

The State of Cloud-Based Biospecimen and Biobank Data Management Tools

Shonali Paul, Aditi Gade, and Sumani Mallipeddi

Biobanks are critical for collecting and managing high-quality biospecimens from donors with appropriate clinical annotation. The high-quality human biospecimens and associated data are required to better understand disease processes. Therefore, biobanks have become an important and essential resource for healthcare research and drug discovery. However, collecting and managing huge volumes of data (biospecimens and associated clinical data) necessitate that biobanks use appropriate data management solutions that can keep pace with the ever-changing requirements of research. To automate biobank data management, biobanks have been investing in traditional Laboratory Information Management Systems (LIMS). However, there are a myriad of challenges faced by biobanks in acquiring traditional LIMS. Traditional LIMS are cost-intensive and often lack the flexibility to accommodate changes in data sources and workflows. Cloud technology is emerging as an alternative that provides the opportunity to small and medium-sized biobanks to automate their operations in a cost-effective manner, even without IT personnel. Cloud-based solutions offer the advantage of heightened security, rapid scalability, dynamic allocation of services, and can facilitate collaboration between different research groups by using a shared environment on a “pay-as-you-go” basis. The benefits offered by cloud technology have resulted in the development of cloud-based data management solutions as an alternative to traditional on-premise software. After evaluating the advantages offered by cloud technology, several biobanks have started adopting cloud-based tools. Cloud-based tools provide biobanks with easy access to biospecimen data for real-time sharing with clinicians. Another major benefit realized by biobanks by implementing cloud-based applications is unlimited data storage on the cloud and automatic backups for protecting any data loss in the face of natural calamities.

Keywords: biobank, biospecimen, data management challenges, research, cloud-based tools

Introduction

IN THE ERA of personalized medicine, biobanks are critical for collecting and managing high-quality biospecimens from donors, including proper clinical annotation.¹ High-quality human biospecimens and associated data are required to better understand the biology of patients and their diseases.¹ This creates a huge demand for quality human biospecimens among the research community. The type of biospecimens selected by researchers depends on the quality framework established by a biobank.

The human biospecimens must be collected and processed based on standards that support quality, in terms of both storage and clinical annotation of relevant patient and biospecimen information. Furthermore, the use of high-quality, well-annotated biospecimens by researchers typically depends on two factors, their availability and the organization of associated clinical data, including medical, genealogical, and lifestyle information in a biobank.²

Therefore, biobanks have become an important and essential resource for healthcare research and drug discovery.

Today, biobanks are regarded as data repositories for raw data (the unprocessed samples), data associated with samples (processing and storage conditions), and supplementary data (such as clinical annotations). Collecting such huge volumes of data requires biobanks to use appropriate data management solutions that can keep pace with the ever-changing requirements of research.

Maximizing the use of biobanks

Biobanks play a pivotal role in elucidating disease etiology, translational research, and advancing public health and patient care.³ However, to meet these objectives, there is a need for building a global biobanking strategy.

Moreover, the cooperation and extensive collaboration among a network of multiple organizations such as biobanks, research institutes, governmental bodies, funding agencies,

public and private science enterprises, and other stakeholders, including patients, must be encouraged to enable the streamlined exchange of biospecimens and associated data. To accomplish this level of cooperation, biobanking informatics solutions are essential. These informatics solutions facilitate managing and sharing very large amounts of data. There are several powerful traditional on-premise data management tools available in the market, assisting biobanks in managing and tracking their data. However, traditional tools are unable to completely address the challenges that the biobanks encounter as they scale up their operations.

Materials

Data management challenges faced by biobanks

Some of the critical challenges faced by biobanks include the following.

Big data management. Biobank data—which include both biospecimen-derived data from multiple sources and the associated health data—are increasing at a rapidly unpredictable rate.⁴ This increase can be attributed to rapid progress in scientific advancements and development of high-end technologies such as microarrays and next-generation sequencing. Besides this, the available biospecimen data are multifactorial, both large in size and heterogeneous in context. This results in a myriad of data management challenges, which are, by nature, high volume, variety, velocity, etc.

These challenges need to identify novel ways to store, manage, and process big data. Thus, it further presents a compelling need to develop an innovative scalable big data infrastructure that can enable healthcare providers access knowledge for every patient, yielding better decisions and quicker outcomes.

Switching to cloud-based biobanking informatics solutions provides a flexibility to scale up or down based on operational need, thereby minimizing large capital expenditures at the very outset. Furthermore, outsourcing biobank data storage services to commercial cloud service providers (CSPs) can relieve researchers from encumbrances such as costs incurred in establishing an IT infrastructure and managing an in-house software, including a significant measure of security oversight, potentially making research more cost-effective.

Managing clinically annotated biospecimens. Optimal value and utility of any biospecimen requires accumulated and meticulous annotation of relevant clinical data obtained from a patient.⁵ All the data surrounding procurement, processing, and distribution of the biospecimen right from collection to stabilization and storage are diligently tracked and recorded.⁵ This process demands a great effort in terms of personnel and time, and requires a data collection tool that eases operations such as managing and tracking biospecimen inventory in real-time, patient consent forms, and linking biospecimens with accurate clinical and diagnostic data.

Sharing of data sets in real time to facilitate collaboration. Research-focused biobanks involve collection of human tissues linked with genetic, genealogical, health, and personal information (demographics, etc.), which are used for a number of research purposes. From several research projects, a multitude of diverse data sets are extracted. Research-focused biobanks facilitate research efforts, as

clinical researchers do not have to expend their valuable time and funds on collection, storage, and curation of human tissues and data.⁶ In addition, the research capabilities of biobanks are further enhanced by combining or sharing equivalent data sets from other biobanks, provided they follow uniform standards in the way biospecimens are collected, extracted, and coded, appropriately taking into account the ethical, legal, and social implications.

There is an important need across the globe to make more effective use of tissues, blood, and other biospecimens donated by patients for research.² The only possible way to fully understand the biology of a disease, including the mechanisms of relapse, recurrence, and resistance, is when researchers have access to biological materials and are capable of linking them with other data sets, including patient clinical data and the cause of the disease.⁷

However, unfortunately, failure in sharing biospecimen data and hoarding of precious biomaterials are leaving clinicians and patient advocates increasingly frustrated, leading to lack of coordination in research.⁷ As a result, patients who consent the use of their tissues and/or blood products often have no clue whether their biospecimens have been effectively utilized by a single research team or a pharmaceutical company.⁸ Lack of collaboration is, therefore, a principal factor hindering progress in treating diseases.

Stringent ethical and regulatory compliance. The data available in biobanks include personal/protected health information (PHI), that is, both biospecimen data and patient-associated phenotypic and clinical data. Several federal laws and guidelines implemented by biobanks dictate regulated biospecimen collection, use, and disclosure of PHI.⁹ Typically, biobanks need a donor's consent for disclosing the PHI to third-party researchers and clinicians. Sharing this information through the web raises a number of privacy/security concerns such as risk of unauthorized use of data, loss of control, a multitenant environment, and lack of clear service level agreements.¹⁰ Besides these factors, privacy legislation such as the EU Data Protection Directive and the U.S. Health Insurance Portability and Accountability Act (HIPAA) have mandated strict requirements for processing sensitive personal health data.¹⁰ All of these factors impede the biobank data management and, in turn, drug discovery.

Disaster (natural or man-made) management. Complexity and costs of creating biobank infrastructures are increasing as the need for high-quality specimens and biobanking practices becomes more important for precision medicine and public health. Every year, across the globe, biobanks are established with new and advanced technologies incorporated into their operations. At the same time, there is also an increase in the incidence of natural and man-made disasters across the globe.¹¹

To cope with such unpredictable circumstances, biobanks must ensure safe inventory and data management systems. During a disaster, a biobank's fundamental objective is to protect every stored biospecimen and whole collections including related electronic data, in addition to safeguarding the employees. Often referred to as “disaster planning” or “preparedness,” methods encompass all efforts to ensure employee safety and reduce negative impacts such as loss of business operations and property damage. A key component for successful preparedness is a commitment to planning from an organization's management team and financial support for the preparedness program.¹² However, the current

trends of saving a biobank's data using paper-based formats and spreadsheets do not largely support the planning for disaster management, leading to loss of biospecimen data, and thereby limiting research frontiers.

To overcome the mentioned challenges and automate biobank data management, many biobanks have invested in traditional Laboratory Information Management Systems (LIMS). However, these biobanks are unable to manage sustained laboratory operations while acquiring traditional LIMS because of their major drawbacks. Traditional LIMS are cost-intensive, do not support real-time data sharing, and often lack the flexibility to accommodate changes in data sources and workflows as demanded by the ever changing needs of biobanks and research projects. As a result, biobanks end up using spreadsheets, laboratory notebooks, and other in-house software for storing and managing biospecimen data. These conventional methods are ineffective as they are typically error prone, time consuming, and require manual intervention, making data retrieval difficult.

Cloud-based tools enable disaster recovery by enabling automatic data backups, mirror servers used to host data, data storage at remote locations makes it possible for electronic data to be protected in case of fire, theft, and in case of natural calamities such as hurricanes, tsunamis, earthquakes, or floods.

Methods

The effective solution: Harnessing cloud technology to overcome biobanking data management challenges

Recent advances in information technology and genome sequencing technology have triggered significant changes in the ways science is carried out and provided a means to share data on a wider scale. The drawbacks and challenges posed by traditional methods and on-premise data management tools led to an emerging alternative known as "Cloud Technology," that is effective in addressing all these challenges.

Cloud technology generally refers to remote networking business practices that offer scalable, on-demand access to a configurable pool of computing resources (e.g., networks, servers, storage, applications, and services) over the web. Cloud computing is low cost, quickly deployable, requires no hardware, includes automatic upgrades, available as a

monthly subscription service, and easily accessible with a simple login similar to Gmail or Facebook. The benefits offered by cloud technology have resulted in the development of a cloud-based data management solution as an alternative to traditional on-premise software.

Currently, several R&D organizations are rapidly adopting cloud-based IT infrastructures as they externalize a growing segment of operations spanning product research and development. This is particularly evident in the life sciences (pharmaceutical and biotech) industries where companies are moving from centralized, corporate-based facilities to a virtual network of contract research, development, and manufacturing organizations (CRDMOs), and/or academic institutions and corporate partners that enable product discovery and development.¹³ The migration to an external research ecosystem has been driven by the economic realities of R&D over the past decade and the total cost of ownership advantages of cloud-based IT infrastructures over traditional models.

Great efforts have been invested by programmers and CSPs in employing a wide range of security mechanisms to enhance data privacy and make cloud platforms more reliable. Several techniques have been used, including data encryption, trusted platform module, secure multiparty computing, homomorphic encryption, anonymization, container, and sandboxing technologies.¹⁴

Why consider cloud-based solutions for biobank data management

Cloud-based solutions offer the advantage of heightened security, dynamic allocation of services, flexible costing, and quick deployment. Rapid scalability of the cloud offers data storage elasticity as one's operation needs to grow and thereby enables efficient big data management, while maintaining high data accuracy for biospecimen data. Cloud storage is secure and reliable as it maintains data integrity with periodic, automatic data backups at the server level during natural or man-made disasters, ensuring no data loss. The digital information can be deposited on the cloud and securely shared with distantly located research groups on a "pay-as-you-go" basis to facilitate real-time collaboration, which otherwise is not possible with on-premise LIMS or paper-based systems.

TABLE 1. LIST OF SOFTWARE VENDORS OFFERING CLOUD-BASED LABORATORY INFORMATION MANAGEMENT SYSTEMS FOR BIOBANKS

<i>Vendor</i>	<i>Software</i>	<i>Description</i>
CloudLIMS	BioTracer CloudLIMS Lite	Sample management software and LIMS simplify laboratory operations in biobanks/biorepositories/biostorages, research organizations, clinical laboratories, pharmaceuticals, and hospitals, by managing, tracking, and efficiently organizing samples and freezer locations at a very affordable price.
LabLynx	WebLIMS	A full-function LIMS that integrates with instruments, Electronic Medical Records (EMRs)/Electronic Health Records (EHRs), and hand-held and tablet devices. It supports access from multiple locations and maintains audit trails for all types of laboratories.
RURO	FreezerPro	Sample management solution that can be used by any modern scientific or pharmaceutical laboratory.
Blaze Systems	BlazeLIMS	Comprehensive and configurable features to support a wide variety of laboratory functions to ensure maximum benefits and minimum costs.

Cloud technology offers the following benefits:

- No operational expenses: There is no need to purchase, set up, and maintain hardware.
- Secure and reliable: Software vendors provide high-end security measures with encrypted data transmission between the client and the server, packet sniffing, IP spoofing, malware protection, and access-based credentials.
- Access using any browser, anytime anywhere as used in Internet banking, Gmail, Facebook, and Twitter.
- Extenuates barriers to instrument integration and data sharing.
- Easy configuration at no extra cost, to mirror laboratory workflows, without the need to invest in costly customization.
- HIPAA-compliant cloud services to ensure data privacy at the application level, thereby restricting unauthorized access to sensitive PHI.
- Uninterrupted biobanking for speeding up research.
- Reliability and stability in the eventuality of system breakdown or server failure.

