Scientific Academia Journal
Vol. 3 (2020)
https://academicpinnacle.com/index.php/SAJ

Risk Assessment Framework for Cybersecurity in Genetic Data Repositories

Aravind Kumar Kalusivalingam
Northeastern University, Boston, USA
Corresponding: karavindkumar1993@gmail.com

Abstract

The abstract of the Risk Assessment Framework for Cybersecurity in Genetic Data Repositories outlines a structured approach to evaluate and mitigate risks associated with storing and managing genetic data in digital repositories. This framework integrates principles of cybersecurity with the unique challenges posed by the sensitive nature of genetic information. By identifying potential threats, vulnerabilities, and impacts, the framework enables repository administrators to prioritize security measures effectively. Key components include threat modeling, vulnerability assessment, risk analysis, and the development of tailored mitigation strategies. Ultimately, this framework aims to enhance the protection of genetic data, ensuring confidentiality, integrity, and availability while facilitating responsible research and data sharing in the genomic era.

Keywords: Risk Assessment Framework, Cybersecurity, Genetic Data Repositories, Threat Modeling

1. Introduction

Genetic data repositories play a pivotal role in advancing biomedical research and personalized medicine by facilitating the storage, sharing, and analysis of genetic information. However, the proliferation of these repositories has raised significant concerns regarding the security and privacy of sensitive genetic data. Cybersecurity threats, ranging from unauthorized access to data breaches and malicious attacks, pose substantial risks to the confidentiality, integrity, and availability of genetic data stored within these repositories [1]. As the volume and complexity of genetic data continue to grow, there is an urgent need for robust risk assessment frameworks tailored to the unique challenges of cybersecurity in genetic data repositories. The Risk Assessment Framework for Cybersecurity in Genetic Data Repositories addresses this pressing need by providing a structured approach to evaluate and mitigate cybersecurity risks

associated with the storage and management of genetic data [2]. This framework integrates principles of cybersecurity with the intricacies of genetic data management, offering repository administrators a systematic methodology to identify, assess, and address potential threats and vulnerabilities. By employing this framework, genetic data repositories can enhance their security posture, safeguarding sensitive genetic information against evolving cyber threats and ensuring compliance with regulatory requirements and ethical standards.

Genetic data repositories serve as crucial hubs for the aggregation, storage, and dissemination of genetic information collected from diverse sources, including research studies, clinical trials, and biobanks. These repositories house vast amounts of genomic data, encompassing DNA sequences, genetic variants, phenotypic data, and associated metadata. They facilitate collaborative research efforts, enable data sharing among scientists and clinicians, and support the development of novel insights into the genetic basis of diseases, drug responses, and human traits [3]. Cybersecurity is paramount in genetic data management due to the sensitive and confidential nature of genomic information. Genetic data repositories store highly personal data that can reveal sensitive information about individuals' health, ancestry, and predispositions to diseases. Unauthorized access, data breaches, or malicious tampering with genetic data could have profound consequences, including privacy violations, discrimination, and identity theft. Moreover, genetic data repositories are prime targets for cyber-attacks due to the potential for financial gain or malicious intent, highlighting the critical need for robust cybersecurity measures [4]. Effective cybersecurity safeguards not only protect the confidentiality, integrity, and availability of genetic data but also ensure compliance with regulatory mandates, such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR). As genetic data continues to underpin advancements in biomedical research and clinical practice, ensuring the security of genetic data repositories is essential to maintaining public trust and fostering innovation in genomics [5].

The Risk Assessment Framework for Cybersecurity in Genetic Data Repositories is a structured methodology designed to systematically evaluate and manage cybersecurity risks associated with the storage, processing, and sharing of genetic data. This framework provides a comprehensive approach to identifying potential threats, vulnerabilities, and impacts on the confidentiality, integrity, and availability of genetic data within repositories [6]. By employing a systematic and iterative process, this framework enables repository

administrators to prioritize security measures effectively, thereby reducing the likelihood and impact of cyber threats. Threat modeling is a foundational component of the Risk Assessment Framework, aiming to systematically identify and categorize potential threats to genetic data repositories. This process involves analyzing the various actors, assets, and potential attack vectors within the repository ecosystem. Threat actors may include malicious insiders, external hackers, or unauthorized users seeking to exploit vulnerabilities for personal gain or malicious intent. Asset identification involves mapping the critical components of genetic data repositories, such as databases, servers, and communication channels. Threat modeling also considers potential attack vectors, including network intrusions, malware infections, and social engineering tactics. By comprehensively evaluating threats, repository administrators can develop proactive security measures to mitigate risks effectively.

2. Understanding Genetic Data Security Risks

Genetic data present distinct challenges compared to other types of data due to their inherently sensitive and personal nature. Unlike traditional forms of information, genetic data contain highly intimate details about an individual's biological makeup, including predispositions to diseases, familial relationships, and ancestry. This inherently personal nature raises concerns regarding privacy, consent, and ethical use, requiring special considerations in data management practices. Furthermore, genetic data are inherently complex, consisting of vast amounts of genomic sequences, variations, and phenotypic information, which present challenges in terms of storage, analysis, and interpretation [7]. The dynamic and evolving nature of genetic data, influenced by factors such as genetic mutations, environmental exposures, and epigenetic modifications, further complicates their management and security. Genetic data repositories face a myriad of potential threats that jeopardize the confidentiality, integrity, and availability of stored genetic information. One of the primary threats is unauthorized access, where malicious actors gain illegitimate entry to genetic data repositories either through exploiting vulnerabilities or leveraging insider privileges. Unauthorized access could lead to privacy breaches, exposing sensitive genetic information to unauthorized individuals or entities. Another significant threat is data breaches, which occur when genetic data repositories are compromised, resulting in the unauthorized disclosure or theft of genetic information [8]. Data breaches not only violate individual privacy rights but also undermine public trust in the security of genetic data repositories. Additionally, genetic data repositories are susceptible

to cyber-attacks, including malware infections, ransomware attacks, and denial-of-service (DoS) attacks, which can disrupt normal operations and compromise the integrity of stored genetic data. Social engineering attacks, such as phishing attempts or pretexting, pose further threats by exploiting human vulnerabilities to gain unauthorized access to genetic data repositories [9]. Moreover, the increasing digitization and interconnectedness of genetic data repositories amplify the risk of supply chain attacks, where attackers target third-party vendors or service providers to gain access to sensitive genetic information. Vulnerabilities in genetic data management systems represent potential weaknesses or gaps that could be exploited by malicious actors to compromise the security of genetic data repositories [10]. Common vulnerabilities include software vulnerabilities, such as unpatched software or insecure coding practices, which could be exploited by attackers to gain unauthorized access or execute malicious code within genetic data repositories. Configuration vulnerabilities, such as misconfigured access controls or inadequate encryption protocols, also pose significant risks by exposing genetic data to unauthorized access or interception during transmission. Additionally, human factors, such as insufficient training or awareness among personnel, can introduce vulnerabilities by increasing the likelihood of human error or susceptibility to social engineering attacks. It's critical for genetic data repositories to regularly assess and address vulnerabilities to mitigate the risk of security breaches and safeguard the confidentiality and integrity of stored genetic data [11].

3. Implementation of the Risk Assessment Framework

Data collection and inventory are foundational steps in the risk assessment framework for cybersecurity in genetic data repositories. This process involves identifying and cataloging all types of data stored within the repository, including genetic sequences, phenotypic information, metadata, and associated documentation. Additionally, repository administrators must assess the sources and origins of the data, such as research studies, clinical trials, or biobanks, to understand the context and sensitivity of the information. Establishing a comprehensive inventory of genetic data enables repository administrators to gain visibility into the scope and scale of the repository's data assets, facilitating subsequent risk assessment and mitigation efforts. Threat identification and classification are essential components of the risk assessment framework, aiming to systematically identify and categorize potential threats to genetic data repositories. This process involves analyzing the various actors, motives, and methods that pose risks to the confidentiality,

integrity, and availability of genetic data [12]. Threat actors may include malicious insiders, external hackers, or unauthorized users seeking to exploit vulnerabilities for personal gain or malicious intent. By classifying threats based on their severity, likelihood, and potential impact, repository administrators can prioritize security measures and allocate resources effectively to mitigate the most significant risks. Vulnerability assessment techniques are employed to identify weaknesses or gaps in the security posture of genetic data repositories. This process involves conducting technical assessments, such as penetration testing, vulnerability scanning, and code review, to identify vulnerabilities in software, hardware, and infrastructure components [13]. Additionally, vulnerability assessment considers human factors, such as employee training and security awareness, to identify potential weaknesses in organizational processes and procedures. By systematically evaluating vulnerabilities, repository administrators can prioritize remediation efforts and implement appropriate controls to mitigate potential risks effectively. Risk analysis involves assessing the likelihood and potential impact of identified threats and vulnerabilities on the security and integrity of genetic data repositories. This process considers various factors, including the severity of potential breaches, the likelihood of occurrence, and the potential impact on organizational objectives and stakeholders. Risk analysis utilizes qualitative and quantitative techniques, such as risk matrices, probabilistic models, and scenario analysis, to evaluate risks effectively. By conducting risk analysis, repository administrators can prioritize security investments, allocate resources efficiently, and develop risk mitigation strategies tailored to the specific needs and priorities of genetic data repositories.

Figure 1, illustrates the data flow and associated threat landscape of a genetic information system, which is divided into three main phases with human interaction throughout. Pre-Analytical Phase: The left section represents the collection, storage, and distribution of biological samples, highlighting potential threats such as sample mislabeling and unauthorized access. Analytical Phase: The middle section covers wet laboratory preparation, DNA sequencing, and bioinformatics processes. Here, threats include equipment tampering, data corruption during sequencing, and unauthorized data access during analysis [14]. Post-Analytical Phase: The right section depicts the storage, sharing, and further analysis of genetic data. Key threats in this phase include data breaches, unauthorized sharing, and data integrity issues. Throughout the diagram, arrows indicate the flow of data between phases, and icons represent specific threats at each stage. The figure emphasizes the need for robust security measures to protect genetic data from collection to final storage and

analysis. Stakeholders should consider the storage, transit, and destruction of sensitive biological material as crucial aspects of overall genetic information security and cybersecurity.

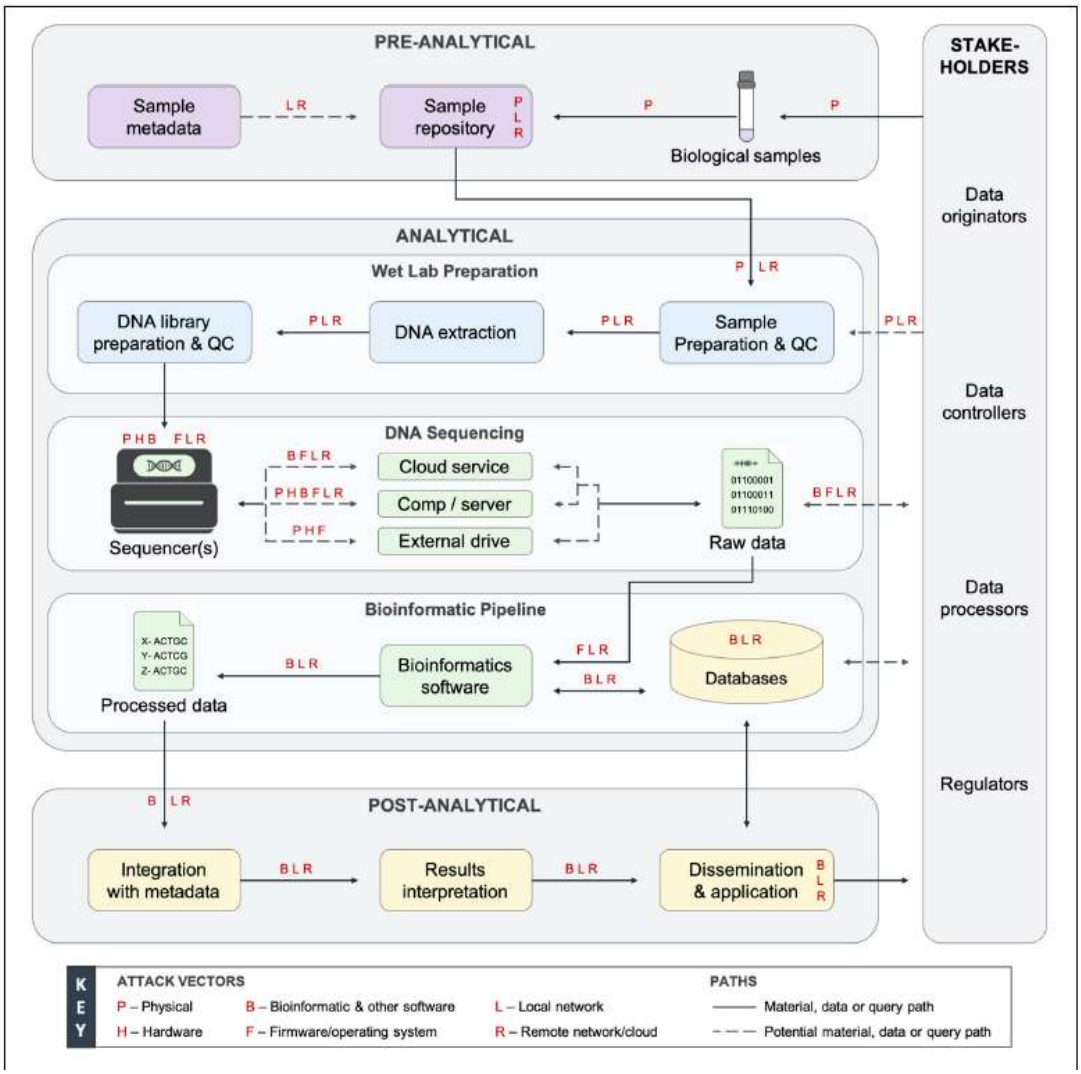

Figure 1: Data flow diagram of a genetic information system and its threat landscape. These systems have three phases with human interaction throughout. The pre-analytical phase involves collecting, storing, and distributing samples. The analytical phase covers lab preparation, DNA sequencing, and bioinformatics, where data is generated, analyzed, transmitted, and stored.

Mitigation strategy development involves developing and implementing measures to reduce the likelihood and impact of identified risks to genetic data repositories. This process encompasses a range of technical, administrative, and procedural controls aimed at addressing vulnerabilities, deterring threats, and enhancing the overall security posture of the repository [15]. Mitigation strategies may include implementing access controls, encryption mechanisms, intrusion detection systems, and incident response plans to protect genetic

data from unauthorized access, disclosure, or manipulation. Additionally, ongoing monitoring, evaluation, and refinement of mitigation strategies are essential to adapt to evolving threats and ensure the continued security of genetic data repositories.

4. Case Studies and Challenges

The Risk Assessment Framework for Cybersecurity in Genetic Data Repositories provides a systematic approach to enhancing the security of genetic data repositories. By applying this framework, repository administrators can systematically identify, assess, and mitigate cybersecurity risks, thereby safeguarding the confidentiality, integrity, and availability of genetic data. The framework enables administrators to tailor security measures to the unique challenges posed by genetic data, ensuring compliance with regulatory requirements and ethical standards while facilitating responsible data sharing and research. Successful risk mitigation strategies in genetic data repositories often involve a combination of technical, administrative, and procedural controls. These strategies may include implementing robust access controls to restrict unauthorized access to sensitive data, deploying encryption mechanisms to protect data during transmission and storage, and establishing regular security audits and monitoring processes to detect and respond to security incidents promptly. Additionally, comprehensive employee training and awareness programs can help mitigate human-related risks, such as social engineering attacks or inadvertent data breaches. By adopting a multi-layered approach to risk mitigation, genetic data repositories can effectively reduce the likelihood and impact of cybersecurity threats. Challenges in securing genetic data repositories include balancing the need for data accessibility with stringent security requirements, addressing evolving cyber threats and regulatory mandates, and ensuring interoperability and compatibility with existing data management systems. Lessons learned from implementing the risk assessment framework include the importance of stakeholder engagement and collaboration, the need for continuous monitoring and evaluation of security measures, and the value of adopting a proactive and adaptive approach to cybersecurity. Additionally, challenges related to data governance, consent management, and ethical considerations underscore the need for ongoing dialogue and collaboration among researchers, policymakers, and stakeholders to address emerging challenges and ensure the responsible use and sharing of genetic data.

