Cybersecurity for digital libraries: an interview with Emanuele Bellini

One of the main challenges and risks associated with the digital preservation of digital libraries is ensuring their security and privacy. Data breaches, cyberattacks, unauthorized access and data loss are some of the potential threats that can compromise data integrity, confidentiality and availability. The recent attack on the British Library [1] has highlighted the need to prioritize the cybersecurity of digital libraries. In addition to protecting digital libraries, strong cybersecurity practices increase user trust, promote collaboration between institutions and support users' privacy in a secure environment.

We have interviewed Emanuele Bellini to understand cybersecurity risks, strategies and best practices.

Emanuele Bellini is a Professor at the University of Rome Tre (Italy) and Chair of the IEEE SMC (Man and Cybernetics Society) TC (Technical Committee) Cyber Humanities and Chair of the IEEE CSR (Cybersecurity Resilience) conference series. He has been involved in many research projects on various aspects of digital libraries, such as digital preservation, persistent identifiers and metadata. His current research interest is on cybersecurity resilience of digital heritage institutions as critical infrastructures for society. The emerging field of cyber resilience can be understood as a blend of strategies, methods and techniques to support the complex adaptive capacity of the digital libraries during cyberattacks.

Q1. What are the most common cybersecurity threats faced by digital libraries today?

The digital library is no longer simply a repository for storing and cataloging digital materials. In our increasingly interconnected world, where the lines between physical and digital are blurred, digital libraries play a crucial role in preserving and sharing knowledge. They are integral parts of a complex human-cyber-physical system, tasked with safeguarding information and ensuring its reliability and accessibility. As such, they should be recognized as critical infrastructure, akin to transportation and energy systems. However, like any infrastructure, digital libraries face various threats. Malicious software, such as malware and ransomware, can compromise data integrity and restrict access to resources. Denial-of-service attacks can disrupt services by overwhelming servers with traffic. Data breaches pose risks of sensitive information exposure, whereas phishing schemes target operators to gain unauthorized access. In addition, insider threats, where individuals misuse their access privileges, are a concern. Failure to promptly update systems can leave libraries vulnerable, as seen in the 2023 breach at the British Library.

Digital Library Perspectives
Vol. 40 No. 2, 2024
pp. 348-355

2059-5816
DOI 10.1108/DLP-05-2024-147

Q2. What specific vulnerabilities make digital libraries attractive targets for cyberattacks?

The primary vulnerability of digital libraries lies in their intricate network of interdependencies, tightly woven into the fabric of critical infrastructure. These libraries often share resources, metadata and authentication systems with academic and research institutions, creating opportunities for attackers to exploit these connections and infiltrate multiple networks or data repositories simultaneously. Linked Open Data, a cornerstone of digital library success, relies on the presumption of continued accessibility and reliability. However, an attack on one node within this system can trigger cascading effects, both physical (e.g. server failures and data corruption) and social (e.g. loss of trust in information and susceptibility to manipulated data), with profound implications for research and society.

Furthermore, digital libraries represent enticing targets for cyberattacks because of the wealth of valuable data they harbor, including research papers, historical documents, personal information and intellectual property. Attackers may seek financial gain, ideological objectives, espionage or simply aim to disrupt operations. As integral components of the human-cyber-physical system, digital libraries frequently collaborate with various entities, expanding their network complexity and susceptibility to supply chain attacks or insider threats. Compounded by the use of outdated software and legacy systems with known vulnerabilities, digital libraries face heightened risks of unauthorized access, data manipulation and other malicious activities. The aftermath of such attacks can be financially burdensome and impede the return to normal operations.

In addition, a false sense of security surrounds open access resources, as their openness does not render them immune to tampering. In fact, altered open access content can spread rapidly across the global digital landscape, posing challenges for its detection and the related mitigation actions. Managing the security of diverse content formats, including text, images, audio and video, further strains cybersecurity resources and expertise within digital libraries. Moreover, libraries of significant social or political importance become prime targets for politically motivated cyberattacks seeking to manipulate, censor or disrupt information dissemination.

Q3. What are the potential consequences of a successful cyberattack on a digital library?

The potential consequences of a successful cyberattack on a digital library span several critical domains. One of the primary concerns is the potential loss or theft of valuable heritage collection/data, which encompasses research papers, historical documents, personal user information, intellectual property and other sensitive materials. In instances where the digital copy serves as the sole remaining version of a lost physical object, any loss or alteration of this copy results in its irreversible destruction or alteration of heritage.

Another significant impact is corruption, leading to inaccuracies, compromised integrity or rendering the data unusable, but one of the critical emerging risks is related to heritage manipulation (basically using artificial intelligence [AI]). Even if the immediate impact is undermining the trustworthiness of the digital library's resources, the long-term impact could be altering the heritage of a community with consequences that need to be carefully assessed. Furthermore, a cyberattack can disrupt digital services, rendering collections inaccessible to users and affecting research, education and other activities reliant on digital resources' availability.

Financial implications may arise due to the expenses associated with incident response, data recovery, legal fees, regulatory fines and potential loss of revenue or funding.

Moreover, a cyberattack can tarnish the digital library's reputation, diminishing its credibility and trustworthiness among users, researchers, academic institutions, funding agencies and the general public. Depending on the breach's nature and relevant regulations, such as general data protection regulation (GDPR), the digital library may face legal consequences, including fines, lawsuits and regulatory sanctions for failing to adequately safeguard sensitive information.

Finally, digital libraries often house valuable intellectual property, including copyrighted materials, proprietary research and unpublished works, which could be targeted for theft or unauthorized distribution, resulting in financial losses and reputational harm to content creators and rights holders.

Q4. How do digital libraries typically detect and respond to security incidents?

The attack on the British Library demonstrates that many digital libraries, even prestigious ones, are not equipped to deal with the growing threats coming from the cyber domain. Often technologies already in use in other critical infrastructures are not used such as security information and event management systems or advanced anomaly detection systems or vulnerability scanning tools. Generally, the organizations are mainly focused on business continuity to mitigate the effects of accidental system failure. It is evident that this approach is no longer valid to address the emerging security threats related to the digital assets managed.

The paradigm shift required by the last cyber incidents (but also by the Ukrainian 's warfare) will have to make it possible to improve both the technological equipment but also the organization and skills for cybersecurity and defense. For example, cybersecurity skills could be developed in librarians by training new profiles, or by pairing them with experts in the sector.

Q5. What measures and strategies can digital libraries take to protect against insider threats or unauthorized access to sensitive communities' information?

Digital libraries ought to establish access control policies delineating roles, permissions and privileges for personnel, contractors and users, adhering to the principle of least privilege. Access should be granted solely on a need-to-know basis, restricting individuals' access to sensitive data to the essentials of their respective roles.

Implementing robust authentication mechanisms, like multifactor authentication (MFA) or biometric authentication, can verify users' identities when accessing sensitive information, enhancing security measures.

Moreover, staff members and users should undergo training to familiarize themselves with security best practices, emphasize the importance of safeguarding sensitive data and educate them on identifying and reporting suspicious activities or insider threats.

To effectively monitor and audit user activities, including those of vendors and third-party collaborators, as well as access patterns and alterations to sensitive data, implementing comprehensive monitoring and auditing mechanisms is essential. This involves logging user actions, monitoring network traffic and conducting routine audits of access permissions and utilization.

Furthermore, deploying data loss prevention (DLP) solutions to oversee and regulate the movement of sensitive data within the digital library's environment is critical. These solutions can identify and thwart unauthorized access, transmission or storage of sensitive information.

Enforcing role-based access control mechanisms is another vital aspect, ensuring granular access controls based on users' roles, responsibilities and organizational hierarchy. Role-based access control guarantees that individuals access only the information pertinent to their job functions. For more advanced solutions such as behavioral analytics, their implementation should be assessed only if the risk analysis justifies it.

Q6. How important is user awareness and education in preventing security breaches in digital libraries?

The human element within a human-cyber-physical system holds paramount importance. Indeed, as indicated by an IBM report, nearly all cyberattacks stem from the intentional or unintentional actions of individuals operating within the system.

Users, comprising library staff, researchers, students and various stakeholders, can inadvertently introduce vulnerabilities through actions such as clicking on malicious links, falling victim to phishing scams or using weak passwords. While the system must impose safeguards to minimize errors, excessive restrictions may hinder normal operations. Hence, it becomes imperative to identify measures aimed at enhancing awareness and competence in cybersecurity.

By increasing awareness regarding phishing tactics and educating users on identifying and responding to suspicious emails or messages, digital libraries can mitigate the risk of successful phishing attacks. Moreover, educating users on the significance of employing robust, unique passwords and implementing MFA serves to fortify authentication mechanisms and safeguard against unauthorized access.

These initiatives should be integrated into the promotion of a culture of cybersecurity awareness and accountability within the organization. Such a culture fosters proactive risk management, continual training and learning, and a collective responsibility for upholding security standards.

Q7. What role does encryption play in securing data within digital libraries?

Encryption plays a pivotal role in managing data within digital libraries. Numerous regulatory frameworks and industry standards, such as GDPR, HIPAA, FERPA and PCI DSS, mandate organizations to incorporate encryption into their data protection strategies.

By converting sensitive data into an unreadable format that can only be deciphered with the correct cryptographic key, encryption ensures the confidentiality of information. This effectively prevents unauthorized access by cyberattackers or malicious insiders, even if they manage to breach storage or transmission channels. However, implementing encryption necessitates a thorough assessment of its cost-effectiveness.

For example, digital libraries typically house vast amounts of sensitive data, including research papers, manuscripts, personal information and intellectual property. Data-at-rest encryption safeguards this data while it resides on servers, databases or storage devices, shielding it from unauthorized access in the event of a security breach or physical theft.

Conversely, during data transmission across networks like the internet or internal infrastructure, data is susceptible to interception by eavesdroppers. Data-in-transit encryption, often achieved through protocols like transport layer security or secure sockets layer, encrypts data during its journey, ensuring both confidentiality and integrity.

While the adoption of encryption may yield evident benefits for closed digital libraries (e.g. IEEExplore), it could be perceived as a hindrance to the free flow of knowledge for open-access data. Hence, decisions regarding encryption must be evaluated based on the specific business case at hand.

Q8. How can digital libraries ensure compliance with relevant data protection and privacy regulations?

Ensuring adherence to pertinent regulations such as common regulations including the GDPR stands as one of the foremost responsibilities for organizations. Digital libraries must thoroughly grasp the data protection and privacy regulations applicable to their operations, considering factors like geographic location, user jurisdictions and the nature of data collected and processed. A pivotal initial step involves conducting a comprehensive data inventory and assessment to identify and categorize the personal data collected, stored, processed and transmitted by the digital library.

Subsequently, it becomes imperative to craft and uphold clear and transparent privacy policies and notices outlining the minimal personal data necessary to fulfill collection purposes, as well as data retention policies stipulating duration and criteria for deletion or anonymization when no longer required. In addition, informing users about data collection types, usage, sharing recipients and user data rights is essential.

Furthermore, it is crucial to develop consent forms to secure explicit user consent before collecting, processing or sharing personal data, where mandated by regulations. Clearly articulating data usage purposes and providing opt-in or opt-out choices for users form integral parts of compliance assessments.

Regular monitoring of compliance with data protection and privacy regulations through audits, assessments and policy, procedure and data processing activity reviews is paramount. Establishing mechanisms for prompt and transparent reporting and resolution of data breaches or noncompliance incidents is indispensable. In this vein, engaging legal and compliance expertise to obtain guidance and remain abreast of regulatory changes affecting digital library operations is advisable.

Q9. How should digital librarians prioritize cybersecurity investments and resources to effectively mitigate risks?

Effectively prioritizing cybersecurity investments and resources is a multicriteria decision-making process that demands digital librarians to evaluate their risks, vulnerabilities and operational needs. The objective is to strike a balance among investments in preventive, detective and responsive security measures to establish a robust and layered defense against cyber threats.

Resource allocation should stem from an integrated and dynamic risk assessment, aimed at identifying and ranking cybersecurity risks, both known and unknown, specific to the digital library's operations, assets and data. This involves gauging the likelihood and potential impact of various known, hybrid and emerging threats; vulnerabilities; and security incidents on the confidentiality, integrity and availability of library resources. To achieve this, creating an inventory of digital library assets, encompassing hardware, software, networks, data repositories, applications and human resources, is crucial, considering the criticality and sensitivity of these assets to the library's mission and operations.

Furthermore, it is imperative to recognize applicable data protection and privacy regulations, industry standards and best practices pertinent to the digital library's operations to mitigate legal and financial risks associated with noncompliance.

Implementing foundational security controls, tools and best practices to tackle common and specific cybersecurity risks and vulnerabilities also warrants consideration.

In addition, evaluating the level of user awareness and the effectiveness of training programs aimed at educating library staff, researchers, students and other stakeholders

about cybersecurity risks, as well as their roles and responsibilities in safeguarding library resources and data, provides valuable insights for making informed decisions.

Q10. What are some best practices for establishing a robust cybersecurity strategy in a digital library environment?

The organization's primary recommendation is to embrace a cyber resilience perspective, aiming for a comprehensive understanding of its security landscape. Numerous frameworks, such as those developed by National Institute of Standards and Technology or National Academy of Science, are available in the literature to assist organizations in formulating a cohesive strategy. The rationale behind this approach is to bridge the gap between sustaining digital library operations and fulfilling its mission. Cyber resilience, therefore, transcends mere resistance to breaches; it entails learning from breach attempts and adapting to the ever-evolving threat landscape. Shifting the focus from security compliance to proactive security measures and adaptability is imperative.

Cyber resilience typically comprises four phases: prepare, absorb, recover and adapt, each necessitating specific actions. In the preparation phase, conducting risk and capacity assessments becomes essential. These assessments evaluate various factors, including the types of stored data, access points, network infrastructure, third-party dependencies, regulatory obligations, organizational awareness, expertise levels and potential threat actors.

Three pillars emerge from this framework: organization, awareness and technology. Establishing a robust cybersecurity governance structure within the organization, defining policies, roles and responsibilities and appointing a chief information security officer or cybersecurity lead to oversee initiatives, coordinate efforts, review policies and manage stakeholder communication is paramount – the first pillar.

The second pillar emphasizes security awareness and training. Regular training sessions for library staff, researchers, students and stakeholders are essential to increase awareness of cybersecurity risks, best practices and individual responsibilities in safeguarding library resources and data. Training focuses on identifying phishing attacks, adopting safe browsing habits and promptly reporting suspicious activities, bolstering the organization's overall security posture.

The third pillar revolves around implementing appropriate technologies. Access control and identity management ensure only authorized users access library resources, while network security measures like firewalls and intrusion detection and prevention systems safeguard against external threats. Endpoint security measures protect devices, and developing an incident response plan outlines procedures for detecting, responding to and recovering from security incidents and breaches.

By adhering to these principles and adopting a holistic cybersecurity approach, digital libraries can craft a robust strategy to safeguard their collections, services and users' information against cyber threats. This ensures the integrity, confidentiality and availability of library resources in the face of evolving cyber risks.

Q11. How can digital libraries collaborate with other organizations or institutions to enhance cybersecurity resilience?

A digital library is part of a complex human-cyber-physical system with many interdependencies, hence establishing reliable and effective collaborations with the other nodes is crucial to enhance cybersecurity resilience for digital libraries. Assuming that the information sharing is at the core of each collaboration, here a several ways digital libraries can collaborate are proposed:

- Collaborate with universities and research institutions, to leverage their expertise, resources and research capabilities in cybersecurity.
- Collaborate with government agencies, law enforcement, regulatory bodies and industry associations to address cybersecurity challenges and share insights into cyber threats, trends and mitigation strategies.
- Establish partnerships for vendor risk management, supply chain security assessments and information sharing on security threats, vulnerabilities and best practices.
- Establish partnerships with other organizations or institutions, such as Computer Security Incident Response Teams or Information Sharing and Analysis Centers, to facilitate collaboration and coordination in responding to cyber threats.
- Conduct joint cyber exercises, tabletop simulations and red team–blue team exercises with other organizations or institutions to test and improve cybersecurity resilience.
- Explore opportunities to share security services, tools and resources with other organizations or institutions to enhance cybersecurity resilience cost effectively.

Q12. Are there any emerging trends or technologies in cybersecurity that are particularly relevant to digital libraries?

Several emerging trends are reshaping the landscape of cybersecurity and four technologies stand out as potentially relevant for digital libraries.

First, zero trust architecture (ZTA) embodies an approach that assumes no inherent trust, mandating rigorous identity verification and access controls for every user, device and application seeking access to network resources. ZTA presents an opportunity for digital libraries to bolster their security posture by mitigating the risk of unauthorized access to sensitive information and resources.

Second, DLP solutions serve as a vital tool in thwarting the unauthorized disclosure or leakage of sensitive data by monitoring, detecting and enforcing policies governing data use, sharing and storage. By using DLP technologies, digital libraries can ensure compliance with data protection regulations, prevent data breaches and safeguard intellectual property and user privacy.

Third, machine learning (ML) and AI technologies are increasingly finding utility in cybersecurity, offering capabilities for threat detection, anomaly detection, behavior analysis and automated response. Digital libraries can harness ML and AI tools to augment their security measures, enabling more effective threat detection and response while adapting to evolving cyber threats in real time.

Finally, blockchain technology presents opportunities for decentralized and immutable data storage, authentication and tamper-proof audit trails. This technology holds promise for enhancing the security and integrity of digital library collections, particularly in preserving and verifying the authenticity of historical records, research data and intellectual property.

Emanuele Bellini
Department of Humanistic Studies, Roma Tre University, Rome, Italy, and

Anna Maria Tammaro
Department of Information Engineering, University of Parma, Parma, Italy

Note

1. The British Library suffered one of the most critical attacks on a cultural heritage institution. The criminal gang responsible for the attack copied and exfiltrated (illegally removed) some 600GB of files, including personal data of library users and staff. As well as the exfiltration of data for ransom, the attackers' methods included the encryption of data and systems, and the destruction of some servers to inhibit system recovery and to cover their tracks. The latter has had the most damaging impact on the library: while they have secure copies of all their digital collections – both born-digital and digitized content and the metadata that describes it – they have been hampered by the lack of viable infrastructure on which to restore it because of its reliance on a significant number of aging legacy applications. Read British Library (2023) *Learning lessons from the cyber-attack* www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf