

PRIVACY AND SECURITY CHALLENGES TO USER DATA DIGITAL LIBRARIES: A LITERATURE REVIEW

Journals of Arts & Humanities Studies

ISSN: 3069-325X (Online)

Vol. 1: Issue 5

Page 23–37 © The Author(s) 2025

Received: 01 November 2025

Accepted: 16 November 2025

Published: 31 December 2025

Omer Hassan Abdelrahman - *Department of Library and Information
Science, University of Khartoum*



ABSTRACT

This review article explores the challenges to user data privacy and security in digital libraries. It examines various privacy violations and security risks, such as data breaches, along with their consequences on user data and digital library services. A narrative literature review approach was adopted for data collection. Various information resources were reviewed including journal articles and websites. Best practices for protecting user data privacy and security are explained and real-world case studies of privacy violations and cybersecurity attacks are discussed. This is followed by presenting the strategies implemented by affected libraries to mitigate the impact of violations, and the lessons learned. Finally, the article reviews key emerging technologies implemented by digital libraries, such as Artificial Intelligence (AI) and Blockchain technology, focusing on their benefits and risks to user data privacy and security.

KEY WORDS

Digital Libraries, User Data Privacy, User Data Security, Cyberattacks, Data Breaches, Regulatory Compliance, General Data Protection Regulation (GDPR), Mitigation Strategies, Emerging Technologies, Artificial Intelligence (AI), Blockchain Technology

1. Introduction

Digital libraries have been providing access to information for scholars, researchers, students, and a wide spectrum of users over the last three decades. Advances in communications technologies have extended the global coverage of digital libraries. This has been accompanied by a parallel increase in the number of users and services provided by these libraries. Consequently, digital libraries have increased their collection and processing of user data for the purpose of providing personalized experiences and efficient content delivery. This, in turn, has raised significant concerns regarding user data privacy and security in digital libraries. This review article addresses the critical issue of challenges to user data privacy and security within the digital library environment. Drawing on a literature review of a variety of information resources, such as journal articles and websites, it explores key topics related to user data such as the types of data collected, common privacy violations encountered, and the security risks to which this data is vulnerable. The impact of these challenges on the digital library user experience is also highlighted, while potential solutions, best practices, and mitigation strategies for safeguarding user data are examined. In this context, the article discusses technical, policy, and organizational measures as part of best practices. It underscores the importance of regulatory compliance, highlighting the role of legal and regulatory frameworks in safeguarding user data privacy, with a particular focus on the General Data Protection Regulation (GDPR). Real-world case

studies are presented to illustrate various privacy and security violations. This is followed by presenting mitigation strategies implemented by affected libraries, and lessons learned. The article concludes by discussing five key emerging technologies – Artificial Intelligence, Internet of Things, Cloud Computing, Blockchain, and Big Data analytics – and their impact on user data privacy and security in the digital library environment.

2. Objectives

The General objective of this narrative literature review is to explore privacy and security challenges affecting user data in digital libraries, with the following specific objectives:

- To identify various types of user data collected in digital libraries and examine the different types of privacy and security violations this data faces.
- To examine different types of cybersecurity threats in the digital library environment.
- To assess the impact of regulatory frameworks, such as the General Data Protection Regulation (GDPR), on privacy practices within digital libraries.
- To highlight best practices for safeguarding user data privacy and security in digital libraries.
- To review notable case studies of data breaches and the efforts exerted by the affected libraries to mitigate their impact.
- To review and assess the benefits and risks of five emerging technologies – Artificial Intelligence (AI), Internet of Things (IoT), Cloud Computing, Blockchain technology, and Big data- on user data privacy and security.

3. Methodology

This article adopts a narrative literature review approach to explore the privacy and security challenges to user data in digital libraries. The purpose of the review is to collect and synthesize key findings from various types of credible information resources and reputable web-based sources relevant to the topic. Due to lack of access to subscription-based academic databases, the search was conducted using Google Scholar and general Google search as the main searching tools for literature. Both tools were used to gain access to high-quality sources including peer-reviewed open access journals, institutional repositories, conference papers, and official reports from authoritative organizations such as the International Federation of Library Associations and Institutions (IFLA) and other professional library and cybersecurity bodies. Sources also included reputable web-based materials such as university websites, institutional blogs, and selected personal blogs by recognized experts or practitioners. Searches were carried out using keywords such as “user data privacy,” “user data security,” “digital libraries,” “emerging technologies in digital libraries,” “data security,” “cybersecurity breaches,” “cyberattacks,” “data breaches” and “privacy best practices.” Searches were confined to sources in the English language and published during a ten-year period from 2015 to 2025. Search results were verified, and thematically categorized according to key aspects of the review. These aspects included: types of data collected, violations this data faces, consequences of these violations on users and libraries, best practices for safeguarding user data, real-world case studies of cyberattacks and mitigation strategies, and the impact of emerging technologies on user data privacy and security in digital libraries.

4. Findings and Discussion

4.1. Types of User Data Collected by Digital Libraries

Digital libraries can gain valuable insights into user behavior, preferences, and needs by collecting and analyzing various types of user data. These insights allow the libraries to allocate resources more efficiently, tailor programs and services to meet the specific needs of different user groups, and enhance overall user satisfaction. [1] This section categorizes the types of data collected within the digital library environments.

