



Management of the Life cycle of Laboratory Electronic information-GAP Analysis.

Raghuram Pannala, Narayanareddy Papadasu

**ScieGen Pharmaceuticals Inc, Hauppauge, NY-11788, USA*

ABSTRACT

Electronic data management in an analytical laboratory extends beyond chromatographic analysis and result issuance. The concept of data integrity, as outlined by the FDA, encompasses data generation, processing, storage, backup, retrieval, and dissemination. The term data integrity refers to the accuracy, consistency and reliability throughout its life cycle. This article primarily delves into electronic data storage, backup, archiving, retrieval, and restoration, shedding light on common issues and complexities associated with the process. Maintaining data integrity in the pharmaceutical sector is essential for meeting regulatory requirements. Regulatory agencies like US Food and Drug Administration (USFDA), European medicines agency (EMA), Health Canada and several regulatory agencies emphasize the importance of data integrity in the field of pharmaceutical and life sciences sector. Regulatory agencies implementing more stringent regulations and guidelines to guarantee that the entire life cycle of pharmaceutical products –ranging from research and development to Quality control, Quality assurance, Manufacturing and distribution- is dependable, precise and uniform. Adhering to regulatory standards, including good laboratory practice (GLP), and good manufacturing practices (GMP) is essential for maintaining data integrity and ensuring compliance with regulations during every stage of product development to commercialization. Breaches in data integrity can severely affect Company's reputation, stake holder trust, and lead to substantial regulatory consequences, including fines, product ban or legal proceedings. In addition to the above consequences, regulatory agencies may delay or deny the approval of new pharmaceuticals. Based on the above issues, this article primarily delves into electronic data storage, backup, archiving, retrieval, and restoration, shedding light on common issues and complexities associated with the process. The information provided in this article aids in identifying unauthorized data tampering, deletion, and in enhancing the implementation of data life cycle management to ensure compliance with ALCOA+ principles (Attributable, Legible, Contemporaneous, Original, Accurate, Complete, Consistent, Enduring, Available).

Key Words: Electronic data, Stand-alone system, data integrity, data backup, data retrieval, data threats.

1. Introduction:

In recent years, the pharmaceutical industry has undergone a significant shift in electronic data management. The data integrity guidance released in December 2018¹ marks the final addition to the data integrity toolkit. Various publications and guidance documents have been published on this subject, including the 'PDA Technical Report 80', Data Integrity Management System for Pharmaceutical Laboratories², WHO guidance³, GAMP 5⁴, MHRA Q&A guidance^{5,6}, and several other reference documents⁷⁻¹². These resources have provided valuable guidance on the life cycle management of laboratory electronic data. It is important to note that PDA TR 80 specifically focuses on laboratory data, while the other guidance documents are applicable to electronic data generated in pharmaceutical manufacturing. Upon reviewing the aforementioned literature and considering current knowledge and applications, it has become evident that certain areas require further clarification and discussion. This article primarily focuses on electronic data storage, backup, archive, and retrieval, as well as addressing common issues and intricacies of the process. It also outlines the necessary steps to ensure the safeguarding of data integrity against potential threats.

2. Life cycle of Electronic data in an analytical laboratory

A typical analytical laboratory consists of several instruments in addition to the commonly discussed HPLC and GC instruments. These instruments include a UV Spectrophotometer, FTIR, Tiltrotor, Particle size analyser, XRD, TOC analyser, ICP MS, and DSC. The majority of labs use chromatographic software that is compliance with 21 CFR Part 11. However, there are often gaps in understanding the software, which are identified during inspections. These analytical instruments come in different models and versions of vendor software, giving users various options to choose from. It is crucial to select a software version that meets the organizational requirements, as vendors may not disclose all potential shortcomings during initial discussions. The life cycle of electronic data in an analytical laboratory as shown in **Figure 1** and the common classification of analytical instruments in any analytical lab is shown in **Figure 2**.

