Graph Databases and Crawler Infrastructures as a Common Workspace for the Unified Management of Multi-Source Archaeological Data in the Service of Combating Illicit Trafficking of Cultural Goods

Haris Zacharatos[1][000−0002−0164−5607], Neofytos Vlotomas[1], Charalampos Georgiadis[2][0000−0001−5742−8049], Petros Patias[2][0000−0002−2309−085X], and Themistocles Roustanis[3]

[1] CELLOCK LTD, Nicosia, Cyprus
[2] The Aristotle University of Thessaloniki, Thessaloniki, Greece
[3] KIKLO IKE, Thessaloniki, Greece
haris@cellock.com

Abstract. The illicit trafficking of cultural artifacts poses a significant challenge to global heritage preservation, with artifacts frequently being stolen, traded, and sold through unregulated markets. Managing and analyzing multi-source archaeological data is essential for tracking and preventing such activities, yet the fragmentation and heterogeneity of available data sources hinder effective monitoring. This paper presents ENIGMA, a unified framework that integrates graph databases and web crawler infrastructures to enhance the detection and investigation of illicit transactions involving cultural goods.

The proposed system leverages a graph database schema to model relationships between artifacts, ownership histories, transactions, and locations, facilitating provenance tracking and anomaly detection. Complementing this, AI-powered web crawlers extract and structure data from auction websites, institutional archives, and online marketplaces. Techniques such as Natural Language Processing (NLP) and Named Entity Recognition (NER) are used to classify and tag extracted information, improving data consistency and reliability.

By providing a scalable and interoperable digital infrastructure, ENIGMA enhances collaboration among law enforcement agencies, cultural institutions, and policymakers. This work highlights the potential of combining graph-based analytics and automated web monitoring to combat the illicit trade of cultural goods and outlines future research directions in AI-driven artifact detection, cross-border legal harmonization, and real-time monitoring frameworks.

Keywords: Graph Databases · Web Crawling · Provenance Tracking.

1 Introduction

The illicit trafficking of cultural goods is a major global issue, significantly threatening the preservation of historical and archaeological artifacts. The black-market trade of looted cultural heritage is estimated to be worth billions of dollars annually, with organizations such as UNESCO [1]and INTERPOL reporting that it is one of the most lucrative forms of international crime, alongside drug and arms trafficking. The digitization of the art and antiquities market, coupled with the rise of online auctions and social media platforms, has further complicated efforts to track stolen artifacts, enabling anonymous transactions and rapid resale of looted objects. These activities not only result in irreversible cultural loss but also serve as potential funding sources for organized crime and terrorist networks.

A key challenge in combating the illicit trade of cultural goods is the fragmentation and heterogeneity of available data sources. Information related to artifact provenance, ownership history, and trade networks is often dispersed across museum databases, auction house records, government registries, and online marketplaces, making it difficult for law enforcement and heritage professionals to identify and track illicit transactions. Traditional database systems struggle to integrate, process, and analyze this vast and complex data landscape, highlighting the need for advanced digital solutions.

This paper introduces ENIGMA, a unified framework that integrates graph databases and AI-powered web crawlers to enhance the detection and investigation of illicit transactions involving cultural goods. By employing graph-based modeling, the system effectively represents artifacts, ownership histories, transactions, and locations, allowing for efficient provenance tracking and anomaly detection. Additionally, automated web crawlers extract and structure data from auction websites, institutional archives, and online marketplaces, using Natural Language Processing (NLP) and Named Entity Recognition (NER) [2] techniques to improve data classification and tagging.

The contributions of this study include:

- A novel graph-based database architecture that models the provenance and trade networks of cultural artifacts, enabling real-time data integration from multiple sources.
- An AI-powered web crawling infrastructure that automates the monitoring of online auction sites, darknet marketplaces, and social media platforms for suspicious transactions.
- A scalable and interoperable digital workspace that facilitates collaboration between law enforcement agencies, cultural institutions, and policymakers to combat illicit trade more effectively.