The types of user data typically collected in digital libraries include:

4.1.1. Usage Data and Activity Logs. This encompasses:

- Searches and history
- Views and downloads of content
- Duration of time spent on resources
 - o Session lengths
 - o Login and logout timestamps
 - o Error reports

4.1.2. Technical Data including the following:

- IP addresses
- Browser type and version
- Operating system
- Types of devices used such as computer, mobile, tablet....etc.
- Referral sources i.e. how users arrived at the library site

4.1.3. Authentication and Account Data including:

- Usernames and Identification Documents (IDs)
- Email addresses
- Potential demographic data if collected during registration
- Membership status

4.1.4. Personalization Data including reading habits and preferences. [2, 3]***4.2. Violations to User Privacy and Security***

In digital libraries, “Privacy” refers to the user's ability to control how their collected data can be accessed, managed, and shared. This notion encompasses several aspects such as information privacy, communication privacy, and individual privacy. Information privacy deals with protecting the personal data that is collected and stored by websites including digital libraries. Communication privacy deals with the unauthorized use and abuse of private conversations. Individual privacy centers on protecting a user's identity in the virtual space. [4] As pointed out by the International Federation of Library Associations and Institutions (IFLA), there are challenges to users' privacy in libraries and that their data, activities, communications, and transactions may be collected by content and service providers hired by library and information services. The IFLA statement also points out that these providers may require that libraries collect data as a condition of providing their content or services. [5]

4.2.1. Types of Violations to User Data Privacy and Security

User data is crucial for digital libraries and platforms as it facilitates key functions such as search optimization, content recommendations, and access management. Users on their part expect these platforms to handle their data such as "personal information, search histories, and reading habits" in a confidential manner. [6, 7] However, the various aspects of user data collected by digital libraries and platforms undergo a number of various violations and threats as detailed below.

4.2.1.1. Unauthorized Access and Data Breaches

There are various unauthorized and illicit methods through which user data in digital libraries can be compromised. Below is a breakdown of these methods.

4.2.1.2. Cyberattacks

Digital libraries are targets for cyberattacks, causing unauthorized access and data theft. The aim of these attacks which use sophisticated methods to bypass security measures is to compromise users' personal information. There are several types of cybersecurity threats to digital libraries including:

- **Data Breaches:** This pertains to illegal access to confidential user data like personal information and borrowing history. Data breaches happen when someone who is not authorized accesses a database which contains private information. Access to personal information during data breaches can lead to severe infringements such as identity theft and financial damage. The serious ramifications of these breaches involve the illicit exposure of vast volumes of data.
- **Malware:** refers to the installation of harmful software, which can disrupt library functions or jeopardize data security.
- **Phishing:** This refers to activities that attempt to trick library staff or users into revealing sensitive information or login credentials.
- **Distributed Denial of Service (DDoS) Attacks:** These attacks aim to overload library servers with heavy traffic disrupting their online services.
- **Insider Threats:** These threats originate from individuals within the organization who misuse their access rights to steal or manipulate the library or its users' information.
- **Exploiting software vulnerabilities or weak passwords:** This involves obtaining unauthorized access to systems, networks, or facilities. [4, 6, 8-12]

4.2.1.3. Third-Party Data Sharing and Integration:

This encompasses the following:

- **Integration of third-party tools:** Some emerging technologies such as big data analytics, cloud storage, and authentication providers may collect or share user data beyond the library's control, occasionally for targeted advertising or financial gain.
- **Lack of transparency with vendors:** Libraries often lack knowledge about how third-party digital service providers manage and disseminate user data. [6, 13]
- **4.2.1.4. Insufficient Policies and Practices:**
- **Underdeveloped policies for Internet of Things collections:** emerging technologies introduce privacy challenges that may not be adequately addressed by existing policies and guidelines.
- **Non-disclosure of data usage to users or compliance with regulations:** Many libraries fail to openly disclose how user personal information is utilized and stored. [7]

4.3. Consequences of the Violations

The challenges and threats to user data discussed above have profound consequences on digital libraries and their users. They include the following:

4.3.1. Loss of Trust and Confidence, leading to Reduced Usage: when users feel their information is monitored or prone to leakage, they may be reluctant to use library services, negatively affecting the library's objectives. [14, 15]

4.3.2. "Chilling Effect" on Freedom of Speech and Expression: Knowledge of extensive data gathering and surveillance may induce self-censorship, as users worry about unforeseen consequences. This can compromise democracy and civil engagement. [5]

4.3.3. Identity Theft and Financial Fraud: Breached personal data can lead to severe outcomes for individuals, including identity theft and financial fraud. [12, 15]

4.3.4. Legal and Financial Consequences, including:

- **Fines and penalties:** Failing to adhere to data protection and privacy regulations by digital libraries can lead to significant legal repercussions, including fines.
- **Lawsuits:** Individuals impacted can file lawsuits for compensation due to the damages arising from infringing the privacy of their personal data.
- **Financial losses:** Libraries may endure financial losses from data breaches, which can be expensive to recover from. [12, 15]

4.3.5. Disruption of Library Services and Operations: Cyberattacks of the Denial of Services (DoS) and ransomware type can inhibit workflow in a library and also make the documents unavailable even to legitimate users [10, 16]

4.4. Best Practices for Privacy and Security in Digital Libraries

Digital libraries need to follow best practices for the purpose of securing and protecting user data. These best practices encompass a range of activities addressing technical measures, policies, as well as educational and awareness issues. This section highlights these measures and activities, with a focus on regulatory compliance and its widely implemented regulations enacted by the General Data Protection Regulation (GDPR).

4.4.1. Technical Measures

4.4.1.1. Encryption: Employ robust encryption protocols for data both at rest (stored data) and in transit (data being transferred). This ensures that if unauthorized access takes place, the intercepted data remains unintelligible.

4.4.1.2. Authentication and Access Control: Use strong authentication and access control procedures such as multi-factor authentication to limit data access and block interaction with digital assets to designated users only.

4.4.1.3. Regular Updates and Patching: Ensure that all systems and software are consistently updated to protect against known vulnerabilities and emerging threats.

4.4.1.4. Continuous Security Monitoring and Incident Response: Deploy real-time monitoring solutions to observe user activities, spot irregularities, and facilitate swift action in response to potential security threats and incidents.

4.4.1.5. Risk Assessment: Regularly evaluate and address vulnerabilities within the digital library's infrastructure. [6, 11, 15]

4.4.2. Policy Measures

4.4.2.1. Transparency in Data Policies: Libraries must create openly available documents that detail their privacy policy which tells users how their data will be collected, stored, utilized, and protected. Users must be empowered to manage their data.

4.4.2.2. Privacy by Design: It is important to integrate privacy considerations into the design and structure of digital library systems from the outset, rather than treating them as an afterthought. This includes making privacy the default setting and providing pseudonymity as an option.

