IDBS

The evolving role of the GxP Archivist:

ensuring data protection and integrity among the changing demands new technologies bring

Stuart Ward, Bob McDowall and Damien Tiller

Customers and regulators expect the highest level of data integrity, and many organizations rely on an archivist to maintain and protect this most important asset.

This need to protect the data encompasses intellectual property and/or regulatory data integral for marketing submissions. Therefore, ensuring the archivist can guarantee adequate protection and control for regulatory data is critical for organizations to meet their business goals. According to the AGIT Guidelines for the Archiving of Electronic Raw Data in a GLP Environment published in 2018, “the archivist is responsible for all aspects of electronic archiving and should have full control of all activities within the archiving process. If there is external IT involvement (e.g. a service provider), the archivist has to ensure that the procedures are followed as described in the relevant service level agreements or contracts.”

But will regulators accept the storage of Good Laboratory Practice (GLP) data using cloud technology?

This white paper will assert that storing GLP data in the cloud is acceptable, provided certain conditions are met. Using a cloud-based system provides the following benefits:

Rapid retrieval
Efficient and controlled data access without being limited to having to visit a specific physical archive; located either on site or retrieval from an off-site repository.

Security
Tighter security controls and clear separation of duties through the entire data management process.

Complete Data
Ability to capture the ever-increasing amounts of “complete” data, required for intellectual property, regulatory and business purposes without significant storage changes.

Record retention
A clear understanding of where the data is located for regulatory/legal purposes.

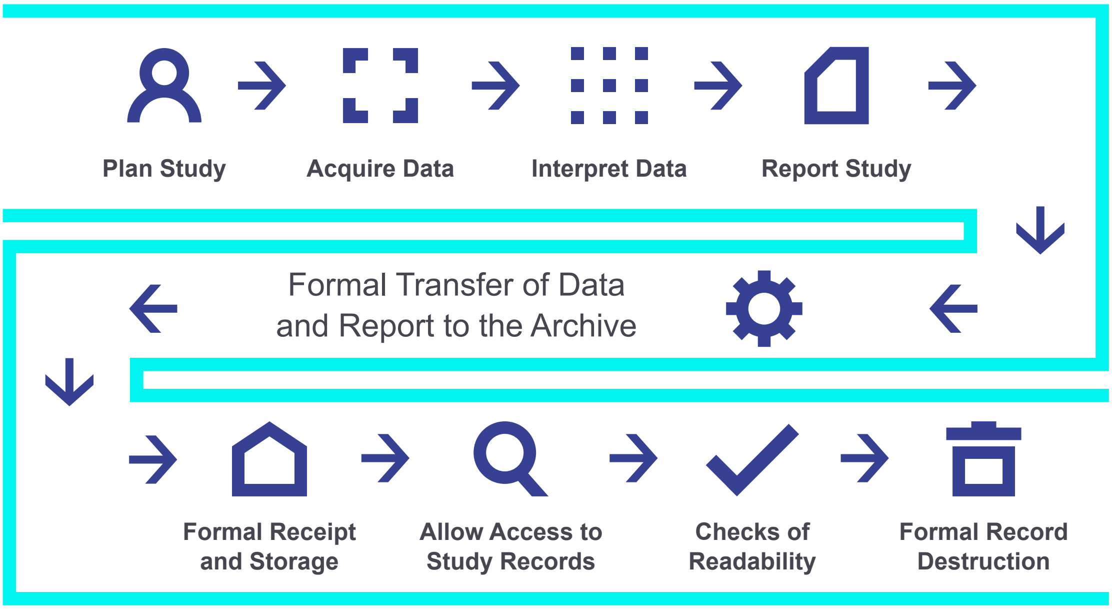
Backup
Tested backup and disaster recovery processes.

To demonstrate this, we begin with the regulatory requirements for a GLP Archivist that are found in the OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring - Number 15: Establishment and Control of Archives that

Operate in Compliance with the Principles of GLP (2007):

- ensure that access to the archive is controlled;
- ensure that the orderly storage and retrieval of records and materials is facilitated by a system of indexing; and
- ensure that movement of records and materials in and out of the archives is properly controlled and documented.
- location

Distilling this down further, the archivist is responsible for ensuring that the data integrity and retention period needs are met so that the archived data can be retrieved for future decision making and regulatory review with minimal delay.


