Metadata in Present and Future: A Comparative Analysis

Konstantinos Zafeiropoulos
Department of Informatics and Computer Engineering
University of West Attica
Athens, Greece
ice20390293@uniwa.gr

ABSTRACT

Metadata is a critical and fundamental element of modern digital systems as it is required for the proper management, retrieval and simultaneous preservation of data between various digital structures. In this particular article, the results of an extensive bibliographic comparative analysis are recorded on the subject of metadata application fields, their past and their future. More specifically, research was conducted on the practical applications of metadata in digital libraries, science, research, as well as Artificial Intelligence (AI), cloud services and cybersecurity. At the same time, the role of metadata in ensuring the integrity of data was investigated by analyzing its regulatory standards as well as its adaptability in the modern era were technologies like AI and Blockchain conquer. Finally, concerns, challenges and proposals for the future of metadata, like the need for interoperability, scalability and legal frameworks, in the rapidly developing epoch were listed.

KEYWORDS

Metadata Management, Data Integrity, Metadata Standards, AI, Metadata in Cybersecurity

1. Introduction

Metadata is commonly defined as "data about data" and is essential for organizing, retrieving and preserving digital content. Also known as information about information, standardized and structured data accessible and easily processed by machines. It is common and logical to observe misinterpretations and to incorrectly coincide some definitions such as that of XML. Specifically, XML is a popular standard for structuring and exchanging data and identifiers, which provides unique names for specific content or intellectual property (Gill, 1998).

Moreover, metadata, as applied within digital information systems, is categorized into several types based on their function: Administrative, Descriptive, Preservation, Use, and Technical as noted by Gilliland (1998). Administrative metadata aids in the management of information resources, covering aspects like rights tracking, version control and data acquisition. On the other hand, Descriptive metadata identifies and describes resources, including cataloging records and user annotations. Preservation metadata documents the physical state and preservation actions for both digital and physical resources, like data migration efforts. Technical metadata pertains to system functionalities, including information on hardware, software and system performance and finally, Use metadata tracks the usage and user interaction with resources, supporting exhibit records and user tracking systems. Nevertheless, metadata are also characterized by other attributes such as their source, which can be internal (generated during creation) or external (added later by third parties) and their status, that may be static or dynamic, adapting to interactions and modifications over time Gilliland (1998). For instance, structured metadata follows a predictable format, like MARC or TEI, while unstructured metadata lacks such conformity. Additional significant attributes are the Semantics where controlled metadata adheres to standardized vocabularies, unlike uncontrolled free-text notes and the level, if we refer to entire collections or focusing on individual items Gilliland (1998).

In the constantly developing digital world where information continues to grow in volume and complexity, metadata provides the foundation for managing a vast array of content effectively. Only by referring to its categories it is obvious why it plays a crucial role across domains like cultural heritage institutions, digital libraries and modern sciences (Gill, 1998), (Gilliland, 1998), (Swedlow, et. al. 2010). More specifically, metadata is important in cultural heritage institutions, because it ensures that digital representations of artifacts and collections are properly cataloged and preserved. Standardized metadata practices are key to avoiding isolated data silos and enabling institutions to share resources and improve their accessibility across digital platforms based on Gill (1998). Gilliland (1998) also emphasizes the dual need for metadata which translates to being both machine-readable and simultaneously user-friendly, allowing institutions with varying levels of technical expertise to manage their digital content effectively. In this case metadata acts as a bridge between different systems, facilitating the exchange of information between institutions being responsible for sharing resources seamlessly across platforms and large-scale digital environments (Caplan, 2001). Arms (2000) echoes the aforementioned sentiment, arguing that without consistent metadata standards, digital libraries risk fragmentation, which could lead to inefficiencies in managing or accessing digital resources.

Metadata also prevail in the sciences and for instance in biological sciences where the management of large datasets like image data is required. Well-structured metadata is essential for ensuring the usability and reproducibility of complex data in research and scientific authors argue that without proper metadata valuable research data could become difficult to reuse or even inaccessible (Swedlow, et. al. 2010). Despite the above, metadata's role doesn't stop there since it's also critical in other fields of science like AI, by providing needed structured data for machine learning algorithms (Baca, 2008). It is highlighted that while AI technology evolves, metadata systems must also evolve and adapt to be able to accommodate the increased complexity of data processing tasks (Brand et. al., 2003), (Swedlow, et. al., 2010).