4.4.2.3. Clear Data Sharing Agreements: Develop explicit data-sharing agreements with third-party providers, ensuring they comply with the same privacy standards observed by the library. [6, 11]

4.4.3. Organizational Measures

4.4.3.1. Security Awareness Training: Train both staff and users on safe online practices. This should include creating strong passwords, recognizing phishing attempts, and protecting sensitive information.

4.4.3.2. Incident response plan: In the event of a data breach, libraries must have a well-practiced incident response plan which includes promptly notifying the relevant supervisory authority and the affected individuals. Consequently, it is crucial to create an incident response plan that details the actions to be taken in the aftermath of a cybersecurity incident. This plan should include protocols for reporting incidents, mitigating and managing damage, and restoring normalcy. The plan should be regularly tested and updated as needed. [15, 16]

4.4.4. Regulatory Compliance

Regulatory compliance in cybersecurity pertains to an organization's commitment to meet certain legal, regulatory, and industry-specific cybersecurity regulatory standards that dictate methods for data protection and ways of implementing security controls. These standards assist organizations and institutions on how to safeguard sensitive information, maintain systems' integrity, and shield systems from cyber-attacks. They mandate that organizations comply with their cybersecurity practices with certain requirements established by the government, industry bodies or contracts. [17]

Libraries need to comply with available data protection regulations like the General Data Protection Regulation (GDPR) in the European Union or the California Consumer Privacy Act (CCPA) in the United States.

4.4.4.1. Key Principles of Data Privacy Regulations Worldwide

There are a number of principles related to data privacy at the international level. They include the following:

- Lawfulness, fairness, and transparency. This principle implies that data must be processed legally, respectfully, and with complete transparency towards the data subject. This principle is reflected in laws that require organizations to provide clear legal advice and remain transparent about their data processing activities.
- Data minimization. This principle involves limitation on data collection and discourages excessive collection and retention of data. It promotes the notion that organizations should only collect and retain data essential for the specific purpose declared to the data subject.
- Purpose limitation: This principle stresses that personal data should be collected for specific, clear, and lawful purposes, and should not be further processed in ways that are incompatible with these purposes.
- Security and confidentiality. This principle requires organizations to implement appropriate technical and organizational measures to safeguard personal data against unauthorized or unlawful processing and accidental loss, destruction, or damage. [18, 19]

4.4.4.2. The General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) was enacted by the European Union in 2018. It provides a comprehensive and clear framework that acknowledges that various categories of personal data necessitate differing levels of security. The GDPR applies to all organizations that collect and process personal data on European Union (EU) residents, regardless of their geographical location. Non-EU organizations have to appoint a GDPR representative and will be liable for all fines and sanctions. [17]

Some of the key requirements of the GDPR include:

- Consent: Organizations must obtain consent for the collection of personal data, with the level of consent varying for different types of personal data collected.
- Data minimization: The GDPR requires that organizations can only collect personal data that is directly relevant to a well-defined business purpose. If an organization collects personal data for one reason but later chooses to use it for another purpose, this may be considered as a violation of compliance.
- Individual rights: Individuals whose data is processed have the right to be informed about the reasons for data collection and how it is processed. They also have the right to object, amend, and request the deletion of their data. Moreover, they should be individually informed if their personal data has been compromised in a way that could threaten their freedoms and rights. [18, 19]

4.5. Case Studies

There are a number of notable case studies which demonstrate how user data privacy and security in digital libraries could be violated and threatened. These cases illustrate instances of data breaches and ransomware attacks. They also show some of the methods of mitigating such violations and threats. Four of the most notable recent cases are detailed below:

4.5.1. Internet Archive / Wayback Machine Breach

The Internet Archive Wayback Machine suffered a cyberattack that resulted in the shutdown of the digital library and Wayback. On October 9th, 2024, a data breach coupled with a DDoS (Distributed Denial of Service) attack rendered the site offline. The hackers injected malicious JavaScript that led to stealing of a database containing 31 million email addresses, usernames, and passwords. Consequently, the Archive was temporarily offline. Services were resumed in read-only mode, with new archiving disabled for a period of time. [20]

4.5.2. The British Library

In October 2023, The British Library, the national library of the United Kingdom, fell victim of a critical ransomware attack by a hacker group known as Rhysida. This attack resulted in the library's website, catalogue, and other digital collections remaining inaccessible for a long time. After the British Library refused to pay the ransom, the attackers released around 600GB of data, including personal information of library users and employees, on the dark web. The British Library estimated that it would need to allocate roughly 40% of its financial reserves—about £6–7 million—to recover from this incident. The library later confirmed that the hackers stole employee and user data, including at least names and email addresses. Other reports suggested that the hacked data included internal Human Resources documents, some invoices, and other documents containing salary information. This incident demonstrates the impact of large-scale cyberattacks on resource availability and privacy violation of library staff and users. This incident particularly illustrates the operational damage that ransomware attacks can cause besides the uncontrolled leakage of sensitive data, even when such exposure is not intended by the attackers. [21, 22]

4.5.3. Seattle Public Library

On May 25, 2024, the Seattle Public Library in the USA was attacked by ransomware hackers. As a result, a number of essential services went missing. For example, borrowers were unable to return the physical materials they had previously borrowed or place holds on any new items. Interlibrary Loans were inaccessible, the in-building Wi-Fi was nonfunctional, and public computers were also down. Moreover, users could not checkout mobile hotspot, the catalog failed to update, and library users and employees could not use the printers. [16]

4.5.4. Toronto Public Library

Toronto Public Library in Canada was also targeted by a ransomware attack on October 28, 2023. This attack took down the most essential technical systems, including the library's website, its internal network, and public computers. As a result, for a period of two months, users were unable to access their library accounts online or use the library's computers. [16]

4.6. Mitigation strategies implemented by affected digital libraries and Lessons Learned

This section details key mitigation strategies implemented by three of the affected libraries to protect user data and improve system resilience after the breaches they suffered. The libraries are the Internet Archive/ Wayback Machine, the British Library, and Toronto Public Library. These libraries are selected to represent the four attacked libraries and illustrate a variety of mitigation strategies.

4.6.1. The Internet Archive/ Wayback Machine

4.6.1.1. Mitigation Strategies Implemented

- Immediate Shutdown and Assessment: The Internet Archive took its website and services offline as soon as the breach and the harm done were detected to prevent further damage and assess the situation.
- Restoration in Stages: Essential services such as the Wayback Machine, Archive-It, and national library crawls were gradually restored, with some features initially available only in read-only mode to ensure safety during the restoration process.
- Security Upgrades, this included:
 - Enhanced firewall technologies and adjustments to data flow for improved monitoring and control.
 - Upgrades and patches to out-dated software systems to close vulnerabilities.

- Regular rotation of credentials and API keys, especially after learning that secrets from the development server had been exposed.
- Enhanced Monitoring: Increased vigilance and monitoring of systems to detect and respond to on-going or new attacks.
- Community and Vendor Support: Received assistance from the open source community and business vendors regarding advanced security tools.
- Staff Expansion: Increased the technical staff to focus on cybersecurity and service availability.
- Communication: Kept users informed through blog updates and social media, and apologized for any inconvenience caused by service interruptions. [23 - 26]

4.6.1.2. Lessons Learned

- No Organization Is Immune: Even well-intentioned, non-profit organizations like the Internet Archive are susceptible to advanced cyberattacks and must stay alert.
- Value of Proactive Security Measures: Regular security evaluation, patch and upgrade management, and unreservedly maintaining proactive improvements are necessary to eliminate breaches, especially with attackers taking advantage of unpatched systems and credential leaks.
- Data Integrity and Trust: Gaining the trust of users and safeguarding the integrity of digital archives became even more critical for the system. Loss of trust, even if temporary, is extremely damaging to public repository services.
- Transparency and Communication: Maintaining clear and timely communication with users and stakeholders is essential during and after a breach to maintain trust and offer advice on potential risks.
- Importance of Community Support: Assistance from the wider library, open source, and security communities is vital for responding to and recovering from large-scale attacks.
- Continuous Enhancement: The incident prompted the Internet Archive to accelerate long-planned upgrades and adopt a more security-oriented approach, which will enhance its resilience over time. [23 - 28]

4.6.2. The British Library Ransomware Attack

4.6.2.1. Mitigation Strategies Implemented

- Complete transparency through public reporting on the attack and recovery process.
- Activation of crisis management teams along with consultation with National Cyber Security Centre and specialists.
- Immediate and long-term recovery plan called "Rebuild and Renew aimed at modernizing infrastructure and security frameworks.
- Adoption of multi-factor authentication (MFA) after the attack as it was absent on the breached server. [29-31]

4.6.2.2. Lessons Learned

- Comprehensive and accredited security measures may still have vulnerabilities; continuous reassessment is essential.
- Multi-factor authentication (MFA) is crucial, especially for remote access servers utilized by trusted partners.
- Early detection and response to reconnaissance activities can avert extensive attacks.
- Transparency helps maintain trust and offers valuable guidance to peer institutions.
- Legacy systems and complex technological setups increase vulnerability and complicate recovery efforts. [30-31]

4.6.3. Toronto Public Library

4.6.3.1. Mitigation Strategies

- Immediate shutdown of all internal and external systems and networks.
- Activation of Major Cyber Security Incident Playbook and Privacy Breach Protocol.
- Involvement of third-party legal advisers and cybersecurity experts for containment, forensics, and impact evaluation.

- Launch of the Incident Management System alongside the Library Operations Centre (LOC) to ensure business recovery and continuity.
- Communication with staff, stakeholders, and the general public.
- Renovating and modernizing the IT environment, including updated hardware and software, enhanced security controls, and patching. [32-34]
- Mandatory cybersecurity training for all staff.

4.6.3.2. Lessons Learned

- Highlighting the importance of a rapid, organized incident response and effective business continuity strategies.
- Need for regular updates and patching of software and hardware.
- The advantages of engaging external specialists (legal and cybersecurity) in the oversight and examination of security breaches.
- Improved staff training and communication protocols.
- Enhanced governance, risk management, and teamwork with city cybersecurity representatives. [32-34]

4.7. The Impact of Emerging Technologies on User Data Privacy and Security

With the rapid expansion and development of the digital library landscape worldwide, several emerging technologies have been impacting user data privacy and security. These technological advancements serve as a double-edged sword; while they are beneficial, they also have their risks; they often involve extensive user data collection, processing, and sharing. Without proper management this will increase the possibility of unauthorized access and data breaches. [35]

This section examines how emerging technologies impact user data privacy in digital libraries. Five emerging technologies that are widely used are presented, namely Artificial Intelligence (AI), Internet of Things (IoT), Cloud Computing, Blockchain technology, and Big data. Each one of these emerging technologies is defined and its positive and negative impacts on user data privacy and security are highlighted and explained.

4.7.1. Artificial Intelligence (AI)

Artificial Intelligence (AI) is a technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy. This enables them to learn from data, make independent decisions, and carry out tasks without human intervention. [36]

4.7.1.1. Benefits

Libraries are introducing AI-powered virtual assistants and chatbots that offer continuous support and deal with user inquiries. This allows library staff to focus on more meaningful interactions. Benefits of AI in digital libraries include:

- **Personalized Recommendations and Enhanced Discovery:** AI analyses user borrowing history, search queries, and content preferences to suggest relevant resources, improving user experience and helping users discover new materials.
- **Improved Information Retrieval and Accessibility:** AI-powered search engines can better understand complex queries, leading to more precise results. Speech recognition technology can improve accessibility for users with disabilities, enabling them to search and access electronic collections via voice commands.
- **Automated Cataloguing and Metadata Generation:** AI can facilitate the automatic classification and tagging of digital resources and content. This makes resources more organized and searchable, which helps users find and manage data more easily. [37, 38]

4.7.1.2. Risks

- **Data Collection and Privacy Issues:** AI systems often depend on extensive personal data for training and operation, which raises concerns about the collection, usage, and protection of this data, especially when it includes sensitive details such as search histories and reading preferences.
- **Algorithmic Bias:** AI systems may inherit and accentuate biases present in their training data, which can result in biased search results, recommendations, and potentially discriminatory effects on users.