2 Background and Related Work

The use of advanced data management and analysis technologies has become critical in combating the illicit trafficking of cultural heritage artifacts. In particular, graph databases [3], web crawlers [4], and multi-source data integration

techniques have shown significant potential in enhancing provenance tracking, fraud detection, and knowledge discovery. These technologies, when combined, create a comprehensive digital infrastructure that enables real-time monitoring, structured data organization, and actionable insights for stakeholders such as law enforcement agencies, cultural institutions, and policymakers.

2.1 Graph Databases in Cultural Heritage Protection

Graph databases have gained widespread adoption across various domains due to their ability to model and analyze complex relationships efficiently [3]. In the cultural heritage sector, they facilitate the representation of interconnected entities, such as artifacts, historical records, legal ownership documents, and previous transactions. Unlike traditional relational databases, which rely on rigid schemas and tabular data structures, graph databases excel in handling highly interconnected datasets, making them particularly suitable for modeling provenance relationships, tracking ownership histories, and uncovering illicit trade patterns [5]

For example, projects like the ARIADNE Knowledge Graph [6] and the Louvre Museum's Provenance Research Database [7] leverage graph-based models to visualize relationships between artifacts, collectors, and trade networks. These models facilitate advanced querying and inference mechanisms, enabling investigators to detect suspicious artifacts linked to known trafficking routes. Additionally, graph-based AI techniques, such as Graph Neural Networks (GNNs), are being explored to predict hidden relationships between cultural assets and illicit actors [8].

2.2 Web Crawlers for Data Collection

Automated web crawlers play a fundamental role in monitoring online platforms for illicit transactions by systematically extracting data from diverse digital sources. These include auction websites, social media, dark web marketplaces, and institutional repositories. By continuously updating a centralized database, web crawlers ensure that newly listed cultural items, suspicious listings, and ownership claims are constantly tracked, reducing the risk of fraudulent sales and unauthorized exports.

Modern web crawlers are enhanced with Natural Language Processing (NLP) and image recognition capabilities, allowing them to classify artifacts, detect counterfeit descriptions, and match online listings with stolen items. For example, the Interpol Stolen Works of Art Database integrates AI-powered web crawlers to scan auction sites and identify artifacts that resemble those in its global registry of stolen cultural assets [9].

However, challenges persist in web crawling within restricted or non-indexed environments, such as private sales platforms and encrypted networks. Advanced methods, including semantic data extraction, entity resolution, and deep learning-based fraud detection, are being explored to enhance the efficiency of web-based monitoring systems.

2.3 Challenges in Multi-Source Data Management

The integration of multi-source data from auction houses, government registries, academic research, and online trade platforms introduces several challenges, including:

- Semantic inconsistencies: Different sources use varied terminologies, classifications, and metadata structures, making it difficult to standardize datasets.
- Data redundancy: Repeated information across multiple sources can lead to duplication, affecting data accuracy and storage efficiency.
- Reliability concerns: Verifying the authenticity of data from user-generated content, informal trade networks, and unverifiable sources is a persistent issue.

To address these challenges, researchers are developing common data models and ontologies that align different datasets under standardized schemas. Initiatives such as the CIDOC Conceptual Reference Model (CIDOC-CRM) and the Linked Art Data Model provide a structured framework for harmonizing metadata across museums, law enforcement databases, and online archives. Furthermore, blockchain-based provenance tracking is emerging as a solution to enhance data integrity, transparency, and immutability in cultural heritage transactions.

The integration of graph databases, AI-powered web crawlers, and robust multi-source data management strategies is revolutionizing cultural heritage protection efforts. Future research should focus on scalable machine learning algorithms for artifact detection, cross-border legal harmonization for digital evidence collection, and real-time monitoring frameworks that proactively detect and prevent illicit transactions.

3 Methodology

The proposed framework employs a graph database schema to systematically represent key entities, including artifacts, owners, transactions, and locations. By modeling the relationships between these entities, the system effectively traces provenance, historical exchanges, and potential illicit trafficking risks.

To populate the database, web crawlers continuously extract data from predefined sources, structuring and filtering information before integration. Advanced techniques such as Natural Language Processing (NLP) and Named Entity Recognition (NER) enhance data classification, enabling accurate tagging and improved contextual understanding of extracted information.

4 Implementation

The Web Crawler is a sophisticated tool designed to extract and process metadata from web pages, aiding in the fight against illicit trafficking of cultural goods. Through careful design, strategic crawling approaches, and rigorous testing, the Web Crawler ensures efficient, ethical, and secure data extraction, contributing to the broader objectives of the ENIGMA platform.

4.1 Design and Architecture

The Web Crawler is composed of several integral components, each with specific roles in ensuring effective and efficient web crawling. These components work together to extract, process, and store relevant data from target websites.

- Frontier: Manages the list of URLs to be crawled, ensuring that URLs are processed efficiently and effectively.
- Downloader: Fetches web pages and other resources from the internet.
- Parser: Analyzes the downloaded content to extract useful information.
- Scheduler: Coordinates the crawling process, managing tasks and resources to optimize performance.

4.2 Crawling Strategies

Effective crawling strategies are essential for optimizing the performance and efficiency of web crawlers. The approach taken depends on the objectives of data collection, the nature of the target websites, and computational constraints. The ENIGMA web crawler integrates multiple strategies for efficient data extraction:

- Hybrid Depth-First and Breadth-First Search (DFS and BFS): To optimize data retrieval, the web crawler integrates both DFS and BFS strategies. DFS is used when deeper exploration of individual sites is needed, such as tracing detailed object provenance records. BFS, on the other hand, ensures a broad collection of related entries from different sources, helping gather a wide spectrum of data from various repositories.
- Focused Crawling: The crawler prioritizes relevant web pages based on predefined topic models, keywords, and heuristics. By filtering out irrelevant content early, focused crawling ensures that only the most pertinent cultural heritage-related data is processed, reducing storage costs and computational overhead.
- Adaptive Crawling: Given the dynamic nature of online marketplaces and archives, adaptive crawling techniques are implemented. This involves periodic re-crawling of previously indexed pages to capture updates, such as newly listed artifacts, altered provenance records, or flagged stolen goods.
- Machine Learning-Assisted Crawling: AI models analyze metadata patterns from previously crawled pages to predict high-value targets for future crawls. This learning-based approach refines search paths over time, improving efficiency by focusing on sources that historically provided valuable data.
- Dynamic Page Handling: Since many modern websites generate content dynamically using JavaScript and AJAX, the crawler includes capabilities for rendering web pages. This ensures that it captures dynamically loaded images, descriptions, and metadata that might otherwise be inaccessible through traditional crawlers.

4.3 Scalability and Performance Optimization

As the volume of cultural heritage data continues to expand, ensuring the scalability and efficiency of the ENIGMA web crawler is essential for maintaining high-performance large-scale data extraction. To achieve this, the crawler employs a distributed architecture, allowing multiple crawling nodes to operate in parallel, significantly improving processing speed while enabling the simultaneous querying of multiple data sources. Load balancing mechanisms further optimize resource utilization by distributing tasks efficiently across nodes, preventing bottlenecks. Additionally, an intelligent task scheduling system dynamically prioritizes crawling tasks based on source reliability, data freshness, and computational cost, ensuring that high-value targets are processed first while minimizing redundant requests.

To facilitate efficient data retrieval, extracted metadata is stored in the ENIGMA graph database using an optimized indexing strategy, incorporating techniques such as inverted indexes and B-tree indexing to enhance search performance and maintain low latency. Caching mechanisms further improve efficiency by temporarily storing frequently accessed metadata, reducing redundant data requests and minimizing network bandwidth consumption. Instead of performing full website re-crawling, the crawler leverages incremental crawling and change detection algorithms to monitor and prioritize updates, ensuring that only modified or newly listed artifacts are reprocessed. This approach conserves computational resources while keeping the database continuously updated.

To further enhance performance, the ENIGMA crawler integrates asynchronous data processing, where data extraction, parsing, and storage occur concurrently. This parallel processing prevents delays, allowing large volumes of data to be analyzed in real-time without overwhelming system resources. By implementing these scalability and performance optimizations, the ENIGMA web crawler ensures efficient, accurate, and high-speed data extraction, making the platform adaptable to the evolving landscape of cultural heritage protection and illicit trade monitoring.