Through the following analysis, this paper analyzes even more the defined terms of introduction and provides a comparative evaluation of articles examining challenges and opportunities associated with metadata while presenting a comprehensive understanding of how metadata can be optimized meeting the demands of modern digital systems.

2. Metadata Standards and Interoperability

Metadata's power and uses are intertwined with the existence of constantly evolving standards. The development of metadata standards has been a central concern especially for institutions managing large digital collections. These standards ensure that data can be effectively organized and retrieved, regardless of the platform or the institution using it. The challenge lies in achieving interoperability while simultaneously ensuring that different systems, using varied metadata schemas, can communicate and share information efficiently and securely.

Gill (1998) discusses how metadata is essential for organizing large digital collections and points out that metadata functions like a catalog, offering concise and well-structured descriptions making it easier to manage information. The effectiveness of metadata hinges on its consistency and standardization, which allows different systems to share and access data seamlessly resulting in a more structured and uniform approach (Gill, 1998). The complexity of managing artifacts in digital form demands a well-defined metadata standard that can support the long-term preservation and accessibility of these digital assets while providing foundation for consistent retrieval and interaction across collections, reducing the fragmentation of data systems (Gill, 1998). In an analogous way Caplan (2001) discusses the importance of metadata for digital libraries, in the context of cross-institutional collaboration. More specifically notes that the lack of standardization across different metadata systems can lead, to incompatibilities, which in turn hinders the ability of institutions to share and access data (Caplan, 2001).

Till now key efforts in establishing web metadata standards have been made including the implementation of search engine meta tags, the Dublin Core Metadata Initiative (DCMI) and the Resource Description Framework (RDF). More specifically search engine meta tags, popularized by AltaVista, introduced simple metadata elements like "keywords" and "description" within HTML to enhance search relevance and result displays (Gill, 1998). However, due to issues like meta tag spamming, many search engines have deprioritized or ignored meta tags for ranking by focusing on the HTML title tag as a primary factor. Nevertheless, some search engines like Inktomi still employ strategies to detect and penalize tag misuse. The Dublin Core Metadata Initiative, comprising 15 elements among others: contributor, creator and subject, is designed for cross-disciplinary resource discovery and encourages extensibility for complex metadata needs (Gill, 1998). While it enables embedding within HTML, global adoption remains limited due to slow agreement on qualifiers and insufficient support from major search engines. In response to the need for richer, interoperable metadata, the Resource Description Framework (RDF) was developed as an XML-based application that builds on the principles of the Warwick Framework. Moreover, RDF enables complex metadata descriptions by using a structured data model that supports diverse semantic vocabularies through an XML namespace mechanism, allowing for flexible and robust metadata that can describe web resources in a unified manner. These efforts collectively showcase the evolving initiatives aimed at creating practical and adaptable metadata frameworks enhancing efficiency, interoperability and automation in the present and future web environments (Gill, 1998).

Despite the above, sadly the issue of interoperability still extends even beyond library systems. Metadata standards, or the lack thereof, create challenges in managing digital content across a broader network of institutions (Arms, 2000). Standardized metadata not only facilitates data sharing but also ensures that users can efficiently retrieve information from various repositories, regardless of the system they are interacting with (Arms, 2000). Therefore, metadata interoperability is not solely a technical concern but also an operational one. As Day (2005) points out, different institutions prioritize different elements within metadata schemas, leading to discrepancies in how data is structured and retrieved. Libraries may focus on descriptive metadata, while research institutions might emphasize technical metadata (Day, 2005).

The 2003 Dublin Core Conference focused on how the coveted adaptable frameworks support communities of practice and integrate metadata effectively in diverse digital environments (Dublin Core Conference, 2003). In parallel, the IFLA Working Group on Metadata Schemas noted the need for detailed and structured metadata guidelines and especially for complex digital collections. Their guidance suggests that metadata records should capture essential information such as provenance, technical details and rights management, all of which are crucial for the long-term preservation of digital resources (IFLA, 2003). The described approach contrasts with the broader and more general framework of Dublin Core, offering a more in-depth and technical perspective for specialized digital collections.