- Lack of Transparency and Control: Users often lack clear options to consent to or withdraw from data collection. Moreover, it can be difficult to understand how their personal information is managed or by whom. Inconsistent data deletion practices further exacerbate these concerns.
- Data Breaches and Security Vulnerabilities: The large datasets required by AI increase the potential for attacks, making AI-dependent systems vulnerable to data breaches that can endanger sensitive user information and damage public trust. [39-42]

4.7.2. Internet of Things (IoT)

The Internet of Things (IoT) refers to the many physical objects equipped with sensors and software that enable them to interact with little human intervention by collecting and exchanging data via a network. [43]

4.7.2.1. Benefits

- Active tracking of environmental conditions, the number of users present in the library, and equipment usage within library buildings.
- The IoT sensors assist librarians by monitoring temperature, humidity levels, and air quality, thus ensuring optimal conditions for users and library material.
- Enhancement of user experience and resource management: IoT has the potential to capture spatial data concerning the library, which could assist in the redesigning of space, management of serviced areas, provision of services in high traffic zones, and other functions, making service delivery much smoother. [44, 45]

4.7.2.2. Risks

- Increasing Attack Surfaces and Security Vulnerabilities: The incorporation of IoT devices in libraries considerably increase the attack surface, as many of these devices lack strong inherent protections or operate under weak security measures, which exposes them to cyber threats and data breaches.
- Extensive Collection of Sensitive Data: IoT devices have the capability to gather highly detailed personal, health, and sensitive information (such as location and movement patterns) that was previously difficult to obtain. This situation creates significant privacy concerns regarding the storage, processing, and potential exploitation of this type of data. [46, 47]

4.7.3. Cloud Computing

Cloud Computing technology refers to providing, via the internet, computing services including servers, storage, databases, networking, software, analytics, and intelligence. [48]

4.7.3.1. Benefits

- Enhanced Accessibility and Remote Access: Cloud-based systems allow users to access library resources from anywhere with an internet connection, breaking down geographical barriers. This improves the value of user data significantly by making it accessible when and where it is needed.
- Scalability and Reliable Data Storage: libraries can now digitize and store increased volumes of user data without restrictions of physical servers due to the cloud computing technology's adjustable storage capabilities. This ensures long-term availability and integrity of information resources and user data.
- Cost Effectiveness: Using cloud computing saves for libraries the costs of purchasing, operating, and maintaining physical storage devices and servers. It also reduces expenditure on frequent hardware and software updates. [49-51]

4.7.3.2. Risks:

- Storing sensitive user data on third-party cloud servers raises concerns about its confidentiality, potential for data loss, and unauthorized access, especially given instances of outages and breaches reported by major cloud providers such as Dropbox, Microsoft, Amazon, etc.
- User data may be stored in various geographical locations, possibly falling under different legal jurisdictions, which can complicate legal accessibility for law enforcement and raise concerns about data sovereignty.
- The shared nature of cloud resources (multi-tenancy) raises trust issues, especially if a breach in a single client's account could affect others.

- Library users may be concerned about surveillance and transparency as their libraries may have limited control over the underlying infrastructure and how cloud service providers manage and monitor user data. [52 - 54]

4.7.4. Blockchain Technology

Blockchain is a distributed digital ledger that securely maintains records across a network of computers in a manner that is transparent, permanent, and resistant to tampering. Each "block" holds information, and these blocks are connected in a sequential "chain". This emerging technology is characterized by its merits of decentralization of control, reliability and consistency of data and transactions, immutability, and anonymity." [55, 56]

4.7.4.1. Benefits

- Data integrity and authenticity: Immutable records ensure accurate historical logs and prevent tampering of metadata and user logs.
- Secure digital preservation: Blockchain supports long-term access and verification of archived materials.
- Decentralization: Reduces reliance on centralized library systems, increasing resilience and transparency.
- Smart contracts: Enable automated circulation, digital rights management, and interlibrary loans. [57]

4.7.4.2. Risks

- Most current blockchain technologies prioritize transparency and auditability, resulting in insufficient privacy for sensitive data. Enhancing privacy usually necessitates extra cryptographic frameworks or consent protocols.
- The immutable nature of blockchain, where data is permanently recorded in the ledger, clashes with privacy regulations like the "right to be forgotten", making it difficult to retract or delete personal data once recorded.
- Despite the fact that transactions are pseudonymous, research indicates that it is possible to de-anonymize blockchain data using graph analytics, which could link activities back to individuals.
- Although blockchain is decentralized in principle, certain aspects, particularly within smart contract implementations, can introduce centralization risks that affect security and access control. [58, 59]

4.7.5. Big Data Analytics

Big data is defined as "the Information asset characterized by such a high Volume, Velocity, and Variety to require specific Technology and Analytical Methods for its transformation into value." Therefore, big data analytics involves the examination of vast amounts of data to reveal hidden patterns, correlations and other insights. There is no single technology that contains big data analytics; several technologies work together to assist in getting the most value from information. Examples of these technologies include cloud computing, data management programs, and data mining technology. [60, 61]

4.7.5.1. Benefits:

Big Data can improve library services while maintaining privacy by analyzing data at the system or macro level rather than individual users. Techniques such as the datafication model, Resource Description and Access (RDA), and linked data technologies facilitate meaningful data use without exposing personal details. [61]

4.7.5.2. Risks:

- Privacy risks arise from the collection and analysis of personally identifiable information (PII), which raises concerns about confidentiality.
- Large data repositories and archives have become potential targets for cyberattacks. This may lead to the re-identification of anonymized data, profiling, discrimination, and a loss of control for users over how their data is used.
- Integrating diverse datasets could unintentionally reveal personal information that is sensitive, which can damage a person's reputation or pose threats to their safety. [62-63]