This classification is based on regulatory guidance such as USP chapter Analytical instrument qualification <1058>¹³ and reference documents. According to FDA guidance, electronic records (raw data and meta data) must be archived. Analog signals from a detector are converted to digital form by (chromatographic) data processors (raw data), and then processed using chromatographic software with various algorithms and commands (meta data) to generate analytical electronic data¹⁴⁻¹⁹. This generated data is used for analytical decisions. Once a decision is made, the data (raw data and associated meta data) remains on the hard disk/server until it is needed for a specific purpose. For more information on computer-related terms refer to the 'Glossary of Computer System Software Development Terminology' (8/95)²⁰ and the 'American National Standard for Information Systems, Dictionary for Information Systems, American National Standards Institute', 1991²¹. An overview of "Post data generation travel" was shown in **Figure 3**.

It is highly recommended to restrict access or provide 'view only' rights to the data once it has been released on 21 CFR compliance systems, which restricts accidental or/and undocumented reprocessing of the data. **Figure 4**. illustrates an example of a locked file after processing and sample release on a stand-alone system.

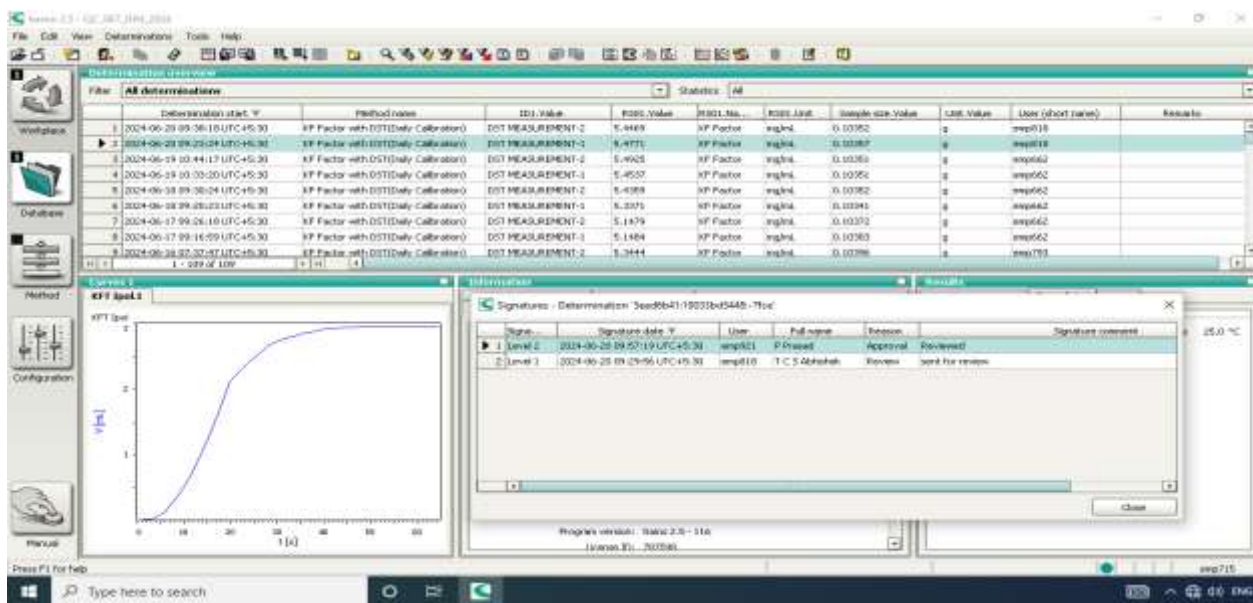


Figure 4. Example of a locked file after processing and sample release on a stand-alone system

Many chromatographic software programs have the capability to lock the results after processing. These locked and electronically signed chromatograms can be presented to auditors. However, not all instrument software provides a feature for locking the data. Once a sample decision has been made, vigilance over the data may decrease, but the usage of the data remains the same. Analysts often open previous analysis data for reference or casual review of old data. Any accidental or undocumented reprocessing of data at this stage poses a significant risk to the firm, as system audit trails are not typically reviewed for these types of events and audit trails related to data may not be reviewed after a decision has been made²². Electronic data generated on analytical instruments in the laboratory may be stored on various types of devices in different locations, as shown in **Figure 5**.