4.4 Crawling Ethics and Robots.txt

Given the legal and ethical considerations involved in web crawling, the ENIGMA web crawler follows strict compliance guidelines, ethical standards, and security protocols to ensure responsible data extraction and usage. It strictly adheres to the robots.txt policies of each website, ensuring that restricted sections remain inaccessible and that web servers are not overloaded, thereby maintaining ethical crawling practices and respecting terms of service agreements. The crawler also upholds intellectual property rights, as it does not store or use copyrighted material beyond metadata extraction, with all collected data being used solely for lawful and academic purposes related to cultural heritage preservation.

To protect data privacy, the system does not collect personal data, such as user credentials, private transactions, or non-public auction information, limiting

retrieval strictly to publicly available metadata essential for provenance tracking and authentication. Security and threat mitigation measures are integrated to detect and prevent the extraction of malicious content, ensuring protection against malware-infected or phishing-related data. The crawler utilizes secure HTTPS connections for data retrieval and follows industry best practices to prevent exposure to harmful sites.

Additionally, the ENIGMA crawler implements rate-limiting techniques to avoid overloading target servers, ensuring fair usage and minimizing its impact on website performance. To ensure legal compliance, it adheres to international data protection regulations, including the General Data Protection Regulation (GDPR), guaranteeing that all retrieved metadata is handled securely and in accordance with legal requirements. Through these comprehensive measures, ENIGMA ensures ethical, secure, and legally compliant web crawling while supporting cultural heritage protection efforts.

4.5 Data Storage: Database Schema and Data Models

The ENIGMA graph database (OrientDB) depict in Figure 1 provides a flexible and scalable schema to store and manage both structured and unstructured metadata, ensuring efficient handling of diverse and evolving datasets. The database is designed to accommodate artifact descriptions, including textual information such as names, historical background, and provenance details, as well as multimedia data, such as images, videos, and 3D models, which are linked to textual metadata for enhanced searchability. Additionally, transactional records, containing details of ownership transfers, auction listings, and legal documents, are integrated to support provenance tracking. Geospatial data is also included, mapping the movement and origin of cultural goods to provide deeper insights into trade routes and potential illicit trafficking patterns.

To optimize data retrieval and analysis, the ENIGMA system employs graph-based relationships, which enable dynamic querying of ownership networks, object history, and illicit trade patterns, providing a more effective alternative to traditional relational databases. The indexing and search optimization mechanisms utilize full-text indexing, geospatial indexing, and hierarchical data storage, ensuring fast and accurate access to relevant information. Additionally, data normalization and deduplication techniques are implemented to automatically detect and merge duplicate records, maintaining data consistency and reducing redundancy. The integration of machine learning-assisted structuring further enhances the system, transforming unstructured metadata—such as descriptions extracted through Natural Language Processing (NLP) algorithms—into structured data points, enriching the database for advanced analytical applications.

Furthermore, ENIGMA ensures data integrity and traceability by incorporating versioning and audit trails, which log all modifications and updates, enabling forensic investigations and ensuring compliance with data governance standards. The system also supports interoperability with external repositories through standardized data exchange formats such as JSON-LD, XML, and RDF, allowing seamless integration with museum databases, law enforcement archives, and

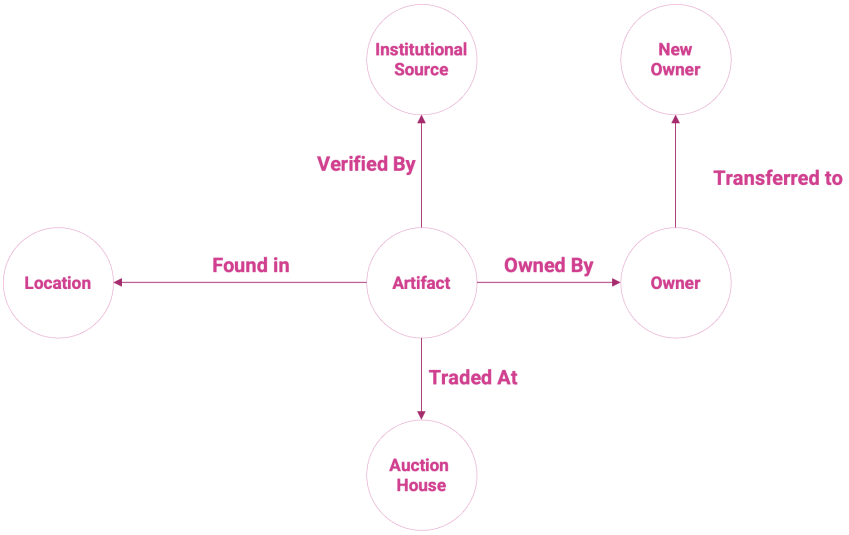

Fig. 1. Enigma sample Graph network diagram

open-access cultural heritage registries. Through these combined capabilities, ENIGMA provides a robust, scalable, and intelligent data management system for combating illicit cultural artifact trafficking.

4.6 Data Extraction Process

The data extraction process is designed to systematically retrieve, clean, and integrate data into the ENIGMA database while maintaining accuracy and efficiency. The extraction workflow consists of multiple stages, each contributing to the integrity and usability of the collected data as depict in figure 2

1. Target Identification and URL Discovery: The web crawler begins by identifying relevant websites and URLs based on predefined criteria such as auction sites, museum databases, and online archives. The crawler dynamically updates its list of sources to ensure continuous coverage of newly emerging data.

2. Web Page Retrieval: The system fetches web pages using HTTP/HTTPS requests, ensuring compatibility with site structures and security protocols. Advanced mechanisms such as JavaScript rendering and AJAX handling are implemented to extract dynamically generated content.

3. Data Parsing and Content Extraction: Extracted web pages are parsed using HTML, JSON, and XML parsing techniques. Metadata such as artifact descriptions, images, timestamps, and provenance records are extracted and structured into machine-readable formats.

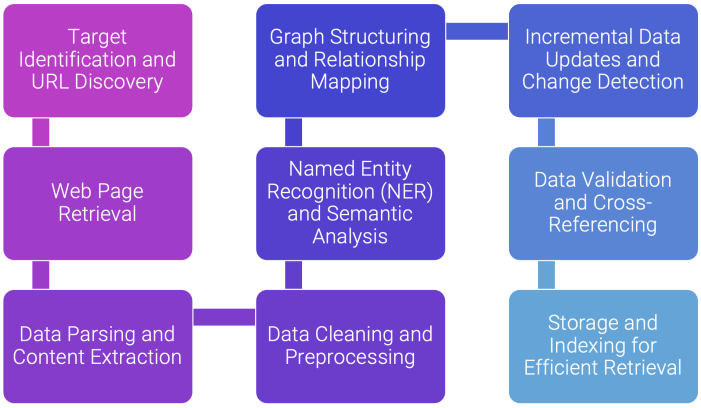

Fig. 2. ENIGMA Web Crawler Workflow Diagram

4. Data Cleaning and Preprocessing: To maintain high-quality datasets, raw extracted data is cleaned by removing duplicate records, normalizing text formats, and correcting inconsistencies. This step also includes image preprocessing, where object recognition and feature extraction techniques are applied to enhance artifact identification.

5. Named Entity Recognition (NER) and Semantic Analysis: Using NLP techniques, key terms such as artifact names, historical events, and geographical locations are identified and mapped to standardized ontologies. This enhances data interoperability and facilitates automated linking between related entries.

6. Graph Structuring and Relationship Mapping: Extracted entities are structured into the graph database, linking artifacts with owners, locations, auction listings, and historical events. This allows for complex queries to trace object provenance and detect suspicious transactions.

7. Incremental Data Updates and Change Detection: The system continuously monitors previously extracted data sources for updates. If an artifact's status changes (e.g., sold in an auction, reported stolen), the database is updated to reflect the latest information.

8. Storage and Indexing for Efficient Retrieval: Once validated, structured data is stored in the graph database, indexed for rapid retrieval, and made available for analysis and visualization through the ENIGMA interface.

5 Data Sharing Framework

The Data Sharing Framework (DSF) is a core component of the ENIGMA platform that enables seamless, secure, and efficient data exchange among various stakeholders, including law enforcement agencies, cultural heritage institutions, and researchers. It ensures data interoperability, integrity, and compliance with regulatory standards.

5.1 Data formats and Standards

The DSF within the ENIGMA platform employs a variety of data formats and standards to ensure seamless data sharing and collaboration among users. Adherence to these standards facilitates interoperability and consistency across the platform. JSON Schema and Validation: Ensures consistency and proper structuring of data using JSON validation rules. XML-RPC and SOAP Protocols: Enable communication with legacy systems and structured data exchange. Europeana Data Model (EDM): Utilized for representing cultural heritage objects to ensure metadata alignment with international archival standards

5.2 Authentication and Authorization

To maintain security and controlled access, the DSF employs:

OAuth and API Keys: To secure access to the DSF, the platform employs OAuth for authorization and API keys for authentication. These mechanisms ensure that only authorized users and applications can access and manipulate the data, providing a secure environment for data sharing.

Role-Based Access Control (RBAC): The DSF implements RBAC, assigning permissions based on user roles. This approach ensures that users have access only to the data and functions necessary for their roles, enhancing security and operational efficiency.

5.3 Secure Data Sharing and Collaboration

The DSF facilitates: Secure API Design: All APIs follow security best practices, including HTTPS encryption and request authentication.

Interoperability with External Databases: The DSF connects with external systems such as Interpol's Stolen Works of Art Database and national heritage registries, allowing real-time data exchange.

Version Control: Ensures changes to datasets are logged and previous versions are retrievable, maintaining historical records.

6 Privacy and Compliance

To ensure adherence to international data protection regulations, the ENIGMA platform implements stringent privacy and security measures. Personally Identifiable Information is anonymized before being shared with external parties, maintaining compliance with General Data Protection Regulation principles to safeguard data privacy and security. To enhance data integrity and accountability, logging and audit trails are maintained, tracking all data interactions to prevent unauthorized modifications and support forensic analysis when needed. The platform employs end-to-end encryption, ensuring that data remains secure both in transit (TLS 1.2/1.3) and at rest (AES-256 encryption). Through these comprehensive measures, the ENIGMA platform ensures robust privacy, compliance, and security, protecting sensitive cultural heritage data while maintaining operational integrity.

7 Challenges and Solutions

7.1 API compatibility Bottlenecks

Integrating various APIs from different AI tools, the Web Crawler, and the DSF presents significant compatibility challenges, primarily due to differences in data formats, communication protocols, and system architectures. These discrepancies can lead to issues such as data mismatches, synchronization failures, and inefficient processing. To ensure seamless interoperability, it is crucial to implement RESTful APIs with standardized data formats like JSON and XML, which provide a widely accepted structure for data exchange and facilitate smooth communication between diverse systems.

Moreover, middleware solutions play a vital role in bridging gaps between varying API specifications and protocols by acting as an intermediary layer that translates and harmonizes data exchanges. These middleware systems can manage protocol conversions, authentication mechanisms, and message routing to enhance compatibility across integrated platforms.

To further mitigate potential bottlenecks, regular API testing and validation should be conducted to identify inconsistencies and resolve them proactively. This involves automated and manual testing techniques, including functional testing, load testing, and security testing, to ensure that all integrated components communicate efficiently and reliably. By adopting these strategies, the system can maintain robust API interoperability, reduce downtime, and enhance the overall performance of data-driven processes.

7.2 Data consistency and Integrity

Ensuring data consistency and integrity is a critical challenge when handling information that flows from extraction by the Web Crawler, through processing by AI tools, and into the DSF (Data Storage Framework). As data transitions across these stages, it is susceptible to corruption, loss, or inconsistency due to transformation errors, integration conflicts, or system failures. These risks can compromise analytical accuracy, lead to incorrect decision-making, and reduce system reliability.

To mitigate these challenges, robust validation mechanisms should be employed, such as JSON Schema and XML validation, which enforce data structure compliance and ensure that incoming data adheres to predefined formats before processing. These validation techniques help detect anomalies early, preventing malformed or incomplete data from propagating through the system.

Additionally, implementing transaction management and concurrency control mechanisms is essential to maintain data consistency, particularly in environments where multiple processes access and modify data simultaneously. These mechanisms ensure that all operations adhere to ACID (Atomicity, Consistency, Isolation, Durability) properties, preventing issues such as race conditions, data duplication, or unintended overwrites.

Furthermore, incorporating audit logs provides a structured way to track modifications, detect anomalies, and enhance traceability. These logs maintain a detailed history of all data changes, enabling administrators to rollback to previous versions if corruption or inconsistencies arise. This approach not only strengthens data governance and security but also enhances system resilience and reliability.

By integrating these solutions, organizations can effectively safeguard the integrity of their data, ensuring that it remains accurate, consistent, and reliable throughout its lifecycle.

7.3 Orchestrating AI processes with WebCrawler and Data Sharing

Orchestrating the interaction between the Web Crawler, AI tools, and the DSF (Data Storage Framework) is a multi-step process that ensures seamless data flow from extraction to final utilization. This process is fundamental to maintaining the efficiency and effectiveness of the ENIGMA platform, enabling structured, enriched, and accessible data for users.

Data Extraction: The process begins with the Web Crawler, which systematically scans targeted websites and extracts metadata from various sources. This raw data, often unstructured, is then stored in the ENIGMA database, capturing a wide array of information relevant to the system's analytical and decision-making processes. Data Processing: Once extracted, the raw metadata undergoes transformation through AI-powered processing tools, leveraging machine learning (ML) and natural language processing (NLP) techniques. These AI tools refine the data by cleaning, enriching, and structuring it into a standardized EDM (Europeana Data Model) format. This transformation is essential to ensuring consistency, accuracy, and usability across different analytical workflows. Data Integration: After processing, the structured data is integrated into the DSF, where it is indexed and made searchable. This step ensures that the transformed metadata is efficiently stored and readily accessible for retrieval, supporting advanced search functionalities, categorization, and interlinking of related information. Data Sharing and Utilization: The final step involves user access and collaboration, where users interact with the DSF to retrieve, analyze, and share data. Through intuitive search and query mechanisms, enriched metadata enables deeper insights, supports data-driven decision-making, and fosters collaborative knowledge-building. This well-orchestrated workflow optimizes the data lifecycle within the ENIGMA platform, enhancing efficiency, accuracy, and accessibility while ensuring that raw, unstructured data evolves into valuable, structured insights for end users. The seamless interaction between these components underscores the robustness of the platform, making it a powerful tool for intelligence gathering, analysis, and collaboration.

8 Joint Workspace Implementation

The Joint Workspace is designed to provide a secure, scalable, and collaborative environment for users to effectively combat the illicit trafficking of cultural goods.

Establishing a robust infrastructure is key to ensuring efficiency, security, and adaptability. One of the critical decisions involves selecting between cloud and on-premises infrastructure. Cloud infrastructure offers flexibility and scalability, allowing dynamic resource allocation and robust disaster recovery without significant upfront investment. However, it comes with ongoing subscription costs and potential security concerns. On the other hand, on-premises infrastructure provides greater control over data security and can be more cost-effective over time, though it requires a higher initial investment and ongoing maintenance. Scalability planning is also essential to support growing data volumes and user demand. Horizontal scaling, which involves adding more system nodes, is highly effective in cloud environments, whereas vertical scaling, increasing resources like CPU and RAM, is more suited for on-premises solutions but has inherent limitations.

Maintaining user access and permissions is critical to securing sensitive data. Implementing Role-Based Access Control (RBAC) ensures users only access necessary data and functions based on their role. Security is further strengthened through multi-factor authentication (MFA) and robust authorization protocols. Additionally, audit trails log all user activities, providing traceability and aiding security audits to detect potential breaches.

Continuous monitoring and logging ensure system reliability and performance. Tools like Prometheus and Grafana are used for performance monitoring, tracking system resources and user activity, with automated alerts notifying administrators of anomalies. Log management solutions, including ELK Stack (Elasticsearch, Logstash, Kibana) and Splunk, facilitate centralized log analysis, helping administrators troubleshoot and resolve issues efficiently.

Effective error handling mechanisms are essential to maintaining a smooth user experience. Exception management prevents system crashes by properly handling errors, while user notifications provide clear messages explaining issues and possible resolutions. Additionally, automated recovery processes allow rapid responses to common failures, such as restarting services or rolling back to stable versions.

