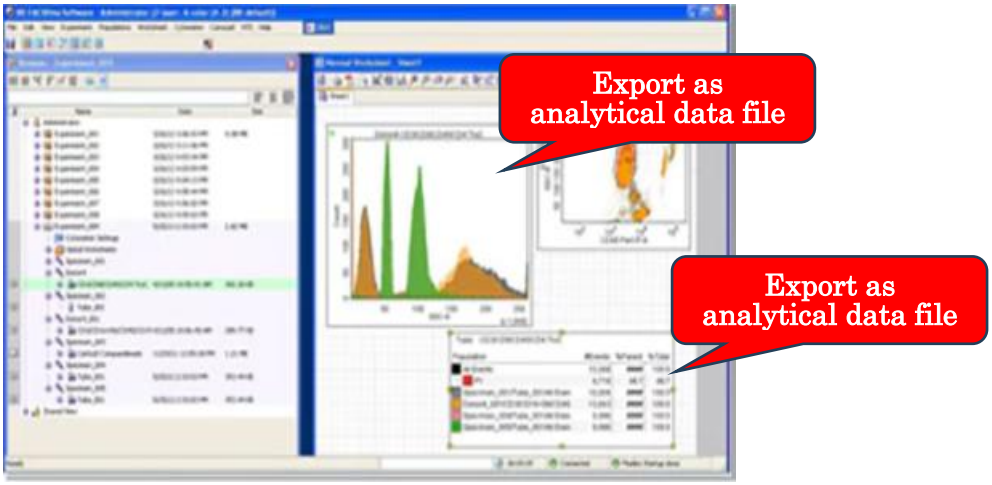
5.4. FCM (Flow Cytometer, FACS, etc.)

Primary Applications	Analysis of heterogeneous cell populations, isolation of specific cell populations, etc.
Measurement Conditions (metadata)	Parameter settings (measurement items, sensitivity, flow rate, threshold/gate settings, etc. (set within the instrument software))
Raw Instrument Data	● Instrument control software (e.g., DiVa) files (unreadable: contains measurement conditions, measurement date/time) ● Measurement data FCS files (unreadable): File save date/time *Readability: Measurement data (FCS files) can be read by analysis software (e.g., FlowJo).
Analysis Conditions	Gate settings
Analytical Data	Cell event data (count), analysis data (cell presence ratio, average fluorescence intensity, etc.) ● Analysis Software (FlowJo) File (Not Readable: Analysis Conditions) ● Analysis Data Files (CSV, JPG, etc.: readable) ● Numerical values of analysis data are in CSV format, and figures are in JPG format, etc., with readability.

Example of occurrence data (display example on instrument control software)



Example of analytical data (display example on analysis software)



5.5. Live cell analysis systems

Primary Applications	Cell proliferation/migration, fluorescent reporter assays, cell death/survival rate assessment
Measurement Conditions	Measurement date, measurement time, measurement mode, analysis parameters
Raw Instrument Data (Format, File Type)	● Measurement Results (Time-lapse Images, Cell Proliferation Curves) ● Measurement data video files (readable): MPEG-4, Windows Media Video, Windows AVI video ● Measurement Data Image Files (Readable): JPEG, PNG, TIFF, etc. (File names may include capture date/time)
Analysis Conditions	Fluorescence channel selection, detection threshold settings
Analytical Data (Format, File Type)	● Analysis Data: Fluorescence Analysis (Count, Total Area, Total Integral Intensity) ● Analysis Data File (Human-Readable): txt (Assay Name, Analysis Date/Time)

5.6. Tissue Slide Imaging System

Multiplex imaging

Primary Applications	Multiplex fluorescence imaging and spatial analysis of tissue sections
Measurement Conditions	Slide information, antibody panel, staining conditions, imaging settings (exposure time, magnification, filters, etc.)
Raw Instrument Data (Format)	● Multispectral image data (raw image: formats such as .qptiff or .im3) ● Instrument software installation files (Phenochart/inForm, including measurement conditions) ● Image data files (.qptiff, .im3, etc.: non-readable, containing date and slide information) *Readability: qptiff and im3 are only viewable in inForm/Phenochart and have low readability, but can be converted to universal formats like tiff.
Analysis Conditions	Region of Interest (ROI) settings, segmentation conditions, parameter settings (cell count per region, cell classification results, marker intensity, spatial distribution information, etc.)
Analytical Data (format, file type)	● Analysis software (inForm) analysis data files Analysis conditions and results can be output in the following formats: cell (non-human-readable), .csv, .txt, .tif, .jpg, etc. (human-readable)

Virtual slide scanner

Primary Applications	Image analysis of tissue sections
Measurement Conditions	Slide information, magnification, illumination conditions, scan mode, etc. (within device software)
Raw Instrument Data (Format)	● Virtual slide images (NDPI format)
Analysis Conditions	ROI settings, color balance, focus information, etc. (Area per region, image measurement values, annotation information, etc.)
Analytical Data (Format)	● Analysis software data files (analysis conditions/results: .csv, .jpg, etc., readable)

5.7. Telemetry (Remote Measurement System)

Temperature logger

Primary Applications	Remote monitoring of temperature and humidity in rooms, storage facilities, and during transport
Measurement Conditions, etc.	Channel No., Channel Name, Temperature (Value, Measurement Date/Time), Humidity (Value, Measurement Date/Time)
Raw Instrument Data (Format, Structure)	● Device Data (TRZ format, CSV format): Channel No., Device Name, Temperature (Value, Measurement Date/Time), Humidity (Value, Measurement Date/Time)
*Information displayed only, not output	● Group name, device name, serial number, number of data records, battery level (software display) ● Data displayed only (device name, serial number, etc.) should be saved via screenshots.

Example of temperature display

Source: Ondotori Web Storage
https://ondotori.webstorage.jp/manual/usage.html

Animal Experiment Telemetry (Remote Measurement System)

Primary Applications	Remote measurement of experimental animals' heart rate, body temperature, activity levels, etc.
Login-Related Information	User ID, login time, logout time
Measurement Condition	Data Acquisition/Save Log: Start time, stop time, file name Measurement conditions: Channel input signal settings, protocol (test number, test substance name, dose, animal number, route of administration), analysis points, sampling rate Measurement schedule: Start date, start time, end date, end time, measurement duration
Raw Instrument Data	● Data Files: File name, protocol information, individual measurement parameters ● Log files: Various logs such as operation, measurement, and analysis
Analysis Conditions	Varies depending on measurement parameters.
Analytical Data (Format, Structure)	● Varies depending on the measurement parameter. Display of waveforms, etc., in the analysis software's proprietary format, and output of analysis values in CSV format, etc.

5.8. Blood analyzers (automated blood cell counters, etc.)

Automatic multiparameter blood cell analyzer

Primary Applications	Blood analysis. Specifically, measurement of blood cell counts (WBC, RBC, Hb, Plt, etc.) and determination of abnormality flags (diagnostic support)
Login-related information	Login User Name
Measurement Conditions	Measurement date, measurement time, animal ID, measurement mode, animal species
Raw Instrument Data	Measurement Results (Scattergram/Particle Size Distribution Chart), Evaluation (Mark) Output files available: ● Measurement Data File: Measurement date/time, analysis results, measurement mode, model, software version (CSV format, FCS format) ● Quality control file: Save date/time, lot number, analysis results Selecting the lot number from the CD-R included with the reagent automatically loads the display values and limits. Readability: The precision control file is device-specific (extension: qcf). However, numerical values can be read as measurement data via CSV output (text output). ● Graphic mode is saved as an image file (bmp or png).

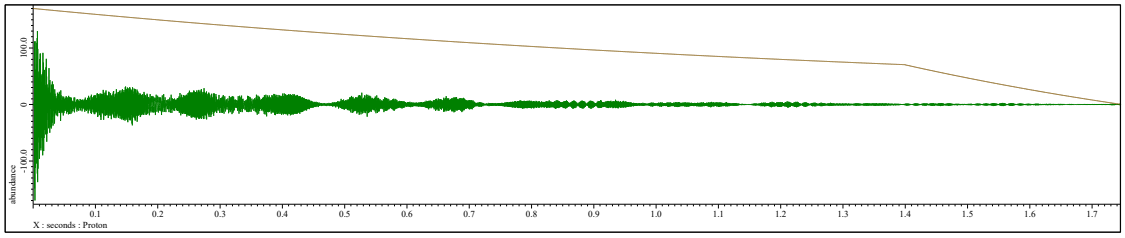
5.9. Colony Counter

Primary Applications	Measurement of colony count, size, and distribution on culture media
Measurement Conditions	Measurement date, detection level, pattern size, frame Automatically counts colony numbers from camera images and aggregates them using processing software
Raw Instrument Data	Image, measurement results (colony count, location information, area) ● Image file: BMP format ● Measurement data file: Colony count, location information, area (CSV format) ● Processing software files: Colony count, aggregation, table creation, graph creation (Excel) Processing software settings can be freely configured and reflected in the input sheet, result evaluation sheet, table sheet, and graph sheet (Excel).

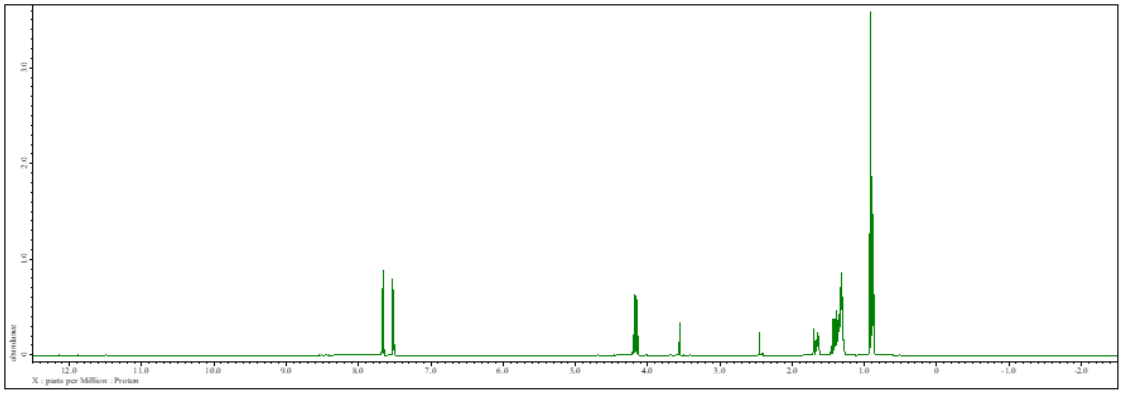
5.10. NMR (Nuclear Magnetic Resonance Device)

Primary Applications	Molecular structure analysis, quantitative NMR (qNMR), purity measurement
Measurement Conditions	Observed nuclide, magnetic field/observation frequency, spectral width, number of data points/dwell time, number of accumulations, pulse sequence, flip angle, relaxation delay, temperature, lock/shim conditions, reference, operator/date, software/version, etc.
Raw Instrument Data (Format, File Type)	● Raw time-domain data: FID for 1D, SER for 2D and higher (Instrument-specific format measurement data files. Unreadable. Reprocessing possible)
Analysis Conditions	● Fourier transform (FT), phase correction, baseline correction, window function (apodization), zero-padding, digital filter removal, chemical shift reference, peak detection/integration/assignment rules, qNMR-specific calibration (internal standard type/amount, relaxation delay)
Analytical Data	● Frequency domain spectrum (real/imaginary, 1D/2D, etc.), peak list, integral values, assignment table File formats: JCAMP DX (.jdx/.dx), nmrML (XML), NMRPipe format, etc.

Example of FID Data



Example of post-FT data



5.11. Temperature and Humidity Control System

Primary Applications	Equipment for monitoring the environment in material storage, experimental, and animal housing facilities. Measurement results are continuously monitored but not always output.
Login-Related Information	Login Information: User ID, Login Time, Logout Time
Measurement Conditions	Data Acquisition/Storage Log: Start time, stop time, number of data points, current status, filename Network Log: SNTP Log (Time, Code), DHCP Log (Time, Type, Content) FTP Log (Time, Flags, File Name) Message Log: Start time, stop time, content Error Information: Type, Content
Raw Instrument Data	● Temperature and Humidity Data Log: File name, Channel ID (temperature sensor, humidity sensor), Channel name, Measurement data (temperature, humidity, data acquisition date/time) Output files for a specified number of days (screen capture/paper record) ● Alert Settings Information: Temperature Sensor (ID, Name), Humidity Sensor (ID, Name), Alert Trigger Value, Alert Trigger Condition (Exceeding or Falling Below Trigger Value) ● Alert Information: Alert trigger condition, affected sensor (ID, name), trigger start to end time ● Data displayed only (e.g., alert setting information) should be saved via screenshots, etc.

6. Glossary

Table 3 provides explanations for terms used in this guidance.

Table 3: Explanation of Terms

Term	Description
Analysis Parameters	Values set for methods such as peak waveform processing. Includes methods for baseline setting and noise removal.
Audit Trail	A secure, computer-generated, time-stamped electronic record that enables the reconstruction of the process of events related to the creation, modification, or deletion of electronic records. Examples include logs of original data generation, analysis audit trails, and system-related audit trails.
Sample Schedule	Injection Schedule
Instrument Parameters	Includes settings such as Analytical Instrument operating conditions, detector type, and solvent type.
Timestamp	・ Date and time automatically stamped by the computer (ERES Guidelines[4]) ・ Technology that attaches information capable of detecting changes and proves whether modifications occurred after that time[5]
Data Integrity	The degree to which data is complete, consistent, accurate, credible, and reliable, and these characteristics are maintained throughout the data lifecycle. The fundamental principles of data integrity (ALCOA++) include Attributable, Legible, Contemporaneous, Original, Accurate, Complete, Consistent, Enduring, and Available.
Dynamic Data and Static Data	The FDA explains the distinction between dynamic and static formats for Analytical Instrument data in Section III d of its guidance, "Data Integrity and Compliance With Drug CGMP Questions and Answers Guidance for Industry," under "[6] ." Static data includes printed paper records and PDFs. "'Static' refers to fixed data records such as paper records or electronic images, while 'dynamic' means the record format allows interaction between the user and the recorded content. For example, in a dynamic chromatography record, the user can change the baseline or reprocess the chromatography data to display peaks smaller or larger. Additionally, users can modify formulas and entries within spreadsheets used to calculate other information, such as test results or computed yields." (Quotation marks indicate text translated from English by the R&D Data Archiving Committee)
PDF	Portable Document Format, a file format for exchanging documents containing text, images, and graphics, specified as ISO 32000

[4] Source: Pharmaceutical and Food Safety Bureau Notice No. 0401022, April 1, 2005: Guidelines on the Use of Electronic Records and Digital Signatures in Applications for Approval or Permission of Pharmaceuticals, etc. (ERES Guidelines)

[5] Time Business Council (https://www.dekyo.or.jp/tbf/index.html) Time Business Glossary

[6] Source: Data Integrity and Compliance With Drug CGMP Questions and Answers Guidance for Industry December 2018, III d (https://www.fda.gov/media/119267/download)

Revision History

Date	Version	Details of revision
2026.2.18	1.0	

R&D Data Archiving Committee Guidance Preparation Members

Director in charge	
Yasushi Okada	Ricoh Japan Co., Ltd.
Chairperson	
Sayuri Uehara	Astellas Pharma, Inc.
Committee Members	
Akira Yamazaki	Kyowa Kirin Co., Ltd.
Satoshi Morino	Eisai Co., Ltd.
Naoto Miyaji	Lang Edge, Inc.
Yoshiaki Hiraishi	Terumo Corporation
Junpei Miura	Business Engineering Corp.
Yoshimasa Haruyama	Business Engineering Corp.
Yoshiaki Yamada	Hitachi High-Tech Analysis Corporation
Michihiro Kimura	Japan Image and Information Management Association
Collaborating Members Japan Society of Quality Assurance (JSQA)	
Hiroyasu Ishiguro	Lead Chemical Co., Ltd.
Akiyoshi Otake	Astellas Pharma, Inc.
Yoshikazu Masaki	Otsuka Pharmaceutical Factory, Inc.