5. Conclusion

In conclusion, the Risk Assessment Framework for Cybersecurity in Genetic Data Repositories represents a vital tool in addressing the complex challenges of securing genetic data repositories. By providing a structured approach to identifying, assessing, and mitigating cybersecurity risks, this framework empowers repository administrators to safeguard the confidentiality, integrity, and availability of sensitive genetic information. Through the application of proactive risk mitigation strategies and continuous monitoring and evaluation, genetic data repositories can enhance their security posture, ensuring compliance with regulatory requirements and ethical standards while fostering responsible data sharing and research. However, as genetic data continues to evolve and expand in scope, stakeholders must remain vigilant in addressing emerging cybersecurity threats and implementing robust security measures. Ultimately, the adoption and implementation of the Risk Assessment Framework for Cybersecurity in Genetic Data Repositories are essential steps toward maintaining public trust, promoting innovation, and advancing biomedical research in the genomic era.

Reference

[1] G. J. Schumacher, S. Sawaya, D. Nelson, and A. J. Hansen, "Genetic information insecurity as state of the art," *Frontiers in bioengineering and biotechnology,* vol. 8, p. 591980, 2020.

[2] D. S. Schabacker, L.-A. Levy, N. J. Evans, J. M. Fowler, and E. A. Dickey, "Assessing cybersecurity vulnerabilities and infrastructure resilience," *Frontiers in bioengineering and biotechnology,* vol. 7, p. 61, 2019.

[3] R. LARATTA, "Tools for the Assessment of Cyber Risk in Healthcare Facilities," 2019.

[4] M. U. Aksu *et al.*, "A quantitative CVSS-based cyber security risk assessment methodology for IT systems," in *2017 International Carnahan Conference on Security Technology (ICCST)*, 2017: IEEE, pp. 1-8.

[5] D. DiEuliis, C. D. Lutes, and J. Giordano, "Biodata risks and synthetic biology: a critical juncture," *J Bioterror Biodef,* vol. 9, no. 1, p. 159, 2018.

[6] J. T. O'Brien and C. Nelson, "Assessing the risks posed by the convergence of artificial intelligence and biotechnology," *Health security,* vol. 18, no. 3, pp. 219-227, 2020.

[7] P. Russo, A. Caponi, M. Leuti, and G. Bianchi, "A web platform for integrated vulnerability assessment and cyber risk management," *Information,* vol. 10, no. 7, p. 242, 2019.

[8] C. Tomulescu, "Cybersecurity. A short review," in *Smart Cities International Conference (SCIC) Proceedings*, 2020, vol. 8, pp. 393-410.

[9] S. Gil, "Network services as risk factors: A genetic epidemiology approach to cyber security," in *Dynamic Networks and Cyber-Security*: World Scientific, 2016, pp. 89-109.

[10] S. Kriaa, "Joint safety and security modeling for risk assessment in cyber-physical systems," Université Paris Saclay (COmUE), 2016.

[11] S.-C. Cha and K.-H. Yeh, "A data-driven security risk assessment scheme for personal data protection," *IEEE Access,* vol. 6, pp. 50510-50517, 2018.

[12] K. Ruan, "Introducing cybernomics: A unifying economic framework for measuring cyber risk," *Computers & Security,* vol. 65, pp. 77-89, 2017.

[13] A. B. Carter, "Considerations for genomic data privacy and security when working in the cloud," *The Journal of Molecular Diagnostics,* vol. 21, no. 4, pp. 542-552, 2019.

[14] Z. M. King, D. S. Henshel, L. Flora, M. G. Cains, B. Hoffman, and C. Sample, "Characterizing and measuring maliciousness for cybersecurity risk assessment," *Frontiers in psychology,* vol. 9, p. 290575, 2018.

[15] J. D. Voss *et al.*, "Operations Security for Emerging Biotechnology Applications," 2020.