Brand et. al. (2003) add that a major challenge in maintaining interoperability lies in balancing the specific needs of different domains such as healthcare, geospatial data and education with the broader, more generalizable frameworks like the Dublin Core. These frameworks advocate for modular metadata standards, which allow individual elements to be customized while maintaining a core set of consistent attributes (Brand et. al., 2003). Sysoyev et. al. (2023) emphasize that by integrating cloud computing and distributed ledger systems such as blockchain with metadata frameworks introduces new opportunities and interesting challenges. On one hand, cloud platforms allow metadata to be more easily shared and updated in real-time across decentralized systems and on the other hand, they ensure that metadata remains secure and traceable. As a result, in fluid environments like this, new standards are required that address and adapt to these constant technological shifts (Sysoiev et. al., 2023).

3. Metadata: Data Integrity and Compliance

Indeed, metadata aids in organizing and retrieving information as well as maintaining and ensuring data integrity following the law and the regulatory standards.

The Federal Geographic Data Committee (2002) emphasizes that metadata in geospatial data frameworks helps maintain the accuracy and authenticity of data by documenting the lineage and the transformations applied to datasets, ensuring that each update or modification is recorded systematically (FGDC, 2002). Similarly, Gilliland, (1998) describes that metadata in digital libraries and archives can track changes and updates in digital content, providing a verifiable record of modifications and even of access rights. Metadata in digital archives should include detailed information about ownership, copyright management and usage restrictions allowing organizations to maintain a clear and legally compliant record of right holders of their digital objects as well as the conditions under which they can be accessed or reused (Gilliland, 1998). According to Caplan (2001), metadata frameworks should be designed in a way to align with the legal requirements of the different jurisdictions, ensuring always that data management practices adhere to national and international standards. For instance, metadata records can document compliance with the General Data Protection Regulation (GDPR) by tracking data consent and usage permissions and as a result offering a verifiable trail of how personal information is handled across digital systems (Caplan, 2001).

Moreover, in the scientific field, Swedlow et. al. (2010) highlight that metadata is essential to maintain continuity of experiments while preserving, significant for the research, details. Despite their common route by ensuring data traceability and accuracy, geospatial and digital library metadata standards share differences focusing on the nature of the data that is being managed. More specifically the geospatial metadata, emphasize on geographical accuracy and environmental impact assessments, as indicated by FGDC (2002). On the contrary, metadata in digital archives usually tend to prioritize the preservation of intellectual property and legal rights over the digital objects (Gilliland, 1998). Finally, Sokolova et. al. (2021) refer to the value and utility of metadata in the field of cybersecurity. By applying security protocols within metadata records, institutions ensure that access permissions, encryption keys as well as data management practices are logged in detail. As a result, data security is maintained and simultaneously transparency is provided (Sokolova et. al., 2021).

4. The Practical Application of Metadata

According to Gill (1998), metadata is ubiquitous and has a variety of practical applications. From simple web catalogs to complex databases that manage vast collections metadata is crucial for searching data and information, helping users locate relevant information in the everyday growing overwhelming sea of digital content. Tools like thesauri and query expansion software are used to manage synonyms, translations as well as cross-classification mappings. In addition, schema registries are essential for defining namespaces, ensuring clear and unambiguous term usage while administrative metadata play major role in managing the maintenance of the currency as well as the rights associated with digital artifacts (Gill, 1998), (Arms 2002).