Available cloud-based biobank data management vendors

Software vendors offering cloud-based biobanking informatics solutions utilize third-party commercial cloud-based infrastructures and services (such as Amazon Web Service, Microsoft Azure, Rackspace, and GoDaddy) to meet the modern evolving needs of the life sciences industry. CSPs provide a strong base that manages the volume and velocity of big data. What is required to effectively manage an externalized research operation is a highly flexible data management platform with the capability of managing a variety of R&D data (e.g., cell-based assays, absorption, distribution, metabolism, and excretion, toxicology, pharmacokinetics, animal studies, gene expression, proteomics, next-generation sequencing) while providing collaboration tools that enable remote access across a range of devices for researchers, and limited/restricted access for partners (i.e., CRDMOs). Recent “platform-based” IT architecture initiatives have made cloud-based informatics a reality for many organizations (Table 1).

Discussion

A shift from traditional methods to cloud-based tools

Evaluating the advantages offered by cloud technology, biobanks gradually have begun switching to cloud-based applications. Cloud-based tools provide biobanks with easy access to biospecimen data for real-time information sharing between clinicians and researchers. Furthermore, cloud applications have improved the efficiency of regulatory committees to approve the use of biospecimen data online rather than using paper-based methods. The need for setting up internal systems that need to be implemented, run, and maintained is no longer required. Other major benefits realized by biobanks with the implementation of cloud-based tools are low cost at an affordable monthly subscription,

unlimited data storage on the cloud, automatic data backups for preventing any data loss, and availability of mirror servers for data recovery in the face of natural calamities, which are all impossible to accomplish using traditional or in-house LIMS.

Author Disclosure Statement

No conflicting financial interests exist.

References

1. Kinkorová J. Biobanks in the era of personalized medicine: Objectives, challenges, and innovation: Overview. *EPMA J* 2015;7:4.
2. Sarojini S, Goy A, Pecora A, et al. Proactive biobanking to improve research and health care. *J Tissue Sci Eng* 2012; 3:116.
3. Harris JR, Burton P, Knoppers BM, et al. Toward a roadmap in global biobanking for health. *Eur J Hum Genet* 2012; 20:1105–1111.
4. Noor AM, Holmberg L, Gillett C, et al. Big Data: The challenge for small research groups in the era of cancer genomics. *Br J Cancer* 2015;113:1405–1412.
5. Macmanus LM, Mitchell RN. *Pathobiology of Human Disease: A Dynamic Encyclopedia of Disease Mechanisms*. Amsterdam, the Netherlands: Elsevier; 2014.
6. Chalmers D, Nicol D, Kaye J, et al. Has the biobank bubble burst? Withstanding the challenges for sustainable biobanking in the digital era. *BMC Med Ethics* 2016;17:39.
7. McIntyre P. Unlocking progress: Why we need to change the culture of biobanking. *Cancer World*, June 21, 2016.
8. Olson S, Berger AC. *Establishing Precompetitive Collaborations to Stimulate Genomics-Driven Product Development*. Washington (DC): National Academies Press (US); 2011.
9. IOM (Institute of Medicine). *Future Uses of the Department of Defense Joint Pathology Center Biorepository*. Washington, DC: The National Academies Press; 2012.
10. Gholamia A, Lind A-S, Reichel J, et al. Privacy threat modeling for emerging biobank clouds. Elsevier 2014;37: 489–496.
11. Henderson MK, Simeon-Dubach D, Zaayenga A. When bad things happen: Lessons learned from effective and not so effective disaster and recovery planning for biobanks. *Biopreserv Biobank* 2013;11:193.
12. Mintzer JL, Kronenthal CJ, Kelly V, et al. Preparedness for a natural disaster: How Coriell planned for hurricane Sandy. *Biopreserv Biobank* 2013;11:216–220.
13. Gwynne P, Heenner G. Laboratory technology trends: Drug discovery: -2, March 29, 2002.
14. Gholami A. *Security and Privacy of Sensitive Data in Cloud Computing Diss*. Stockholm: KTH Royal Institute of Technology; 2016.

Address correspondence to:
Shonali Paul, BTech, MBA
CloudLIMS.com
427N Tatnall Street #42818
Wilmington, DE 19801-2230

E-mail: shonali@cloudlims.com