Plan Study
Acquire Data
Interpret Data
Report Study
Formal Transfer of Data and Report to the Archive
Formal Receipt and Storage
Allow Access to Study Records
Checks of Readability
Formal Record Destruction

The changing technology landscapes

With new technologies such as Infrastructure as a Service (IaaS) / Software as a Service (SaaS) and cloud computing enabling laboratories to generate data in unprecedented numbers, the data volumes have increased while budgetary pressures have remained constant. Organizations have looked at how they provide data management systems to their users. In many cases, this has involved using the public cloud, for example IaaS as provided by Amazon Web Services (AWS) or Microsoft Azure, to host their systems and data including data used for regulatory purposes. This inevitable move to the cloud has further put a strain on companies and regulators as they try and interpret regulations that, in most cases, were created with paper records in mind. For example, the GLP document referring to data archiving mentioned above was published in 2007.

Ensuring regulatory compliance against differing interpretations of the same regulations has resulted in archivists having a number of dilemmas.

The main challenge: how are the regulations being interpreted and what will happen to the studies that are performed using systems running on cloud technology – will the regulators accept them?

For some regulations, like Good Manufacturing Practice (GMP), there are clear statements that drive organizations to use the best technology to meet their business requirements such as the US Food and Drug Administration's (FDA) Technology Modernization Action Plan (2019).

For GLP, the regulations such as OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1: OECD Principles on Good Laboratory Practice (1998) and OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 17: Application of GLP Principles to Computerised Systems (2016), whist the more recent document does mention using service providers to deliver the computerised system, no document explicitly mentions the use of the cloud. Therefore, many regulators and businesses are interpreting the GLP regulations without any clearly published guidance from the regulatory authorities. This makes cloud use difficult. This conservative stance is counterintuitive when many organizations want to use the cloud to provide technology that can meet the ever-changing business needs and maintain competitive advantage. It is estimated that the use of cloud technology in just the life science industry will grow at a compound annual growth rate of approximately 13% over the next five years!

As mentioned above, many of the GLP principles were written when the cloud was not readily available or envisaged as a technology that would provide significant benefit for activities like life science data management. When some of these principles were published, much of the regulated data was still being collected using paper-based or hybrid processes. Therefore, it is understandable that the principles were primarily written with a paper process and a physical archive in mind.

The benefits of electronic data management systems

Since then, many organizations have moved to electronic data management systems, for example electronic laboratory notebooks (ELNs) and laboratory information management systems (LIMS), to streamline their data management workflows, to enable rapid searching and retrieval of records that show complete data and to provide better ways of meeting the regulatory requirements and the ever increasing focus on data integrity.

Electronic data management can have increased functionality such as audit trails and user authentication over previous paper records and associated archiving methods. These electronic systems that tend to be more secure, have better capabilities to maintain data integrity, for example the ability to see the complete history of a record, and improve data accessibility i.e. the user does not need to visit the physical location of the archive. Instead, they can access the data from a computer terminal whilst preserving data security, as mentioned in the requirements stated above. For example, for some systems, the archive is now a logical partition, with read-only access, in the application which is also capturing the data.

Cloud Technologies have the advantage that all the organization's data can be searched and reported-on in one place so comparisons between old and new data can be easily made. In addition, any relationships between captured data can be easily preserved, which is not always the case when the archived data is exported to a separate system. Ultimately, having this combined data accessibility and integrity has resulted in streamlining business processes, including regulatory audits and submissions.

When ELNs, LIMS and other electronic data management applications first came to market there was hesitation around initial adoption because, similar to the situation with the cloud now, organizations were nervous around how the regulators would evaluate this technology against the regulations without explicit guidance.

Up to this point, the various bodies responsible for the GxP regulations have stated that the existing regulations/principles should be followed when using electronic systems. Over time, regulators and organizations have begun to appreciate and understand the benefits that electronic systems can provide, for example for data integrity purposes, and their use is now widespread through the life science industry. Additionally, and highlighted by the COVID-19 pandemic, the ability to perform remote audits means data and studies can continue adding benefits to business continuity.

The amount of data that organizations are wanting to collect is constantly increasing and the expectations of regulated bodies auditing the studies is that all data is safely stored and can be easily accessed for review. To support this expectation, organizations are looking for new technology solutions to optimize data storage and reduce their costs. The main technology being considered by many organizations is the cloud, which can provide scalable resources such as storage and compute power more efficiently than with on-premise data centres, to meet these changing needs.

However, during discussions with other organizations and quality professionals, there are some perceived hurdles when the system is being used to store (GLP) regulated data, such as: OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1: OECD Principles on Good Laboratory Practice (1998) § 9.2.7 Storage. Specifically, organizations and some regulators believe that the street address of the cloud data centre needs to be known and some also question whether the record retention requirements can be met when using the cloud. Once the setup and benefits of the cloud are understood these concerns can be easily allayed, as organizations know where their data is stored and can also control the retention period. More importantly, electronic data can be retrieved rapidly from an off-site hosting centre far faster than physical retrieval of paper records from a records management company.

Location and security of data

As mentioned above, organizations and regulators want to understand where the data is stored and, for GLP studies, this information is required to be disclosed in the final report/submissions to regulators. The GLP principle § 9.2.7, in OECD document No 1 published in 1998, states: “The location(s) where the study plan, samples of test and reference items, specimens, raw data and the final report are to be stored.”

It should be noted that there is no definition of what location means and it is, therefore, down to interpretation. Meaning this could be the cloud URL, the country or geographical location and does not necessarily dictate the computer ID or postal address.

When this particular regulation was published in 1998, the majority of GLP studies were written and/or published on paper. This meant that having the street address of where the study report was archived would be extremely useful so that the report could be viewed for inspection. This would also allow the auditors to review that the report was stored in an appropriate manner as to preserve data integrity and ensure that the data can be viewed for the duration of the retention period.

As organizations have moved to digital solutions, the focus has changed in that auditors now need to know where a computer terminal is available and where data and study reports can be retrieved rapidly for review. Auditors also need to understand whether the data centre being used has the appropriate controls in place to ensure that data integrity and retention periods can be met. For on-premise software deployments, where separate controls underpinned by ISO 27001 Information Security Management and/or Service Organization Control 2 certification, reports on various organizational controls related to security, availability, processing integrity, confidentiality or privacy (SOC-2) may or may not be in place.
Up to this point it has been the standard that these environments are inspected as part of any audit to confirm suitability.

Cloud providers put considerable effort into preserving the security of the cloud – this is their primary focus, and organizations such as AWS (2020) have a plethora of leading security standards such as SOC 1, 2 and 3 and ISO 27001. However, as organizations have moved their deployments to the cloud, it has become clear that most cloud providers will not disclose the exact location of their data centres. The reason for this is simple: security. It is possible to suggest that by providing public knowledge of the data centre location the likelihood of a focused attack increases and hence security is compromised, which would have a big impact on the cloud business.

Therefore, cloud providers, such as AWS, do provide the regional location of where the data is stored, and organizations can select which regions are used, for example London, Paris, Frankfurt, Virginia USA, etc. The use of the regions provides organizations with the control and the details of which country and, ultimately, which legislation/ regulations the data falls under.

In addition, many cloud providers and software vendors have been audited and demonstrated that they meet the security requirements for ISO 27001 and/or SOC-2 certification. This is explored more in the white paper “Is the Cloud a Safe Place for your Data?” (2020). However, standards such as SOC can be used to provide confidence that the infrastructure and software that is being used to store the data meets – and in many cases exceeds – the requirements described in the various GxP regulations since many of the common security standards have been updated to take into account the latest technology available. In particular, ISO 27001 certification for a data centre provides better assurance of security than some physical archives, such as:

Physical security for site and vehicles, biometric security for staff, and visitors have identity checks.

Logical security of the IDBS IT function that supports each virtual instance overlaid with the customer’s user account management.

Cyber security is vital to protect the study data

Clear separation of duties

Authorized access prevents data centre staff from accessing customer data.

Environmental controls

Redundancy in environmental systems that are operational and tested.

It is possible for organizations to get further benefit if software applications have been optimized for cloud are also used.

It is very common for these types of optimized applications to be provided as SaaS to be delivered and managed by the vendor, who in many cases has also created the software in the first place i.e. is an expert in the setup and running of the application. From a regulatory perspective, this type of arrangement can further enhance data integrity and security. There is a clear separation of duties. This starts from the cloud provider (such as AWS) having different personnel maintaining the hardware versus the logical systems running the infrastructure. The personnel maintaining the physical hardware also do not know what is stored on each machine and have no access to the data. Likewise, those running the logical systems do not have physical access to the hardware. In addition, the vendor supplying the software has no access to the cloud infrastructure other than through the IT infrastructure templates and systems that the cloud provider supplies to manage the infrastructure and there should be mechanisms in place to limit the access to customer data, for example, named users for customer agreed investigation purposes.

Finally, the organization using the software will have no access to the application servers/databases setup for the software. Ultimately this means that each party can concentrate on their “specialty” and the security associated with it i.e. the cloud provider maintains the infrastructure, the SaaS vendor can concentrate on the running of the software and finally customer can concentrate on the business process that they are trying to achieve.

Disaster recovery

Another aspect of data security is what happens if a data store fails and/or other issues with the application. An essential element of the application's deployment are mechanisms to deal with issues and be able to recover quickly. This starts with a real-time replication of the primary database so that if an issue occurs there will be automatic failover to the secondary database and minimal impact to the system users. Then systems may have replication across 'availability zones' so a copy of the data is stored in a completely different location. Therefore, providing a restore point if for whatever reason the primary data store location becomes compromised. On top of this, because the storage is easily scalable, the number of backup points can be appropriately defined to meet the organizational needs.

These capabilities, whilst needing to be configured, are part of the standard capabilities of many of today's large cloud providers. Overall, this means that the disaster recovery capabilities of applications optimized for running in the cloud will exceed the provisions that can be practically implemented with an on-premise software deployment and hence further increasing data security.

Record retention

As mentioned above, one concern a GLP regulator might have is the use of the cloud and the impact on record retention. However, the considerations do not seem to be different from if the software is deployed using cloud or on-premise technology. The reason for this statement is that contracts either with the SaaS/cloud vendor or corporate IT suppliers will determine the lifespan of the systems to meet the needs of record retention policies. Therefore, the record retention timespan can easily be defined as part of the software and/or cloud procurement process.

Summary

The question was asked if the regulators will accept the storage of data using cloud technology.

GMP regulations accept the new technology and GLP regulations do not prohibit the use of the cloud. Therefore, any reservations are limited to the interpretations and assumptions that are being made. Reviewing the benefits that applications running on the cloud can provide from security, disaster recovery and the ability to easily access the (archived) data, it is arguable then that it would be difficult to envisage that the use of the cloud to store GxP data will not be accepted.

Organizations need to perform their own due diligence carrying out robust vendor assessments and verification and validation activities to see if they are ready to use applications running on the cloud, however the evidence is clearly showing it is very much the future. The regulators today expect to see complete data stored appropriately to maintain data integrity and these expectations are only going to become higher. Using the right technology will make the processes that organizations need to follow easier and therefore they will have more time to concentrate on the science.

Not having a plan to adopt the cloud may mean losing competitive advantage as well as, with the ever-increasing amounts of data and the extra demands on audits, making the role of the archivist more challenging.

OUR SOFTWARE FITS THE BILL.
CONTACT US TO FIND OUT MORE.

IDBS

info@idbs.com
www.idbs.com

UK (HQ)

Tel: +44 1483 595 000
2 Occam Court,
Surrey Research Park
Guildford, Surrey, GU2 7QB

USA

Tel: +1 781 272 3355
285 Summer Street
Fifth Floor, Boston,
MA 02210

BANGALORE INDIA

43, Residency Rd, Shanthala Nagar,
Ashok Nagar, Bengaluru,
Karnataka 560025 India
india@idbs.com

CONNECT WITH AN EXPERT