Metadata in general provides important contextual information like the date of creation, the format and necessary technical details that ensure future retrieval and easy use of the digital objects (Gill, 2000). Without this, digital materials risk becoming obsolete as formats and systems change and evolve rapidly in the modern era (Caplan, 2001). As a result, metadata systems must be scalable to accommodate the growing collections of digital data (Day, 2005) and metadata frameworks must be designed in a flexible way and be able to evolve alongside digital collections, ensuring that information remains accessible as the volume and variety of data expand (Day, 2005). An additional and inextricably linked function of metadata with the aforementioned is enabling interoperability between different systems and particularly when organizations or institutions need to share data between them. Metadata in this case act as a bridge, ensuring that data can be shared seamlessly across platforms even with different technical frameworks, enhancing collaboration between the two ends (Caplan, 2001), (Brand et. al., 2003). Metadata is also used in managing the legal and rights-related aspects of digital information. More specifically, administrative metadata captures information about ownership, usage rights and licensing agreements, ensuring that institutions comply with legal standards when sharing or reusing digital content. Important fact also for cultural heritage institutions and for digital libraries that handle intellectual property rights (Gill, 1998), (Baca, 2008). By including user-centric elements like abstracts, summaries and tags, digital libraries enable users to make more informed decisions about which resources are most relevant to their needs Baca (2008). Arms (2002) further supports the above by demonstrating how well-structured metadata can streamline search processes, making it easier for users to find and access relevant resources quickly. Naquin et. al. (2023) add that metadata can also be used in streaming services by categorizing the content and tracking user preferences enabling platforms to offer personalized recommendations.

It is important that metadata always retain a balance between simplicity and functionality. Baca (2008) argues that while metadata systems should be detailed enough to ensure effective organization and retrieval, they should not become that complex that they become burdensome to be maintained. This is especially true for smaller institutions with limited technical expertise. Baca later highlights the importance of designing metadata systems that can be applied across various platforms and data types without overwhelming the users tasked with implementing and maintaining them. (Baca, 2008), (Caplan, 2001).

In scientific research and more specifically in biological sciences, where imaging and large datasets are involved, the quality and comprehensiveness of metadata determine whether data can be effectively shared and reused. Specific metadata containing details about the photography process and processed is important for other researchers who want to replicate experiments or build on and work with existing data. Without robust metadata systems, these valuable scientific data can be lost and hinder the scientific progress (Swedlow et. al., 2010).

Additionally, metadata except from cataloging resources could also include educational elements, learning objectives and audience levels, to enhance the relevance of materials for educational and academic uses (IFLA, 2003). At the 2003 Dublin Core Metadata Initiative conference, experts discussed that metadata could be adapted to incorporate user-generated content and community feedback, allowing educational and research institutions to create more dynamic and interactive repositories (Dublin Core Conference, 2003). Caplan (2001) also notes that these international bibliographic control initiatives have led to the establishment of tailored metadata guidelines for educational environments. Creating standards that accommodate the various pedagogical needs of institutions worldwide, ensures that digital collections are accessible, relevant and with future lifespan across diverse educational settings (Caplan, 20001).

Metadata is already increasingly being used to also train AI systems. In AI-driven applications like natural language processing or image recognition, high-quality metadata ensures that AI systems can interpret and categorize data accurately providing necessary structured data for the machine learning algorithms (Baca, 2008), (Brand et. al., 2007). Last but not least, metadata could also be embedded into network traffic to record details such as source, destination and timestamps for each data packet. Analyzing similar metadata, cybersecurity systems could identify patterns that deviate from the norm, such as unusual access times or unexpected locations, pin pointing security breaches and malicious activities (Sokolova et. al., 2021).

5. Metadata for Digital Libraries and Archives

Libraries have historically facilitated various forms of interaction with knowledge, from supporting academic research, to fact-finding and to enhancing personal exploration (Arms, 2000). In the everyday growing digital realm, this interaction has expanded, allowing users to discover resources like images, sound recordings, documents, videos and all kinds of interactive media enhancing personal and genealogical experiences. Unlike the broad array of web resources, digital libraries are structured collections managed with the aim of long-term and sustainable accessibility, distinguished by their use of formal metadata to organize resources effectively (Arms, 2000).

Access in digital libraries is improving through advanced tools for generating, transforming and sharing metadata, alongside search and retrieval functionalities. XML has become a key syntax for metadata exchange, especially in e-commerce and digital libraries, supporting standards like the Bath Profile for search syntax, MPEG-7 for multimedia and Open Archives Initiative for metadata harvesting (Arms, 2000). It enables flexible use and transformation of schemas, with XSL/T facilitating conversions from EAD to HTML for web displays. Despite that, the full adoption of XML's and the development of RDF tools for scalable interoperability still remains a work in progress. Text-indexing engines, relational databases and SGML-based systems, previously offering distinct functionalities, now have integrating capabilities. For example, Oracle's CONTEXT module supports full-text search within databases and some indexing engines even manage XML content directly (Arms, 2000). Other emerging automated tools also integrate thesauri and knowledge bases, as we already referred, enhancing retrieval by compensating for the incomplete metadata.