Regular maintenance procedures ensure security, stability, and performance optimization. Scheduled maintenance includes updates, patches, and system enhancements. A robust backup and recovery strategy guarantees data integrity, with periodic backups and tested recovery plans in place. Lastly, security audits help identify vulnerabilities, enforce compliance with regulations like GDPR, and apply necessary security updates.

9 Conclusion and future work

9.1 Emerging Technologies

As the ENIGMA platform evolves, the integration of emerging technologies will be essential for enhancing its ability to combat the illicit trafficking of cultural

goods. Future advancements will focus on Artificial Intelligence (AI) and Machine Learning (ML) to refine data extraction, enrichment, and analysis processes. This includes the development of more sophisticated Natural Language Processing (NLP) algorithms to improve text interpretation and entity recognition, as well as the integration of advanced image recognition capabilities for identifying artifacts across multiple sources. By leveraging these innovations, ENIGMA aims to increase automation, improve detection accuracy, and enhance real-time monitoring of cultural goods markets.

9.2 Scalability and performance improvements

To ensure the ENIGMA platform remains scalable and efficient as data volumes and user activity increase, ongoing improvements will focus on horizontal scaling, distributing workloads across multiple servers to accommodate growing demands without compromising performance. Database optimization techniques, such as indexing, query optimization, and in-memory databases, will enhance retrieval speeds and processing efficiency. Additionally, cloud integration will provide a flexible and scalable infrastructure, leveraging on-demand resources to dynamically adjust based on usage patterns. To maintain system reliability, load balancing mechanisms will be implemented to evenly distribute network traffic across multiple servers, preventing bottlenecks and ensuring high availability and seamless performance.

9.3 Feedback mechanisms and continuous Integration

To maintain a user-centric approach and ensure the ENIGMA platform remains responsive to evolving needs, future development will focus on robust feedback mechanisms, including surveys, feedback forms, and direct user engagement, to gather insights and refine system functionality. The adoption of Continuous Integration and Deployment (CI/CD) practices will streamline development, enabling automated testing, integration, and frequent updates, minimizing downtime while keeping the platform up to date. Additionally, fostering community collaboration with cultural heritage organizations, law enforcement agencies, and academic institutions will drive innovation and ensure alignment with industry best practices. To further enhance reliability, real-time performance monitoring and analytics tools will be integrated, enabling proactive issue detection and resolution, ensuring an efficient and seamless user experience.

By focusing on these areas, the ENIGMA platform will continue to advance, incorporating the latest technologies and best practices to enhance its capabilities in preserving and protecting cultural heritage. Scalability and performance improvements will ensure the platform can grow with its user base, while feedback mechanisms and continuous integration will keep it responsive and adaptive to user needs.

10 Acknowledgments

This research was funded by the European Union, under the project ENIGMA "Endorsing, safeguarding, protection, and provenance management of cultural heritage" grant number 101094237. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA. Neither the European Union nor the granting authority can be held responsible for them.

References

1. UNESCO. (2024). Culture and development: stop the illicit traffic of cultural property. https://unesdoc.unesco.org/ark:/48223/pf0000225521
2. Yu-Ming Shang, Hongli Mao, Tian Tian, Heyan Huang, Xian-Ling Mao, From local to global: Leveraging document graph for named entity recognition, Knowledge-Based Systems, Volume 312, 2025, 113017, ISSN 0950-7051.
3. VILLAZÓN-TERRAZAS, Boris, et al. Knowledge graphs and semantic web. Communications in Computer and Information Science, 2020, 1232: 1-225
4. BREWSTER, Ben; INGLE, Timothy; RANKIN, Glynn. Crawling open-source data for indicators of human trafficking. In: 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing. IEEE, 2014. p. 714-719.
5. PICCIANO, Paulina Joy. Laundered: an analysis of antiquities trafficking networks and the illicit market. 2023. Master's Thesis. Rutgers The State University of New Jersey, School of Graduate Studies.
6. BARDI, Alessia, et al. The ARIADNEplus Knowledge Base: a Linked Open Data set for archaeological research. 2024
7. https://collections.louvre.fr/en/page/apropos
8. ABATE, D., et al. Significance. Stop illicit heritage trafficking with artificial intelligence. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2022, 43: 729-736.
9. https://www.interpol.int/Crimes/Cultural-heritage-crime/Stolen-Works-of-Art-Database