PLC-based devices such as titrator, pH meters, and balances may store data on the instrument's internal memory or connected USB devices in either CVS or PDF format. The risk associated with this type of data is that it can be easily manipulated²³. According to the FDA Guidance from December 2018 on Data Integrity and Compliance with Drug cGMP, temporary backup copies made in case of a computer crash or other interruption would not meet the requirement to maintain a backup file of data as stated in § 211.68(b).

Data generated on stand-alone computers is stored on local hard drives, which can pose a significant risk, especially for data generated on software lacking proper database architecture with restrictions. This data can be easily deleted from explorer mode²⁴.

Most stand-alone systems and instrument software do not support data protection, allowing users to delete data after it has been generated. If proper reconciliation is not done these events can go unnoticed²⁵ **Figure 6. and Figure 7.** shows a file deletion from windows before applying user restrictions.

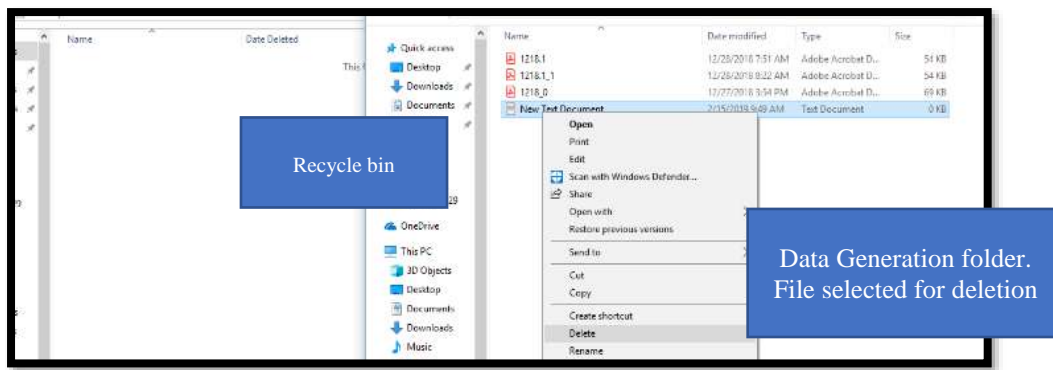


Figure 6. File deletion from windows before applying user restrictions

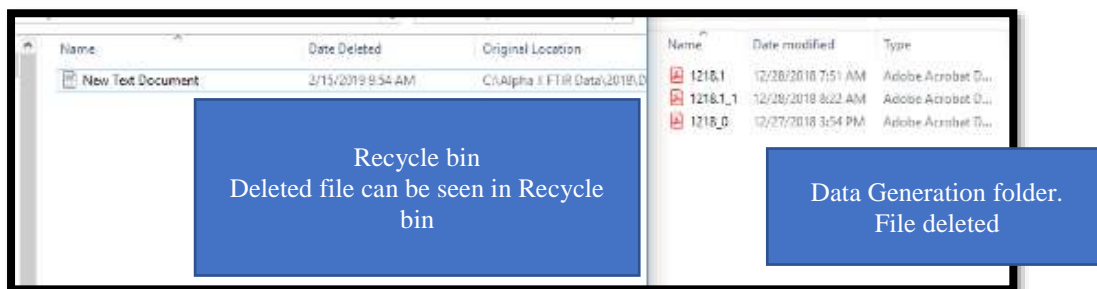


Figure 7. File deletion from windows before applying user restrictions

Windows operating systems offer the ability to track file deletions through the 'Windows Event Viewer', which can detect any queries. Windows codes such as '4660' and '4656' are used to search for and obtain details of deleted files in the event viewer security logs. The results from these queries will also include deletion of 'temp files', moving a file from origin location, renaming a data file name and other similar events on the computer as they run as a global query. Filtering only the file deletions related to a specific analytical instrument program can be extremely challenging. To prevent file deletions on Windows, companies should implement local controls to enforce user restrictions. For example, installing executable files that restrict file modifications (such as Cut/Copy/Paste/Delete/Rename/moving files) and blocking keyboard shortcuts for file deletions. Additionally, disabling the right-click option for the mouse or touchpad can help prevent accidental file deletions. This can be seen in **Figure 8**.

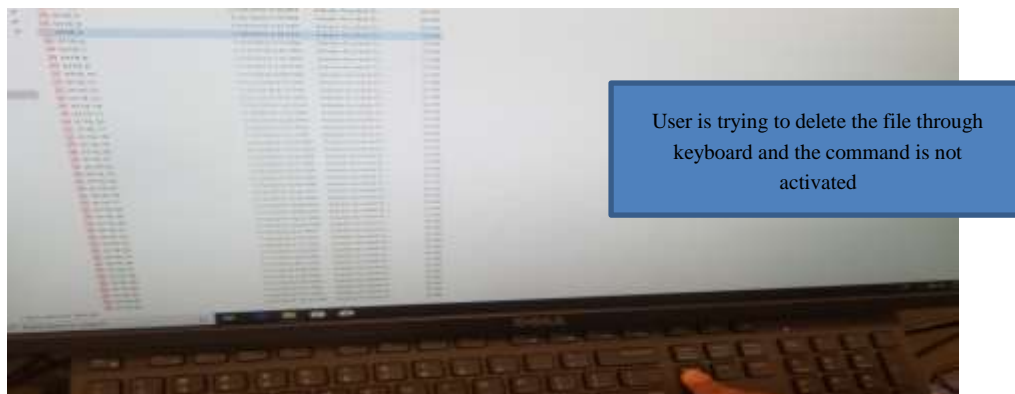


Figure 8. File deletion from windows - after applying user restriction

2.1.1 Case study

It is possible to delete files outside of the instrument software control by using explorer mode. There are two scenarios illustrated in **Figure 9**, regarding instrument software control.

In some cases, a data file with the same name or number may be generated twice on the hard disk, but only one data file is found. To prevent such deletions, it is advisable to lock the data wherever the software allows or implement appropriate user level controls. For added security, it is recommended to disable USB drive and CD/DVD port access on stand-alone computers related to instruments, which helps reduce the risk of data manipulation. Data generated on instruments connected to a server is stored on the server itself, while some instruments have an interim storage device from which data is periodically transferred to the server. Access to data modification on computers connected to instruments that are on the local server should be restricted

to administrators only. This can be achieved through Role Based Access Control (RBAC), a method that limits network access based on the roles of individual users within an organization. RBAC ensures that employees have access only to the information necessary for their job roles, preventing unauthorized access to irrelevant data. The evolution of software development has revolutionized work management in analytical labs, both onsite and offsite, enabling instant data viewing within the organization or even across continents. According to FDA guidance on Data Integrity, shared read-only user accounts are acceptable for viewing data, but they do not meet the requirements of part 211 and 212 for actions such as second person review to be attributable to a specific individual.

Some software store data generated off site on 'cloud' technology. The commencement of cloud-based services requires validation and an agreement to be in place defining the responsibilities of each party involved. 'PDA Technical Report 80' Data Integrity Management System for Pharmaceutical Laboratories details the cloud-based services validation. In general, the firm attaining services must ensure that the data is transferred securely and completely to the cloud ensuring end to end data integrity. If data exchange or data analysis is done on/through cloud, the activity should be validated, user controls and access levels to be defined appropriately.

2.2 Phase: 2 Data backup

The firm determines the frequency of data backup based on its own risk assessment. Many firms conduct data backup activities monthly for PLC-based systems and stand-alone systems. For server-based systems, data backup activities performed by tri-monthly/half-yearly/annual basis. According to FDA guidance on Data integrity, the term 'backup' is synonymous with 'archive' as defined in the General Principles of Software Validation²⁶. Regular data backups play a crucial role in data protection by limiting access.