Speaking for digital libraries and archives, Gill (1998) points out the importance of structured metadata for enabling resource discovery and retrieving efficient information. He argues that metadata should be designed to capture the essential attributes of digital objects while always maintaining their easy-to-use feature for all the types of operators, humans and automated systems. As a result, search engines make metadata an indispensable tool for libraries to manage vast digital collections and ensure the rediscussed longevity of their archives (Gill, 2000). Arms (2000) suggests that metadata should not only describe digital objects but also include contextual and descriptive elements like user engagement and interaction data. This approach would enhance resource discovery by tailoring search results based on user preferences behaviors and cookies (Arms, 2000). With this structured information about the content, users could locate relevant materials even in vast repositories, more efficiently, a fact that is especially important in institutions where digital collections span multiple disciplines and content types (Arms, 2000).

Brand et. al. (2003) point out that while there are many existing metadata schemas, like the Dublin Core, MARC and MODS, these standards are often applied inconsistently across the various institutions. This lack of uniformity complicates data sharing and integration efforts between libraries and more particularly when collections are managed using different systems (Brand et. al., 2003). For libraries with different content types and organizational structures, metadata provides a common language, allowing digital objects to be understood and accessed by users worldwide extending beyond the boundaries of individual institutions (Baca, 2008).

The role of metadata becomes even more critical as digital archives increasingly adopt cloud-based storage and services which ensure that digital objects are stored securely and are retrievable across the different platforms and environments. Cloud-based metadata systems must be flexible enough to accommodate a wide range of data types and formats while remain consistent and accurate to facilitate long-term preservation (Day, 2005). Integrating AI-powered metadata management systems into digital libraries is the future of digital libraries and can enhance their efficiency ensuring that digital resources remain comprehensive and accessible in time. These systems automate metadata generation, filling gaps and updating, defining collections that are continuously aligned with the current metadata standards without relying on manual input (Sysoiev et. al., 2023).

6. Metadata and Technological Advances

Gill (2000) states that in today's era of big data, metadata acquires an even more dominant role in the information management field. He suggests that metadata should constantly evolve to accommodate with the upcoming innovative technologies, which benefit from the structured descriptions by processing information more efficiently, like the artificial intelligence and the blockchain. In other words, metadata should not be static but should continuously adapt to the rapid changing technological landscape (Gill, 2000).

Baca (2008) adds that metadata is part of the technological advances since it is increasingly being used to train AI in natural language processing and image recognition highlighting that without high-quality metadata, AI systems would struggle to interpret, categorize and make sense of unstructured or semi-structured data. The described symbiotic relationship between metadata and AI enhances the performance of AI that generates more complex and refined metadata automatically (Baca, 2008). Naquin et. al. (2023) also notes that AI-driven metadata frameworks are important and are increasingly being integrated into multimedia content management systems, like the streaming platforms. These systems utilize metadata that categorizes content and also personalizes user experiences by adapting metadata tags based on viewers behavior and their unique preferences (Naquin et. al., 2023). Automations like the ones described above significantly reduce the burden institutions and organizations have by managing large digital collections, allowing them to focus on higher-level tasks like data curation and preservation maintaining human oversight and ensuring accuracy and consistency (Brand et. al., 2007).

As previously stated, metadata is not only used to describe the contents of the images but also the context in which they were created referring instruments used and the conditions of the experiment enabling AI to perform accurate comparisons and analyses (Swedlow, et. al. 2010). Cloud environments also demand metadata that is both scalable and dynamic, capable of accommodating a diverse array of data types and formats while ensuring that information is accessible across different systems and interfaces. In this context metadata should be designed to support current and future technologies (Day, 2005).

Moreover, Sysoiev et. al. (2023) discuss the increasing reliance on blockchain technology for managing securely metadata within the decentralized systems. Blockchain provides a secure and tamper-proof way of storing metadata records, always providing authenticity and traceability across the platforms. By embedding metadata into blockchain frameworks, institutions can establish a verifiable chain of custody for their digital objects, enhancing the trustworthiness of their metadata systems, in legal and compliance contexts (Sysoiev et. al., 2023). As AI, blockchain and other advanced technologies continue to develop, the role of metadata will expand and metadata will not only facilitate data discovery and organization but will also enable more sophisticated forms of data analysis, integration and prediction (Caplan, 2001), (Sysoiev et. al., 2023).

7. AI and Metadata: Present Uses

As already mentioned, Artificial Intelligence has increasingly become intertwined with metadata uses like management, enhancing the way digital content is produced, distributed and consumed. This allows categorization, customization and broader accessibility of the content to diverse audiences. For example, AI driven metadata frameworks have been used already in projects like the Tailored Media Project, which enabled broadcasters to adjust content according to user preferences and environments facilitating in the end a more personalized user experience (Weller et. al., 2024). Another practice of AI is imminent in its use to automate the creation of metadata. The traditional manual methods of generating metadata are time-consuming and often prone to inconsistencies, on the contrary AI methods can automate these processes through image, text and audio analysis while generating the appropriate metadata tags. More specifically, using natural language processing (NLP), AI systems can perform sentimental analysis in a document and detect the thematic relevance of multimedia content, making search engines and digital libraries capable of presenting more refined and contextually relevant results (Li et. al., 2022). This ensures that metadata is generated more consistently and at scale, streamlining the described process (Weller et. al., 2024).

AI systems, except from automating management and generation of metadata, also assist in ensuring compliance with the regulatory standards by auditing metadata records in real-time. More specifically machine learning algorithms can identify discrepancies and errors in metadata entries and particularly in contexts requiring strict compliance like healthcare and finance. By ensuring the metadata conforms to specific regulatory requirements (e.g., GDPR or HIPAA), AI tools enhance the reliability and legality of digital content management systems (Sokolova et. al., 2021). Furthermore, AI plays a crucial role in enhancing the quality and relevance of metadata specifically in content-heavy platforms like the streaming services, where users expect personalized experiences. Through these algorithms, metadata can be dynamically updated and adjusted reflecting user interaction and as a result improving content discovery and streaming recommendation systems (Naquin et. al., 2023).

AI and metadata can be used in autonomous systems too, like self-driving cars and smart IoT devices which are known to rely on metadata for real-time decision-making. These systems collect vast amounts of sensory data and AI can process this data to generate metadata that describe the environment, the location, the speed or even present obstacles. This type of metadata is critical in enabling the autonomous systems to make instantaneous decisions and adjustments, ensuring always the safety and the efficiency in complex dynamic environments (Sysoiev et. al., 2023).

Another advancement is AI's role is data integrity and scalability. AI algorithms help identify gaps in metadata, ensuring that digital content can be accessible over time. AI-based metadata management systems can also detect outdated or even incomplete metadata and prompt updates for these cases (Weller et. al., 2024), (Islam et. al., 2024). For example, enabling real-time monitoring and enabling updating of metadata, ensures that information remains consistent and accessible, even in the cloud computing environments where digital content is continuously updated and moved across multiple servers and platforms. (Naquin et al., 2023). The above approach reduces even more the likelihood of data fragmentation and supports interoperability in decentralized systems.

As AI and metadata continue to evolve, their integration creates opportunities beyond content management, extending into emerging fields like cybersecurity. This transition highlights the versatility of metadata systems, which not only support scalable data ecosystems but also play a crucial role in maintaining data security and privacy across diverse digital environments.

8. Metadata in Cybersecurity

As mentioned, metadata can be used for the security of digital environments and in the management of cyber threats. As recorded below, researches have shown that metadata can also be applied in cybersecurity frameworks to strengthen monitoring and protection measures.

Sokolova et. al. (2021) explain that metadata are a practical and necessary tool for monitoring actions, users and non-users, within systems allowing digital defense systems to detect unusual patterns or security breaches. More specifically by implementing metadata that logs access events and system modifications, organizations can build a proactive defense mechanism able to identify, early on, potential threats (Sokolova et. al., 2021). For example, metadata extracted from device activity, software versions and network connections can be analyzed to identify anomalies which could indicate malware or unauthorized presence in the digital systems (Naquin et. al., 2023).

In addition, Sysoiev et. al. (2023) explore the role of metadata in safeguarding personal privacy in today's vast digital repositories. They note that metadata can be structured to include details like user consent, data access permissions and privacy compliance with regulations like GDPR, helping ensure that sensitive information is handled according to legal standards. Metadata acts in this case as a record of how each user's personal data is managed (Sysoiev et. al., 2023).

Furthermore, Li et. al. (2022) discuss metadata's role in establishing secure communication channels within distributed systems by demonstrating that metadata can encrypt data exchanges and verify the authenticity of transactions. This is particularly crucial in decentralized environments where information that moves across nodes, cloud and IoT infrastructures, needs to be securely transferred, maintaining data integrity and traceability (Li et. al., 2022). Moreover, incident response teams can quickly identify the root cause and scope of attacks, improving their response time and effectiveness, by tagging logs with metadata related to time stamps, IP addresses and user actions (Li et. al., 2022).

Comparing the aforementioned approaches, it is clear that some cybersecurity applications of metadata focus on ensuring compliance and privacy, while others prioritize the proactive monitoring of communications. Together, these strategies showcase metadata's adaptability and capability as a cybersecurity tool, highlighting also the need for flexible and robust metadata frameworks.

9. Conclusion and Future Work

In closing, metadata is a vital part of digital systems as it improves the accuracy of searches, supports cross-collection searches, is preserved over time, updates the content of digital objects, records changes in data, tracks rights and manages versions and documents Gilliland (1998). As emphasized by Gill (1998) and Caplan (2001), standardized and interoperable metadata frameworks remain foundational in ensuring that digital assets are accessible and shared effectively across institutions. These frameworks prevent the fragmentation of digital collections, supporting interoperability across diverse systems (Brand et al., 2003). As we already mentioned in the scientific field Swedlow et. al. (2010) emphasize the role of metadata in reusability of complex datasets and experiments, while Baca (2008) and Day (2005) address the importance of scalability and adaptability, as the digital collections grow.

While technological integration deepens, AI and machine learning have also introduced new dimensions to the use of metadata by automating processes that previously were manual. Baca (2008) and Swedlow et. al. (2010) highlight that while AI supports the efficiency and scalability of metadata systems, it requires a robust structure to interpret the increasing volumes of unstructured data. Weller et. al. (2024) and Islam et. al. (2024) illustrate the potential for AI-driven solutions to fill metadata gaps, ensuring the quality and relevance of digital content, in multimedia and large-scale digital ecosystems. However, these advancements underscore the need for human oversight that guarantees accuracy and ethics to the application of the AI-generated metadata (Sokolova et al., 2021). In addition, it was noticed that metadata also affects the security of these systems. Specifically, metadata structures can monitor access events, manage encryption protocols and log user permissions contributing to robust cybersecurity defenses (Li et al., 2022). Moreover, integrating blockchain technology with metadata frameworks, as highlighted by Sysoiev et. al. (2023) offers promising solutions for enhancing data traceability and authenticity, critical factors in cybersecurity and modern compliance efforts.

Looking forward, metadata research should focus on optimizing AI and using blockchain to enhance the accuracy, traceability and security of digital assets. The challenges of standardization and interoperability which ensure that metadata systems can seamlessly adapt to evolving technological landscapes should also be accessed and analyzed. Exploration into the use of automated metadata generation tools and their role in large-scale digital environments will be essential to balance efficiency and accuracy while maintaining user privacy and compliance with international regulations.

Concluding, metadata remains crucial for ensuring accessibility, law and regulations conformity, preservation, cybersecurity and discoverability of digital content, especially as data volumes and technology grow every day.

References

[1] Gill, Tony. "Metadata and the World Wide Web." In Introduction to Metadata: Pathways to Digital Information. Version 2.0. Edited by Murtha Baca. (J. Paul Getty Trust, 1998). https://web.archive.org/web/20030622012421/http://www.getty.edu:80/research/institute/standards/intrometadata/2_articles/gill/index.html

[2] Gilliland-Swetland, Anne J. "Setting the Stage." In Introduction to Metadata: Pathways to Digital Information. Version 2.0. Edited by Murtha Baca. (J. Paul Getty Trust, 1998). https://web.archive.org/web/20030415012713/http:/www.getty.edu:80/research/institute/standards/intrometadata/2_articles/index.html

[3] Swedlow, J. R., et al. (2010). Metadata matters: Access to image data in the real world. *Journal of Cell Biology*, 189(5), 777–782. https://rupress.org/jcb/article/189/5/777/35828/Metadata-matters-access-to-image-data-in-the-real

[4] Caplan, P. (2001). Metadata for digital libraries: Theoretical and practical considerations. *Library of Congress*. http://lcweb.loc.gov/catdir/bibcontrol/caplan_paper.html

[5] Arms, Caroline R. "Some Observations on Metadata and Digital Libraries". In: Conference on Bibliographic Control for the New Millennium, September 2000. https://www.loc.gov/catdir/bibcontrol/arms_paper.html

[6] Baca, M. (2008). Understanding metadata: What is metadata? *Getty Research Institute*. https://d1wqtxts1xzle7.cloudfront.net/82117494/Understanding_20Metadata-libre.pdf

[7] Amy Brand, Frank Daly, Barbara Meyers Metadata Demystified: A guide for publishers. NISO Press, July 2003. https://web.archive.org/web/20080205132020/http:/www.niso.org/standards/resources/Metadata_Demystified.pdf

[8] Day, M. (2005). Metadata principles and practicalities. https://d1wqtxts1xzle7.cloudfront.net/7158886/10.1.1.108.8711-libre.pdf

[9] 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice-Metadata Research and Application, 28 Sept. – 2 Oct. 2003, Washington USA. https://web.archive.org/web/20060831045317/http:/www.ischool.washington.edu/dc2003/index.html

[10] Dublin Core Metadata Initiative (DCMI). http://dublincore.org/

[11] International Federation of Library Associations (IFLA). (2003). Guidance on the Structure, Content, and Application of Metadata Records for Digital Resources and Collections. https://web.archive.org/web/20071216121108/http://www.ifla.org/VII/s13/guide/metaguide03.pdf

[12] Sysoiev, S., Dudnik, O., & Kotsiuba, O. (2023). Metadata Usage for Personal Data Privacy in Digital Repositories. *Information Technology and Security Journal*, 11(2), 45-56. https://its.iszzi.kpi.ua/article/view/222559

[13] International Federation of Library Associations (IFLA). (2003). Guidance on the Structure, Content, and Application of Metadata Records for Digital Resources and Collections. https://web.archive.org/web/20071216121108/http://www.ifla.org/VII/s13/guide/metaguide03.pdf

[14] Federal Geographic Data Committee (FGDC). (2002). Metadata Standards for Geospatial Data. Federal Geographic Data Committee. https://www.fgdc.gov/publications/documents/metadata/metabroc.html

[15] Sokolova, Y., Pavlenko, O., & Shevchenko, A. (2021). Using Metadata for Threat Detection in Cybersecurity Systems. *IEEE Transactions on Dependable and Secure Computing*, 1-10. https://ieeexplore.ieee.org/abstract/document/9347414

[16] Naquin, E., & Duncan, L. (2023). Metadata for Collections as Data in a Multi-Institutional Digital Library. *Journal of Library Metadata*, *23*(1-2), 43–58. https://doi.org/10.1080/19386389.2023.2229229

[17] Weller, A., Bleisteiner, W., Hufnagel, C., & Iber, M. (2024). The Future is Meta: Metadata, Formats, and Perspectives Towards Interactive and Personalized AV Content. arXiv:2407.19590. https://ar5iv.org/abs/2407.19590

[18] Li, X., Wang, Y., & Zhang, L. (2022). Metadata Framework for Secure Communication in Distributed Systems. *Proceedings of the ACM on Secure and Privacy Systems*, 1-9. https://doi.org/10.1145/3546096.3546104

[19] Islam, S., Wishart, G. C., Walls, J., Hall, P., Seco de Herrera, A. G., Gan, J. Q., & Raza, H. (2024). Leveraging AI and patient metadata to develop a novel risk score for skin cancer detection. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-71244-2