5. Conclusion

This review article has explored various challenges to user data privacy and security in digital libraries. It outlined the different types of user data collected in digital libraries and the various types and methods of violations faced by this data. These challenges included unauthorized access and cyberattacks, third-party data sharing and integration, and insufficient policies and practices. The consequences of violations and breaches to user data privacy and security were highlighted and their impact on both user data and digital library services was explained. Best practices for protecting user data were presented, including technical, policy, and organizational measures, alongside a discussion of regulatory compliance. In this regard, key principles of data privacy regulations at the global level were highlighted, and the General Data Protection Regulation (GDPR) was reviewed. To illustrate various user data violations and breaches, four real-world notable case studies were discussed, detailing the types of cyberattacks that occurred in each case and the extent of the damages sustained. Mitigation strategies implemented by the affected libraries and the lessons learned from these cyberattacks (e.g. the importance of regular security audits) were discussed. The article closes by reviewing the impact of five emerging technologies recently implemented in digital libraries, elaborating on their benefits and risks to user data. Examined technologies included Artificial Intelligence (AI), Internet of Things (IoT), Cloud Computing, Blockchain technology, and Big Data analytics. Ultimately, digital libraries must harness the benefits of these technologies while mitigating their risks to effectively safeguard user data privacy and security.

References

- [1] Ashikuzzaman M. Types of data libraries can collect and analyze to enhance their services [Internet]. Lisedu Network; 2024 Jun 2 [cited 2025 May 18]. Available from: <https://www.lisedunetwork.com/types-of-data-libraries-can-collect-and-analyze-to-enhance-their-services/>
- [2] Wang J. Usage statistics [Internet]. SPARC Open. [Date of publication unknown] [cited 2025 May 20]. Available from: <https://sparcopen.org/our-work/negotiation-resources/data-analysis/usage-statistics/>
- [3] Porsche L, Suchá LZ, Martinek J. The potential of Google Analytics for tracking the reading behavior in web books. *Digit Libr Perspect.* 2022;38(4). <https://doi.org/10.1108/dlp-03-2022-0021>
- [4] Institute of Electrical and Electronics Engineers (IEEE). Understanding privacy in the digital age [Internet]. c2025 [cited 2025 May 22]. Available from: <https://digitalprivacy.ieee.org/publications/topics/understanding-privacy-in-the-digital-age/>
- [5] International Federation of Library Associations and Institutions (IFLA). Statement on privacy in the library environment [Internet]. 2015 Aug 15 [cited 2025 May 22]. Available from: <https://www.ifla.org/publications/ifla-statement-on-privacy-in-the-library-environment/>
- [6] Singh U, Chaudhari AM. Evaluating strategies for improving user privacy in digital libraries. *IJFANS Int J Food Nutr Sci.* 2022;11(13):49-55. [Internet]. Available from: <https://www.ijfans.org/issue-content/evaluating-strategies-for-improving-user-privacy-in-digital-libraries-13080>
- [7] Kornfeind M. Public libraries and information privacy policies: a case of the Naperville Public Library and privacy trends in the LIS profession. *World Libr.* 2022;26(2). Available from: <https://worldlibraries.dom.edu/index.php/worldlib/article/view/595>
- [8] Noh Y. A critical literature analysis of library and user privacy. *Int J Knowl Content Dev Technol.* 2017;7(2):53-83. <https://doi.org/10.22744/IKCDT.2017.7.2.053>
- [9] Magsi I, Shaheen N, Channar WA, Ali M, Lakho Z, Ahmed A, et al. Cybersecurity challenges in digital libraries. *Rev J Soc Psychol Soc Works.* 2025;3(1):344-50. [Internet]. Available from: <https://rjpsw.com/pdf/v3/i1/11.pdf>
- [10] Matonkar PV. Library security in the digital age: cyber threats and solutions. In: *The Knowledge Nexus.* Creative Book Publisher; 2024. Available from: https://www.researchgate.net/publication/385092279_Library_Security_in_the_Digital_Age_Cyber_Threats_and_Solutions
- [11] Saha R. Data privacy and cybersecurity in digital library perspective: safeguarding user information. *Int J Sci Res Eng Manag.* 2024;8(4). <https://doi.org/10.55041/IJSREM30761>
- [12] Understanding data breaches: what you need to know [Internet]. EC-Council University. 2025 May [cited 2025 May 25]. Available from: <https://www.eccu.edu/blog/data-breaches-threats-and-consequences/>