It is important not to store multiple copies of the same data for security and regulatory reasons. Possible methods of data storage include USB for PLC devices, tape drives for servers, and mirroring folders for servers. Data backup for PLC-based devices should be saved onto permanent storage devices such as CDs or DVDs, with a minimum of two copies usually created. To mitigate risks and as part of a contingency plan, these two copies should be stored separately at different locations. Stand-alone computers present challenges for data backup, with common issues including untraceable or missing data in backups, incomplete data transfers, and corrupted data. Data backup is typically recommended using software-specific 'backup' commands to help identify any issues during data copying. A successful backup message from a spectrophotometric software is illustrated in **Figure 10**.

```
<Application Log Information>
Number of Media: 14/14
Source DB: Log DB
Destination DB: APL_LOG001.mdb
File Size: 0.6MB
File Date: 10-12-2018 12:05:15(+05:30)
Check Sum: ecabf4fd073da553d1fa325d661b73ec
Number of Backup Records: 657
```

Figure 10. Successful backup message

Some instrument software available in the market lacks this feature, leading firms to resort to using the simple 'copy and paste' command instead. Before deleting the copied data, firms conduct several checks to ensure its integrity. These checks include examining random data files from the copied data to verify complete and accurate copying, utilizing file/folder comparison software to identify any differences between the original and copied files, comparing the total number of files and size of the folders, and employing checksum numbers or check summation to guarantee a complete backup of the intended files. The FDA Glossary of Computer System Software Development Terminology defines check summation as a technique for error detection that ensures the accurate copying or transfer of data or program files. It involves summing groups of digits, such as a file, without considering overflow, and comparing the sum to a previously computed sum to verify accuracy. This technique contrasts with cyclic redundancy check (CRC) and parity check. Additionally, when standalone computers are connected on a local network for access control restrictions, a time-programmed data backup is established to prevent accidental data losses. **Figure 11**, displays a typical message from an automated data backup in audit trails.

Action	Details	Change Date	User
1. Finished Backup Process	Project: MAR_2024 Reason: Taken Project Backup	10-04-2024 12:00:51 IST	emp698MT_Coordinator
2. Project Integrity Successful	Project: MAR_2024 Reason: Taken Project Backup	10-04-2024 12:00:51 IST	emp698MT_Coordinator
3. Backed-up Project	Project: MAR_2024 Reason: Taken Project Backup	10-04-2024 12:00:51 IST	emp698MT_Coordinator
4. Started Backup Process	Project: MAR_2024 Reason: Taken Project Backup	10-04-2024 12:00:48 IST	emp698MT_Coordinator
5. Finished Backup Process	Project: MAR_2024QC_002 Reason: Taken Project Backup	10-04-2024 12:00:48 IST	emp698MT_Coordinator
6. Project Integrity Successful	Project: MAR_2024QC_002 Reason: Taken Project Backup	10-04-2024 12:00:48 IST	emp698MT_Coordinator
7. Backed-up Project	Project: MAR_2024QC_002 Reason: Taken Project Backup	10-04-2024 12:00:45 IST	emp698MT_Coordinator
8. Started Backup Process	Project: MAR_2024QC_002 Reason: Taken Project Backup	10-04-2024 12:00:32 IST	emp698MT_Coordinator
9. Finished Backup Process	Project: MAR_2024QC_002 Reason: Taken Project Backup	10-04-2024 12:00:32 IST	emp698MT_Coordinator
10. Project Integrity Successful	Project: MAR_2024QC_002 Reason: Taken Project Backup	10-04-2024 12:00:32 IST	emp698MT_Coordinator
11. Backed-up Project	Project: MAR_2024QC_002 Reason: Taken Project Backup	10-04-2024 12:00:31 IST	emp698MT_Coordinator
12. Started Backup Process	Project: MAR_2024QC_002 Reason: Taken Project Backup	10-04-2024 12:00:22 IST	emp698MT_Coordinator
13. Finished Backup Process	Project: MAR_2024QC_002 Reason: Taken Project Backup	10-04-2024 12:00:22 IST	emp698MT_Coordinator
14. Project Integrity Successful	Project: MAR_2024QC_002 Reason: Taken Project Backup	10-04-2024 12:00:22 IST	emp698MT_Coordinator

Figure 11. Automated backup message in audit trails

Some firms choose to encrypt the information in their backup or archives, thereby minimizing the potential risks associated with accidental or intentional loss or mishandling incidents. Analytical data stored in the cloud should be archived and transferred to the custody of the sponsor. Any problems discovered during the backup process are handled through the company's quality system and investigated as necessary. Once thorough checks are completed, the original data on the computer/server must be deleted.

2.3 Phase 3 – Data retrieval

This scenario, although hypothetical, involves the need for firms to retrieve and access archived data that is one year or several years old in electronic format using specific software for various reasons. These reasons include presenting to auditors (regulatory or customer), addressing regulatory deficiencies, handling customer complaints, and conducting internal investigations. Some situations may require data reprocessing, which should then be archived again and integrated into 'data lifecycle management'. Unfortunately, firms often treat this as a standalone task and lose sight of these events during data backup routines. It is advised that the custodian, most likely the quality assurance personnel responsible for data backup CDs/DVDs, keep a record of the data issued and document it properly to preserve all data versions under their care for compliance and future use. It is highly recommended to establish a documented procedure, as well as a tracking and tracing system for archived data until its retirement.

2.4 Phase – 4 Data backup and Destroy

Access to the restored and reprocessed data will be granted upon the successful completion of a specific task. It is imperative to repeat the steps outlined in both Phase-1 and Phase-2, followed by creating a backup of the data. One of the main challenges faced by firms in handling this type of data is the reprocessing required for one or a few injections or data files at different time points, making it difficult for firms to manage this in a traceable manner.

Data Tracking for any organization, it is crucial to monitor and trace the data travel and location. The quality and IT departments must collaborate to identify, verify, and validate the data flow from its creation to distribution. While there are several known scenarios in this process, unexpected situations may arise, leading to setbacks in the system. We can see some possible issues in the info graphic presented below Table1.

	Knowns	Unknowns
Known	Data deletions or manipulations detected by Audit trails or routine checks detected internally and addressed as per firms quality systems	Existing issues in the data handlings.
		Few errors in the system audit trails, system lockouts, null values, data not present.
		Should be communicated to instrument/software vendors and work on a risk mitigation plans.
	Issues can be present but not sure when they will happen	'No assignable cause found' situations.
	Connectivity errors	Example: Oracle errors, SQL errors, data base errors which may not affect the quality, integrity of the analytical work.
Unknown	Data file duplications	
	Multiple copies of data file [tapedrive, Hard disk, DVD.	
	Challenge the situations , validate before use	
	Retrospective validation for current software.	

2.5 Validation or fitness test for use

All instruments and software must undergo testing for their intended usage as per guidance and cGMP recommendations. The FDA guidance in emphasizes the importance of validating computer systems for intended use, such as creating an electronic master production and control record (MPCR). The validation process should be in line with the risk associated with the automated system to ensure proper workflow functioning. Failure to validate a computer system for its intended use may result in workflow errors. Validation scenarios may include data file format, database type (SQL, MS Access, Oracle), system audit trail recording, user level audit trail, file deletion restrictions, date and time change restrictions, data file locking, file copying and pasting controls, and file backup and archival procedures. It is recommended for firms to establish a risk assessment and validation protocol for all off-the-shelf (OTS) software to align with intended use and define data life cycle management applications and limitations. While achieving a fool-proof software may be challenging, local controls can be implemented by firms to address any short falls.

3. Conclusion

The management of electronic data involves various stages such as data generation, processing, storage, backup, retrieval, reprocessing, backup restoration, and dissemination. These stages are crucial in ensuring the proper handling and protection of data. The post data accumulation / generation and data handling are very important with the advent of computerization and most analytical instruments being hooked up to computers. The risks associated with data losses and data manipulations increases. To mitigate these risks, it is essential for firms to establish a procedure for risk assessment and software validation before implementing it in the cGMP (current Good Manufacturing Practice) environment. This ensures that the software used meets the necessary standards and requirements. In addition to risk assessment and validation, local controls and supporting software programs are necessary to maintain data integrity throughout the life cycle management of electronic data. These measures ensure compliance with ALCOA and ALCOA+ requirements, which are essential for data integrity and regulatory compliance.

CRedit authorship contribution statement

All the authors contributed significantly to this manuscript, participated in reviewing/editing and approved the final draft for publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors wish to extend their gratitude to the management of a ScieGen Pharmaceuticals Inc, NY, USA. and Chemeca drugs private limited, Andhra Pradesh, INDIA for supporting this work. They also thank their colleagues in the analytical laboratory for their cooperation in carrying out this work.

4. References

- [1] Food and Drug Administration: Data Integrity and Compliance with Drug cGMP Questions and Answers Guidance for Industry. (2018).
- [2] Parenteral Drug Association: Data Integrity Management System for Pharmaceutical Laboratories, Technical Report No.80. ISBN: 978-1-945584-05-3 (2018).
- [3] World Health Organization guidelines: Guidance on Good Data and Record Management Practices. WHO Technical Report Series No. 996, (2016).
- [4] GAMP 5 Guide: A Risk-Based Approach to compliant GxP computerized systems (second edition). (2022).
- [5] Medicines and Healthcare products Regulatory Agency:GMP Data Integrity Definitions and Guidance for Industry. (2015).
- [6] MHRA GxP Data Integrity Definitions and Guidance for Industry. (2016).
- [7] Pharmaceutical Inspection Co-operation Scheme: Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments, (2021).
- [8] Part 11, Electronic Records; Electronic Signatures-Scope and Application. (2003).
- [9] European Union:EudraLex, Volume 4: Good Manufacturing Practice, Annex 11: Computerised Systems. (2011).
- [10] U.S. Food and Drug Administration: 21 CFR Part 211–Current Good Manufacturing Practice for Finished Pharmaceuticals.
- [11] European Medicines Agency;Good manufacturing practice - Data Integrity: Questions and answers. (2016).
- [12] World Health Organization:Good Manufacturing Practices for Pharmaceutical Products: Main Principles. WHO Technical Report Series No. 986.(2014).
- [13] USP <1058>Analytical Instrument Qualification.
- [14] Peter D. Wentzell, Christopher D. Brown, Signal Processing in Analytical Chemistry. (2000) 9764–9800.
- [15] J.W. Cooley, J.W. Tukey, An Algorithm for the Machine Calculation of Complex Fourier Series. (1965) 297–301.

- [16] A.J. Diefenderfer, B.E. Holton, Principles of Electronic Instrumentation. 3 (1994).
- [17] A.V. Oppenheim, R.W. Schafer, Discrete-Time Signal Processing. (1975).
- [18] A.G. Marshall, F.R. Verdun, Fourier Transforms in NMR, Optical and Mass Spectrometry, a User's Handbook. (1990).
- [19] P.J. Treado, M.D. Morris, A Thousand Points of Light: The Hadamard Transform in Chemical Analysis and Instrumentation. 61 (1989).
- [20] FDA: Glossary of Computer System Software Development Terminology (8/95). (2019).
- [21] American national standard for information systems-dictionary for information systems. New York : ANSI/Secretariat: Computer and Business Instrument Manufacturers Association. (1991).
- [22] Special Publication 800-12; An Introduction to Computer Security-The NIST Handbook, Audit trails. (1995).
- [23] U.S. Food and Drug Administration: Part 211 Current Good Manufacturing Practice for Finished Pharmaceuticals. (1995).
- [24] M.J. Cahilly, Data Integrity Training Lessons Learned and Case Studies. (2015).
- [25] U.S. Food and Drug Administration: Warning Letters by year, 1998–current. Warning Letters and Notice of Violation Letters to Pharmaceutical Companies. (2017).
- [26] General Principles of Software Validation; Final Guidance for Industry and FDA Staff. Centre for devices and Radiological health. (2002).

Figures:

Figure -1:

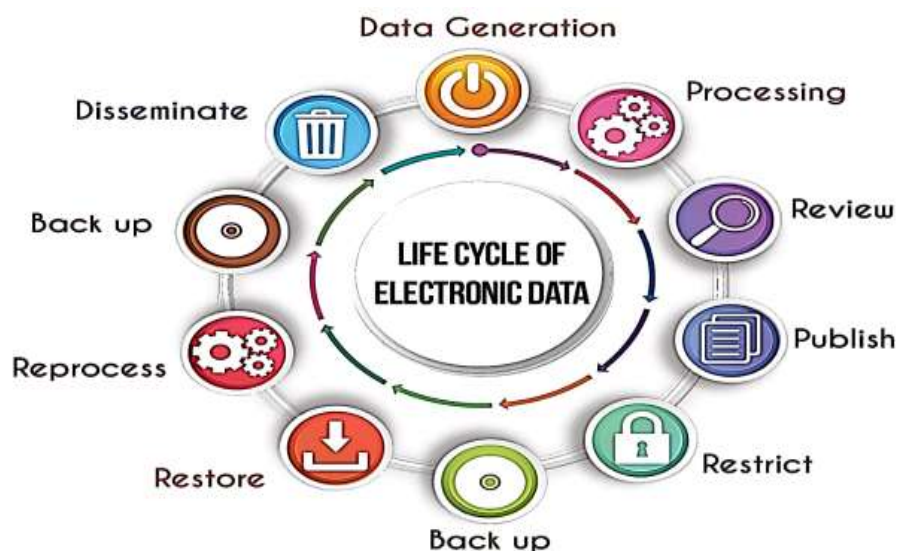


Figure-2:

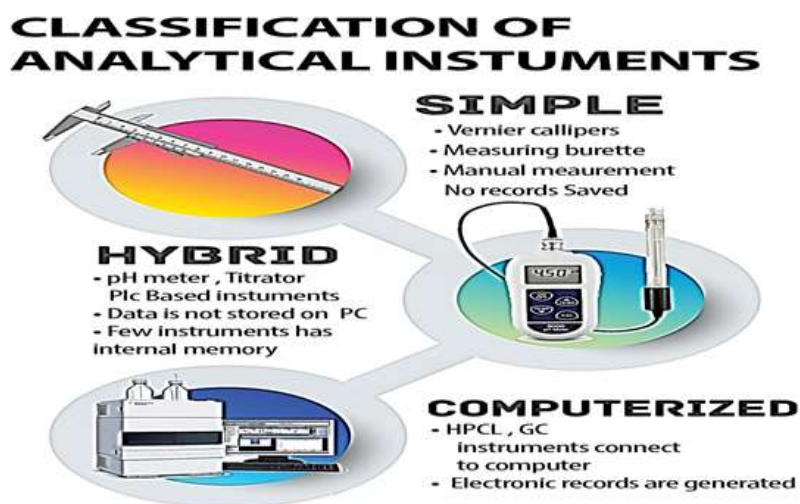


Figure-3:

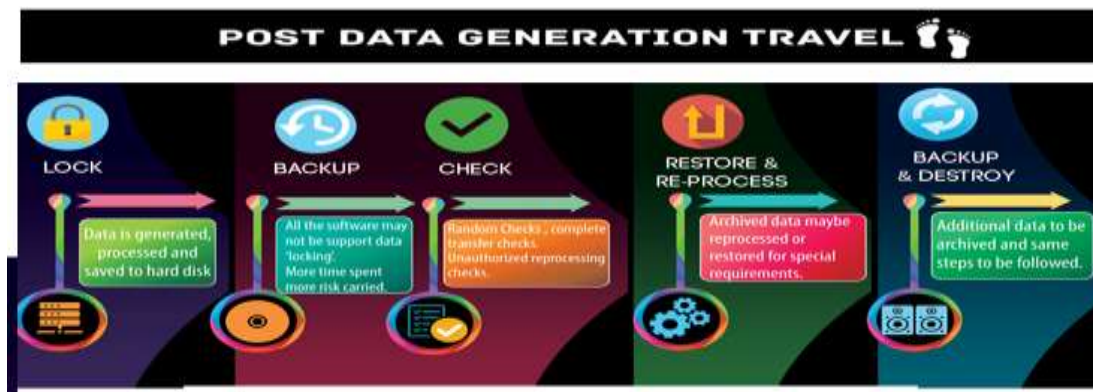


Figure-5:



Figure-9:

