P-257

Best practices for storing electronic data of retired systems -Example of temperature / humidity monitoring system-

Daisuke Sasaki[1,2], Naofumi Hagiwara[1,3], Masahiro Bando[1,4], Fuminori Kondo[1,5]

[1]Japan Society of Quality Assurance, [2]Astellas Pharma Inc., [3]Asahi Kasei Pharma Corporation, [4]Eisai Co., Ltd., [5]Sumitomo Chemical Company, Limited

Abstract

Data Integrity (DI) has been brought to attention since 2016 with 'GXP' Data Integrity Guidance and Definitions, issued by the Medicines and Healthcare products Regulatory Agency, a UK regulatory agency. In line with the increasing attention to DI, its importance in the GLP field has been clarified recently. In addition, the Organization for Economic Co-operation and Development (OECD) issued "OECD series on principles of good laboratory practice and compliance monitoring, Number 22, Advisory Document of the Working Party on Good Laboratory Practice on GLP Data Integrity (OECD AD No.22)" last year. However, even in the multiple documents mentioned above, the method corresponding to each device or system is not specifically described. Even if one of the documents can be referred to the relevant system, described method is not always fit to the level that can be handled by each facility. OECD AD No.22 states, "The effort and resource applied to assure data integrity should be commensurate with the risk and the impact of the associated data integrity failure." It means complete response is not always required, and there is room to select the degree and type of response in consideration of risks. In view of the situation, Japan Society of Quality Assurance sets up a virtual scenario called "Electronic data storage of the system which subject to retirement", identified the issues involved, and examined countermeasures for the issues. In this presentation, we will report practical DI countermeasures considering the level at which it is feasible for most facilities.

What is DI? What should we do for it?

DI means ensuring that the contents of the data and data system are consistent and reliable, and that the data is not abnormal due to an external attack or some error.
Recently, it has been attracting attention in the GLP field as well.
However, it is hard to practice DI by relying only on existing documents.

The ALCOA+ principles ensure DI. Understanding them is important to advance DI.

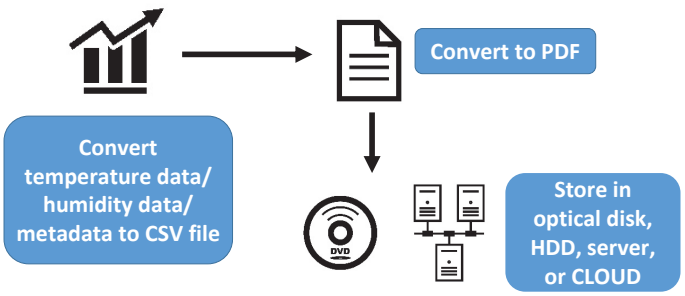
Attributable: It is clear who wrote and when/where the data was collected.
Legible: Literally, it means "readable" and "understandable as data". It means that the data is not meaningless.
Contemporaneous: Recording is immediate. "Later" is not permitted.
Original: The data recorded at the very beginning.
Accurate: Literally, the data is accurate.
Complete: There is a record that can reproduce the event.
Consistent: There is no contradiction in the record.
Enduring: Withstand the storage period.
Available: Can be used when you need it.

Model case: Temperature / humidity monitoring system to be retired

Due to the expected shortage of repair parts within 3 years, it is necessary to update the "Temperature / Humidity Monitoring System".
The detailed situation for the case is as follows.

What should be done	Extract and store entire raw data from the old system.
Years how long the system has been used	10 years
Rooms	100 rooms
Stored data	Minutely data and audit trails
Data output function of the system (NOT validated)	Output report as Minimum / Maximum value for each room on each day. Minutely data and audit trails can be output as a CSV file.
Extended function for data output (NOT implemented and needs to be developed)	Each data (minutely data, daily minimum / maximum value for each room, audit trail) can be output all at once to CSV or Excel files by using Excel macro etc.
Browse old data using successor system	They cannot be browsed.

Measure 1: Output to the electronic file (conversion to Standard format data)



[Points of consider for data migration]
- In principle, all data should be migrated to CSV file.
- Storage of only the necessary part of the data may be acceptable if there is a valid reason and with suitable risk assessment.
- The storage format should be taken into account the risk at the time of conversion and operatablity when the migration and archiving.
- Check if validation has done when using software or Excel macros for conversion.

〔Points of consider for storage〕
- Consider the lifetime of the storage medium, operation during replacement, and how to maintain backup.
- Establishing a method for checking legibility / avoiding tampering after archiving.
  - "Check ALL DATA" vs "Checking sampled data".
  -Use number of folders / amount of data / hash value for assuring the complete conversion.
- Validate in advance the procedure for data migration / verified copy creation after archiving.

Measure 2:Time capsule approach (maintaining the old system)

It is the method to keep the software of the old system and data is maintained as it is.

Pros:
It is the easiest and least expensive way because It is not necessary the file format migration or new system development.
Cons:
It is usually difficult for continuous support from the manufacturer of the system, so that the issue is how to ensure the readability of the data during the retention period.

Feasibility of the method
In the case study of this poster, it is assumed that the manufacturer's support will run out within 3 years. Therefore, this method can be considered as a temporary countermeasure up to that point on the premise that long-term countermeasure by Measure 1 or 3 in the future.
The old systems should be stored in a designated safe place. The frequency of inspections also should be determined appropriately, considering the following factors.
E.g., Consideration of planned retention length, manufacturer support, and data importance.

(Similar method): Using virtual environment
The method is used when the old system does not work the new OS.
As a method close to this method, there is a means to use a virtual environment. It is used when the system does not support the new OS.
Virtually launch the old OS on a new OS and operate the old system under the environment. Since it is not necessary to maintain the old hardware itself, the sustainability period will be longer than the veritable time capsule approach. Even in the case, however, it is necessary to consider the feasibility of the manufacturer's support for operation in a virtual environment.

Measure 3: Implementation or development of a data viewer system

A system only for viewing past data, which cannot be used for new data collection.

Pros:
The electronic data of the old system can be used as it is
Can be searched and aggregated on demand as needed
Cons:
Huge cost for system development and implementation
Requires cooperation by vendors of old system

Important notice
It is safe way that developing a simple read-only system that just excludes data collection function from older systems.
Adding new function means adding more validation items.
When removing functions other than data collection, please consider whether it will lead to inconvenient data hiding.
User management and audit trails are essential when implementing data modification features.

Discussion

The field of electronic data storage is in a transitional period. For the time being, it is necessary to continuously consider the impact of the risks and errors in each event when the mistakes occur, depending on the degree of establishment of storage techniques and methods.
In response to Data Integrity, several documents have already been issued all over the world as presented in this poster. Also in Japan, it's time to measure for data integrity since OECD AD No.22 has also been issued.
In the poster, we suggested the method, at a current realistic levels, as a best practice of countermeasures for storage the electronic data from the systems to be retired. We hope the practical method will be useful for data storage though each organization should select which contents to use.
In addition, the system in the "Model case" was a legacy system.
However, it will be able to partially utilized to resolve the problem for the relevant part, even in the current system when the measures for data storage are insufficient.

Related Documents
Medicines & Healthcare products Regulatory Agency (2016). "'GXP' Data Integrity Guidance and Definitions."
OECD AD No.17: Advisory Document of the Working Group on Good Laboratory Practice, Application of GLP Principles to Computerised Systems
OECD AD No.22: Advisory Document of the Working Party on Good Laboratory Practice on GLP Data Integrity
JIIMA, R&D Data Archiving Committee (2020). "Long-term archiving guidance for analytical instrument data, Version 2.1"
JIIMA, R&D Data Archiving Committee (2021). "Long-term archiving technology guidebook for analytical instrument data, Version 1.0"
(In Japanese)

Abbreviations
CSV: Comma Separated Value, DI: Data Integrity, JIIMA: Japan Image and Information Management Association, MHRA Guidance: "'GXP' Data Integrity Guidance and Definitions.", OECD: Organisation for Economic Co-operation and Development, OECD AD No.17: Advisory Document of the Working Group on Good Laboratory Practice, Application of GLP Principles to Computerised Systems, OECD AD No.22: Advisory Document of the Working Party on Good Laboratory Practice on GLP Data Integrity, OCR: Optical Character Recognition, OS: Operating System, PDF: Portable Document Format, PMDA: Pharmaceuticals and Medical Devices Agency, SOP: Standard Operating Procedure

Issue classification	Items	Details of the issue	Measures	Risk of data loss	Need for manual work	Frequency (When need any manual work) Process name and/or number	Risk in case of poor measures
Decision before storage work	Batch saving of current data (data format might be converted)	Develop a novel dedicated system only for browsing data. Incorporate the old system into the virtual environment in the new device.	Check system specifications by CSV Check data integrity comparing each data before and after data migration	Mid	Yes Difficult to proceed with data migration completely without manual work.	- Selection of data to be migrated - Checking data after migration	Low It may be difficult to output all the original data electronically. Consider a hybrid method, such as outputting data on paper.
	Save in current data format	Store the old system for browsing data.	Owning multiple devices for old system (insurance for breakdown) Equipment inspection regularly The alternative measures are required before all devices fail.	Low	No		Low There is no data omission since the old device is saved as it is. Missing all data remains as a future risk when all devices fail.
	Change definition of "Raw data"	When need to change the definition of raw data.	Define in the system retirement plan and declare changes in the completion report.	Low	No		Low There is no data loss due to this measure.
Target data for storage	Large number of data to be stored	Number of data for Temperature/Humidity 1 data/min. x 10 years x 100 rooms	There is no way other than to save all. (Basically, it is not possible to reduce the data acquired as GLP data.) Alternatively, it may be possible to save the data only during the GLP test period.	High	Yes	- Selection of data to be migrated - Save files and folders separately Since the amount of data is huge, the work frequency is high.	High It is very difficult to migrate all data without omission and save it as the original data.
	Huge data to store	Output data without omission.	The output function should be confirmed by validation.	High	Yes	- Plan the Validation plan - Confirmation during validation - Confirmation after output Since the amount of data is huge, the work frequency is high.	High It must be possible to reliably determine whether all of the original data can be migrated and stored.
	Types of metadata to output	Metadata identification Is there any metadata that cannot be output? What to do in that case.	Output is required (confirm with the vendor for output methods and items that can be output) The output function should be confirmed by validation. Required data e.g. Information accompanying temperature and humidity data (room, sensor No.) Audit trail (login/out and operation) System log (backup, deviation, equipment abnormality) Paper material should be needed for the metadata which cannot be output.	Mid			Low Consider hybrid methods such as outputting data on paper or saving screenshots.
	Unit of temperature (Celsius / Fahrenheit)	Is the unit properly attached to the data?	As the temperature/humidity data, even if the unit is not output with data, it may sufficient to understand correctly what the numerical value means afterwards. If that is difficult, leave the information on paper to help understanding the data correctly.	Mid			Low Is it possible to say that there is little damage if missing information is only unit?
Storage method	Verify the method for data storage	How to conduct validation test for data output.	Randomly specify the period and output the data and confirm if there is any abnormality. It is necessary to confirm the specifications with the vendor before validation tests.	Mid	Yes	- Plan the Validation plan - Conduct validation - Confirm the results of validation	High It is necessary to perform validation tests without omission.
	Output the data	How to data output. Whether user can find the temperature/humidity data with the storage they want to check from the list of numerical values without any problem.	Decide how to output the separated data E.g. Room temperature of Room 001 on Jan 2010. Number of data: 1 point x 60 minutes x 24h x 31 days ≒ 45000 points. Require validation. It is worth discussing whether validation on the extracted data is sufficient.	Mid	Yes	- Select the data to be migrated - Separate files and folders as needed and save them properly Since the amount of data is huge, the work frequency is high.	High
	Data conversion to CSV	Change data format to CSV.	The output function should be confirmed by validation. Especially, it is required confirmation of consistency before and after data conversion.	High	Yes	Selection of data to be migrated. Confirm Completeness and legibleness after format changes.	High Possible metadata loss.
	Data conversion to Microsoft Excel format e.g. .xlsx .xls	May be garbled.	The output function should be confirmed by validation.	High			High
	Output xlsx file using Excel Macro.	Is the Excel Macro working properly?	The output function should be confirmed by validation.	Mid	Yes	- Plan the validation test - Conduct test - Confirm results of validation test	High It is necessary to perform validation tests without omission.
	Output PDF file 1. Directly using function of the system 2. Via CSV to PDF 3. Via xlsx to PDF	The data is listed on a huge page.	Predetermine the search method to be able to reach the desired data. (If the minimum and maximum values of the day are not treated as raw data, saving the data may be optional.)	Low	No		Low It is also necessary to give up if it is difficult to convert to PDF.
	Output to the graph	Can it be displayed as a graph?	Build a system to create graphs. Is it necessary to make a graph when viewing temperature data? Only data can be saved. It would be nice to have a system app for drawing.	Low	No		Low
	Storage the output data	When outputting to Excel format, it is necessary to make the data non-rewritable.	e.g. Storage as Read-only files. Compress, and using password Immediately convert to PDF Using the archive area with limited access rights Set a password Even with such ways, 100% lock is still difficult. Defend with multiple measures.	Mid	Yes		Mid
	Storage location	Whether to put them in the archive area of the server or write them on a disk.	1. Electronic data storage system 2. Electronic archive area (server) 3. CD, DVD, Blu-Lay, etc. 1 or 2: Periodic inspection is required. 2: Migration maybe required when updating the server. 3. The disc needs to be updated according to its durability. CLUOD will be available for 1 and 2.	Mid	Yes	When data migration.	Mid
	Data integrity	Correspondence to data output mistakes and garbled characters.	In order to prevent garbled characters and output mistakes, the primary and secondary are created by two independent workers, and compare the output afterwards.	High	Yes	When confirm the output data.	Mid Confirm with validation test. Consider a different method when garbled characters are found during validation test.
Data migration	Data migration	Migration by vendor.	Vendor may do it (at least the implementer). GLP education for vendors is not mandatory and should be managed by GLP staff (leave the signature by GLP staff).	Mid			Mid When it was conducted by a vendor (expert), early detection of abnormal situations can be expected.
		Design migration method by vendor, and conduct variation by user (GLP staff).	It is not necessary to conduct validation test by ourselves, but at least the user should confirm that the contents of the validation test are consistent with the user's intended use. Even in such case, it is necessary to confirm the plan and the result in the validation plan / report and confirm that there is no problem.	Mid	Yes	When the user validation test.	Mid The validation itself does not result in data loss, but it is important to confirm that there are no abnormalities at this point.
		Data migration procedure Data migration work specifications/report	Inconsistencies often occur when assembled only on a desk. Each document, whether user-created or vendor-created, should be brushed up through rehearsals.	Low			Low
		What to do when a failure occurs during data migration.	Consult with vendor in advance so that the vendor can handle them. It is also necessary to carry out according to the operation of GLP, such as documentation.	Low			Low The original data still remains.
		When problems occur during migration work and data output.	Ask the vendor to keep a record of the cause, response, and extent of impact.	Low			Low The original data still remains.
	(Option) Confirmation that the new system is able to acquire data as before	Output data on each of the old and new systems and confirm the data integrity as a part of validation of the new system.	It is necessary to define in advance which device the raw data comes from, during the period of data acquisition from both systems (declare that the data from new system data will be the raw data from the time of switching of systems). From the point of view of data storage of the old system, it is not no necessary to confirm the identity of the new system with the old system.	Low			Low
Disposal of the old system	Disposal of old server	Identification of required documents.	Follow the method predetermined by SOP Required documents e.g. Change/update of device registration slip Disposal record Write in advance that the old system will be discarded in the retirement plan, and record it in the retirement report.	High	Yes	At the time of system disposal.	High
Governance for data maintenance	Maintenance the data after data migration (Applicable to both original and format changed data storage)	It is necessary the data should be original, legible, and complete. Consideration the need for backup.	Prepare SOP for data maintenance and train operators regularly. - Risk assessment (what to do?) - Compliance with appropriate quality and risk management systems. - Assign the responsible person. - Reliable method to protect the records.	Mid	No		Mid Previous files may not be legible due to the update of the application.