- [13] Heckel J. Study shows challenges to protecting privacy of library users [Internet]. University of Illinois, Strategic Communication and Marketing, News Bureau. 2023 Dec 7 [cited 2025 Jun 20]. Available from: <https://news.illinois.edu/study-shows-challenges-to-protecting-privacy-of-library-users/?hl=en-GB>
- [14] Cetin MB. Evaluating the effects of digital privacy regulations on user trust [master's thesis]. 2024 [cited 2025 May 25]. Available from: <https://arxiv.org/pdf/2409.02614>
- [15] Aregbesola A, Nwaolise EL. Securing digital collections: cyber security best practices for academic libraries in developing countries. *Libr Philos Pract* (e-journal). 2023;7822. Available from: <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=15048&context=libphilprac>
- [16] Public libraries and cybersecurity: keeping patrons' data safe [Internet]. PressReader Blog. 2024 Jul 17 [cited 2025 May 27]. Available from: <https://blog.pressreader.com/libraries-institutions/public-libraries-and-cybersecurity-keeping-patrons-data-safe?hl=en-GB>
- [17] What is regulatory compliance? [Internet]. Cynomi. [Date of publication unknown] [cited 2025 Jun 3]. Available from: <https://cynomi.com/learn/regulatory-compliance/>
- [18] Institute of Electrical and Electronics Engineers (IEEE). Emerging data privacy laws and regulations around the world [Internet]. c2025 [cited 2025 Jun 3]. Available from: <https://digitalprivacy.ieee.org/publications/topics/emerging-data-privacy-laws-and-regulations-around-the-world/>
- [19] Perry Y. Data compliance for regulations around the world [Internet]. NetApp Blog. 2019 Sep 17 [cited 2025 Jun 3]. Available from: <https://www.netapp.com/blog/data-compliance-regulations-hipaa-gdpr-and-pci-dss/>
- [20] Warren T. The Internet Archive is back as a read-only service after cyberattacks. *The Verge* [Internet]. 2024 Oct 14 [cited 2025 Jun 10]. Available from: <https://www.theverge.com/2024/10/14/24269741/internet-archive-online-read-only-data-breach-outage>
- [21] The British Library. British Library cyber incident review. 2024 Mar 8. [Internet]. [cited 2025 June 10]. Available from: <https://cdn.sanity.io/files/v5dwkion/production/99206a2d1e9f07b35712b78f7d75fbb09560c08d.pdf>
- [22] British Library cyber attack explained: what you need to know [Internet]. *ComputerWeekly.com*. 2024 Jan 15 [cited 2025 Jun 10]. Available from: <https://www.computerweekly.com/feature/British-Library-cyber-attack-explained-What-you-need-to-know>
- [23] Kahle B. Internet Archive Services Update: 2024-10-17 [Internet]. *Internet Archive Blogs: a blog from the team at archive.org*. 2024 Oct 18 [cited 2025 Jun 12]. Available from: <https://blog.archive.org/2024/page/4/>
- [24] Kan M. After breach, Internet Archive expects to return within 'days, not weeks' [Internet]. *PCMag*. 2024 Oct 11 [cited 2025 Jun 14]. Available from: <https://www.pcmag.com/news/after-breach-internet-archive-expects-to-return-within-days-not-weeks>
- [25] The Internet Archive breach continues [Internet]. *Help Net Security*. 2024 Oct 21 [cited 2025 Jun 14]. Available from: <https://www.helpnetsecurity.com/2024/10/21/internet-archive-breach-continues/>
- [26] Kahle B. Learning from cyber-attacks [Internet]. *Internet Archive Blogs: a blog from the team at archive.org*. 2024 Nov 14 [cited 2025 Jun 15]. Available from: <https://blog.archive.org/tag/ddos/>
- [27] zgzsus. What I learned from the Internet Archive hack and its aftermath: personal reflections [Internet]. *zgzsus Personal cat edu*. 2024 Nov 10 [cited 2025 Jun 15]. Available from: <https://wp.catedu.es/zgzsus/what-i-learned-from-the-internet-archive-hack-and-its-aftermath/>
- [28] DomainTools. Rogue hackers and the Internet Archive breach: 31 million accounts exposed [podcast]. *Breaking Badness Podcast*. 2024 Oct 23 [cited 2025 Jun 12]. Available from: <https://www.domaintools.com/resources/podcasts/rogue-hackers-and-the-internet-archive-breach-31-million-accounts-exposed/>
- [29] Işık Ö. Full transparency: 10 lessons from the cyber-attack on the British Library [Internet]. *IMD.org*. 2024 Sep 30 [cited 2025 Jun 15]. Available from: <https://www.imd.org/ibyimd/technology/full-transparency-10-lessons-from-the-cyber-attack-on-the-british-library/>
- [30] Learning lessons from the cyber-attack [Internet]. *ISTARI*. 2024 Jul [cited 2025 Jun 14]. Available from: <https://istari-global.com/insights/spotlight/learning-lessons-from-the-cyber-attack/>
- [31] Enhancing cyber resilience: lessons from the British Library's ransomware attack [Internet]. *Ascot London*; 2024 Oct 15 [cited 2025 Jun 17]. Available from: <https://ascot.london/enhancing-cyber-resilience-lessons-from-the-british-librarys-ransomware-attack>

- [32] Toronto Public Library. Cybersecurity response and business continuity [Internet]. 2024. Available from: <https://www.urbanlibraries.org/innovations/cybersecurity-response-and-business-continuity>
- [33] Toronto Public Library cyberattack: a wake-up call for stronger security [Internet]. Information and Privacy Commissioner of Ontario. 2025 Mar 24 [cited 2025 Jun 17]. Available from: <https://www.ipc.on.ca/en/cases-of-note/toronto-public-library-cyberattack>
- [34] City taking action to prevent cyber attacks that targeted our libraries, hospitals and zoo [Internet]. Toronto Newswire. 2024 Feb 12 [cited 2025 Jun 20]. Available from: <https://torontonewswire.com/the-city-taking-action-to-prevent-cyber-attacks-that-has-rocked-our-libraries-and-zoo/>
- [35] Data privacy in 2025: what lies ahead? Trends and predictions [Internet]. TrustCloud Community. 2025 Apr 2 [cited 2025 Jun 20]. Available from: <https://community.trustcloud.ai/article/data-privacy-in-2025-what-lies-ahead-trends-and-predictions/>
- [36] Stryker C, Kavlakoglu E. What is artificial intelligence (AI)? [Internet]. IBM; 2024 Aug 9 [cited 2025 Jun 19]. Available from: <https://www.ibm.com/think/topics/artificial-intelligence>
- [37] Narendra AP, Gunawan LS, Setiawan ASA. Artificial intelligence implementation in library information systems: current trends and future studies. *Vietnam J Comput Sci.* 2024;1(25). <https://doi.org/10.1142/S2196888824300023>
- [38] Prasanna R, Yogendra S. The role of artificial intelligence in enhancing digital library services. *JETIR.* 2023 Dec;10(12):f712-f723. Available from: https://www.researchgate.net/publication/390089610_The_Role_of_Artificial_Intelligence_in_Enhancing_Digital_Library_Services
- [39] Timonera K. Understanding AI privacy: key challenges and solutions [Internet]. eWeek. 2024 Nov 27 [cited 2025 Jun 22]. Available from: <https://www.eweek.com/artificial-intelligence/ai-privacy-issues/?hl=en-GB>
- [40] jrose. Unwelcome AI: examining the negative impacts on libraries [Internet]. Liblime Blog. 2024 Oct 31 [cited 2025 Jun 22]. Available from: <https://liblime.com/2024/10/31/unwelcome-ai-examining-the-negative-impacts-on-libraries/?hl=en-GB>
- [41] Gunter D. AI challenges for librarians [Internet]. Research Information. 2024 Feb 22 [cited 2025 Jun 12]. Available from: <https://www.researchinformation.info/analysis-opinion/ai-challenges-librarians/?hl=en-GB>
- [42] Granados A. AI and personal data: balancing convenience and privacy risks [Internet]. Velaro Blog. 2024 Nov 25 [cited 2025 Jun 19]. Available from: <https://velaro.com/blog/the-privacy-paradox-of-ai-emerging-challenges-on-personal-data?hl=en-GB>
- [43] Greengard S. Internet of Things [Internet]. Britannica. 2025 Jun 7 [cited 2025 Jun 14]. Available from: <https://www.britannica.com/science/Internet-of-Things>
- [44] Kumar BTS, Usha SS. IoT-based services in digital libraries: Innovations, benefits and concerns. In: *Proceedings of the International Conference on Emerging Trends in Information Technology*; 2024 Mar 15–16; Bengaluru, India. Lincoln (NE): *Library Philosophy and Practice (e-journal)*; 2024. Available from: <https://digitalcommons.unl.edu/libphilprac/1479>
- [45] PressReader Team. Enhancing patron and staff experience: the Internet of Things in libraries [Internet]. PressReader Blog; 2024 Nov 12 [cited 2025 Jun 15]. Available from: <https://blog.pressreader.com/libraries-institutions/enhancing-patron-and-staff-experience-the-internet-of-things-in-libraries>
- [46] Tawalbeh L, Muheidat F, Tawalbeh M, Quwaider M. IoT privacy and security: challenges and solutions. *Appl Sci (Basel).* 2020;10(12):4102. <https://doi.org/10.3390/app10124102>
- [47] Office of the Victorian Information Commissioner. Internet of Things and privacy – issues and challenges [Internet]. [Date of publication unknown] [cited 2025 Jun 04]. Available from: <https://ovic.vic.gov.au/privacy/resources-for-organisations/internet-of-things-and-privacy-issues-and-challenges/?hl=en-GB>
- [48] What is cloud computing? [Internet]. Microsoft Azure. c2025 [cited 2025 Jun 20]. Available from: <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cloud-computing/>
- [49] Indraj C, Satishkumar N. Application of cloud computing in libraries: prospects and challenges. *Int J Novel Res Dev.* 2025 Feb;10(2):b315-b322. Available from: <http://eprints.rclis.org/46434/1/IJNRD2502143.ocr.pdf>
- [50] Kahn M. Cloud computing in libraries: revolutionizing the way we store and access knowledge [Internet]. MK Library. 2025 Feb 27 [cited 2025 Jun 22]. Available from:

- <https://www.mklibrary.com/cloud-computing-in-libraries-revolutionizing-the-way-we-store-and-access-knowledge/>
- [51] Yadav P. Cloud computing in university libraries: transforming access, management, and service delivery. *Int Res J Eng Technol*. 2024 Sep;11(9):649-54. Available from: <https://www.irjet.net/archives/V11/i9/IRJET-V11I994.pdf>
- [52] GeeksforGeeks. 7 privacy challenges in cloud computing [Internet]. GeeksforGeeks. 2025 May 24 [cited 2025 Jun 14]. Available from: <https://www.geeksforgeeks.org/privacy-challenges-in-cloud-computing/?hl=en-GB>
- [53] Kadali S, Chary KD. Key issues and challenges in cloud computing for library services. *IOSR J Eng*. 2017 Dec;7(12):80-2. <https://doi.org/10.9790/9622-0712038082>
- [54] The Investopedia Team. Cloud Security: Definition, How Cloud Computing Works, and Safety [Internet]. Investopedia; 2022 Sep 30 [cited 2025 Jul 10]. Available from: <https://www.investopedia.com/terms/c/cloud-security.asp>
- [55] Mullsjafari S, Bechkoum K. Blockchain technology and related security risks: towards a seven-layer perspective and taxonomy. *Sustainability*. 2023;15(18):13401. <https://doi.org/10.3390/su151813401>
- [56] Hanif S, et al. Blockchain technology in libraries: ensuring data security and transparency. *J Appl Linguist Tesol*. 2025;8(1). Available from: <https://jalt.com.pk/index.php/jalt/article/view/423>
- [57] Rao MH, Prasad GS. Enhancing security and privacy in digital libraries using Blockchain technology. In: *Proceedings of the 7th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI 2025)*; 2025 Jan 24–25; Goa, India. Leiden, Netherlands: Atlantis Press; 2025. p. 82–90. doi: 10.2991/978-94-6463-712-0_9
- [58] Usman M, Ullah A. Application of Blockchain Technology in Digital Libraries. In: *Proceedings of the International Conference on Advanced Technologies in Computing and Systems (ATICS 2024)*; 2024 May 10–12; Islamabad, Pakistan. Islamabad: AvePubs; 2024. Available from: <https://avepubs.com/uploads/articles/171633893072904.%20ATICS-09-2024.pdf>
- [59] Xu H, Zhang N. Privacy implications of blockchain systems: a data management perspective. *Organ Cybersec J Pract Process People*. 2023 Sep. <https://doi.org/10.1108/ocj-01-2023-0003>
- [60] Big data analytics: what is it and why it matters [Internet]. SAS. [Date of publication unknown] [cited 2025 Jul 08]. Available from: https://www.sas.com/en_us/insights/analytics/big-data-analytics.html/
- [61] Harper M, Oltmann SM. Big data's impact on privacy for librarians and information professionals. *Bull Assoc Inf Sci Technol*. 2017 Jul-Aug;43(4):19-23. <https://doi.org/10.1002/bul2.2017.1720430406>
- [62] Chancey T. Big data privacy issues: protect your data with advanced analytics and security [Internet]. The Scarlett Group; 2024 Jan 12 [cited 2025 July 8]. Available from: <https://www.scarlettgroup.com/big-data-privacy-concerns/>
- [63] Masinde J, Mugambi F, Muthee DW. Big data and personal information privacy in developing countries: insights from Kenya. *Front Big Data*. 2025 Apr 4;8:1532362. <https://doi.org/10.3389/fdata.2025.1532362>