POLITECNICO DI TORINO
Repository ISTITUZIONALE

Towards EUROGLOSS: an Interoperable Knowledge Infrastructure for Built Heritage

Original
Towards EUROGLOSS: an Interoperable Knowledge Infrastructure for Built Heritage / Cannavo', A., Lucarini, C., Palla, D., Arslan, V., Munteanu, N., Burgassi, V.. - (2026). (2026 IEEE 5th International Conference on Intelligent Reality (ICIR 2026) Pisa (ITA) June 25-26, 2026) [10.1109/ICIR70631.2026.11628138].

Availability:
This version is available at: 11583/3011533 since: 2026-05-28T17:16:08Z

Publisher:
IEEE

Published
DOI:10.1109/ICIR70631.2026.11628138

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

18 November 2026

Towards EUROGLOSS: an Interoperable Knowledge Infrastructure for Built Heritage*

Alberto Cannavò, *Member, IEEE*, Costanza Lucarini, Dominik Palla,
Volkan Arslan, Nicolae Munteanu, and Valentina Burgassi

***Abstract*—This paper presents the methodological and technical requirements of EUROGLOSS, a digital platform designed to host a multi-layered repository of construction documents related to Modern Age European Royal Residences. By aligning with the FAIR principles, EUROGLOSS aims to establish an interoperable framework where standardized construction terminology is transformed from heterogeneous archival sources into ontological entities, serving as a functional tool for real-time intelligent services. Given that Modern Intelligent Reality (IR) applications require high-fidelity data to bridge the gap between physical architectural assets and their historical evolution, the EUROGLOSS repository can act as a semantic backbone for Knowledge Digital Twins, able to provide practitioners and conservation authorities with actionable historical evidence. To ensure a robust framework, a comparative analysis of representative global digital platforms was conducted, evaluating metadata schemas and system architectures.**

***Index Terms*—digital heritage, systematic analysis, comparative evaluation, knowledge infrastructure, EUROGLOSS**

I. Introduction

The digital transition in Cultural Heritage is moving beyond simple archiving toward the creation of Intelligent Reality (IR) ecosystems [1] In this context, Digital Humanities (DH) are no longer a separate domain but provide the essential semantic backbone for context-aware systems [2]. By integrating high-level humanistic knowledge with AI-driven perception [3], Knowledge Digital Twins of the built environment can be developed so that they are not only accurate but also rich in historical and terminological depth [4].

This vision is operationalized through the EUROGLOSS COST Action (CA24102)[1], an on-going European-wide initiative aimed at overcoming the current fragmentation of technical and lexicographical data related to the built heritage of the Modern Age [5]. Specifically, the project focuses on the construction history of European Court Residencies, a field characterized by a large but dispersed corpus of archival documents, including construction site records, architectural drawings, and administrative accounts [6].

Despite this fragmentation, recent studies have demonstrated that a comparative analysis of different contexts reveals a long-term refinement of local construction techniques and traditions, which evolved in synergy with the expertise of the leading figures of Savoyard architecture [7].

To transform this fragmented landscape into a scalable digital asset, EUROGLOSS is engineering an Open Access Common Repository. This infrastructure is designed as a high-interoperability data layer for historical technical knowledge and its primary output consists of a glossary of technical terms used in the construction processes of 17th- and 18th-century European Court Residences, with a particular focus on documents and drawings produced by artisans and craftsmen. Drawing inspiration from established initiatives such as the glossary of Les Cours de Desgodets[2] and the Getty Research Institute Vocabularies[3], EUROGLOSS digital platform intends to go beyond simple digitization: it seeks to harmonize technical terminology across diverse languages and historical periods. By adopting the FAIR principles (Findability, Accessibility, Interoperability, and Reuse) [8], the project establishes a standardized framework where heterogeneous archival sources become machine-readable. In this perspective, EUROGLOSS proposes a cross-lingual formal ontology mapping, providing semantic foundations to an infrastructure that allows for cross-contextual querying and inference. Moreover, to ensure long-term viability and scalability, EUROGLOSS is designed to adopt an extensible modular approach to ontology development. Rather than relying solely on manual curation, the intended infrastructure aims to leverage AI-driven methods for knowledge discovery and semi-automated mapping. By employing machine learning algorithms to analyze heterogeneous archival corpora, the proposed system is intended to help researchers identify new terminological clusters and map them to existing ontological nodes. This AI-augmented framework is designed to not only accelerate the growth of the repository but also ensure that the architecture remains flexible, allowing for the seamless integration of diverse datasets as the European network expands. Such architectural adaptability is planned to ensure both compliance with international interoperability standards and the precision required to capture the nuanced

* This article is based upon work from COST Action "A Glossary of Technical Construction Vocabulary in 17th-18th Century European Court Residences (EUROGLOSS)", CA24102, supported by COST (European Cooperation in Science and Technology).

A. Cannavò and N. Munteanu are with the Dept. of Control and Computer Engineering, Politecnico di Torino, Turin, Italy (e-mail: alberto.cannavo@polito.it, nicolae.munteanu@studenti.polito.it).

C. Lucarini and V. Burgassi are with the Dept. of Architecture and Design, Politecnico di Torino, Turin, Italy (e-mail: costanza.lucarini@polito.it, valentina.burgassi@polito.it)

D. Palla is with the Dept. of Informatics and Quantitative Methods, University of Hradec Kralove, Hradec Kralove, Czech Republic (email: dominik.palla@uhk.cz)

V. Arslan is with the Dept. of Civil Engineering, Zonguldak Bülent Ecevit University, Zonguldak, Türkiye (email: volkanarslan@beun.edu.tr)

[1]EUROGLOSS COST Action: https://www.cost.eu/actions/CA24102/

[2]Les Cours de Desgodets, http://www.desgodets.net/glossaire

[3]Getty Vocabularieshttps://tinyurl.com/mrdc7tw4

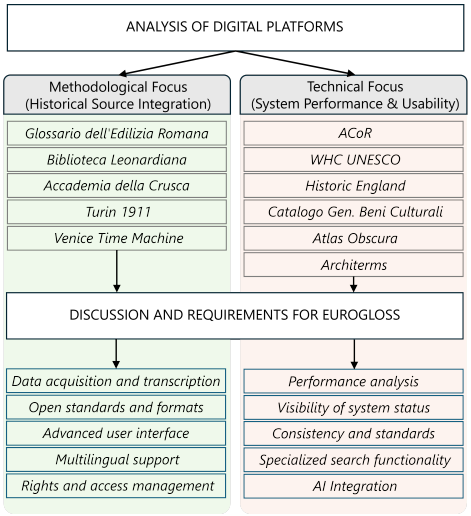


Fig. 1. Systematic analysis of digital platforms for the development of EUROGLOSS requirements.

technical peculiarities of 17th- and 18th-century construction practices.

This paper thus focuses on the methodological and technical foundations of the EUROGLOSS project through a systematic analysis (summarized in Figure 1) of prominent case studies within the field of digital heritage.

By examining European initiatives active in the last ten years, such as the *Glossario dell'Edilizia Romana*, the *e-Leo* and *Accademia della Crusca* digital archives, *Turin 1911*, and the *Venice Time Machine*, it maps an evolutionary trajectory: from early-generation digital lexicography to the current frontiers of "Big Data" for cultural heritage.

Beyond these benchmarks, the research evaluates heterogeneous infrastructures, including the *Atlas des techniques de la Construction Romaine*, *Historic England*, and the *UNESCO World Heritage Convention* database, to assess their data schemas and interoperability standards. The core objective is to define the algorithmic foundations required to harmonize authoritative institutional repositories, such as the *Catalogo Generale dei Beni Culturali*, with non-traditional, crowd-sourced data sources like *Atlas Obscura*. Furthermore, the integration of multilingual terminological frameworks (e.g., *Architerms*) is addressed as a critical component for the development of context-aware systems.

The resulting analysis delineates a strategic roadmap for an architecture capable of integrating authoritative institutional catalogs with primary historical data, marking a first fundamental step toward a scalable, intelligent cyber-physical infrastructure for built heritage.

II. Research Projects Analysis (Methodological Focus)

The selection of case studies analyzed in this section is based on a purposeful sampling strategy designed to map the technological spectrum of digital heritage infrastructures. The rationale for this selection is structured as follows:

- *Domain-Specific Lexicography*: the *Glossario dell'Edilizia Romana* serves as a primary reference for construction history. It represents the state-of-the-art in specialist lexicographic research; however, its current reliance on static digital tables highlights the semantic gap that EUROGLOSS aims to bridge by transitioning from isolated records to a relational knowledge infrastructure.
- *Data Foundation and Linguistic Excellence*: *e-Leo* and the *Accademia della Crusca* were selected as benchmarks for high-fidelity digitization and archival metadata standards.
- *Semantic-Spatial Integration*: *Turin 1911* provides a critical example of integrating controlled vocabularies (e.g., Getty AAT) within spatial (3D/GIS) environments, a key requirement for the proposed Intelligent Reality framework.
- *Scalability and AI Automation*: the *Venice Time Machine* is included to evaluate the potential of large-scale data processing and semi-automated Knowledge Graph construction.

Based on this rationale, the following sections provide a technical evaluation of the platforms. The analysis focuses on architectural features and metadata schemas to define the EUROGLOSS design requirements.

A. Glossario dell'Edilizia Romana

The *Glossario dell'edilizia romana tra Rinascimento e Barocco*[4] was implemented as an illustrated dictionary of over 550 lemmas of historical construction. The project integrates entries in Italian and Latin (Italian glossary, glossary of obscure terms, and Latin glossary), accompanied by photos and illustrations. The IT architecture used is that of a web portal hosting the glossary pages; the data appear to be stored in simple indexed tables and static files, with no evidence of an advanced semantic infrastructure.

In practice, the system behaves like a lexicographic data entry queryable via a browser, with basic search functions on lemmas (the use of controlled vocabularies or ontologies is not documented). The data collection method relies on specialists (manual research of terms in treatises and archives), and the content is presented via HTML pages showing definitions, images, and bibliographic references.

B. Biblioteca Leonardiana

Similarly, the *e-Leo digital archive of the Biblioteca Leonardiana*[5] provides a clear example of an infrastructure based on texts and images. This system manages the complete digitization of Leonardo's works and related authors. The web interface offers various features (Browse mode, index of drawings, integrated glossary, and text search). The images of the works (originals and facsimiles) are acquired with high-quality scanners and the resulting texts are largely transcribed

[4]Glossario dell'Edilizia Romana: https://tinyurl.com/2d7jh5ev
[5]Biblioteca Leonardiana: https://www.leonardodigitale.com/

via optical character recognition (OCR). From a technical standpoint, the data (image files and OCR text) are managed by a web back-end (front-end analysis and metadata inspection suggest an architecture based on a database or XML); the portal does not allow content to be downloaded, in order to protect copyright. In summary, *e-Leo* demonstrates the use of digitization pipelines (scanning + OCR) and a web interface with an integrated text search engine.

C. Accademia della Crusca

The *digital archive of the Accademia della Crusca*[6] represents another case study: it collects descriptive metadata and digital reproductions of over 1450 archival units (volumes, bundles, letters, notebooks, card catalogs). From an IT perspective, it appears as a portal based on dynamic web pages: each document has a descriptive record with metadata fields (series, subseries, year, subject) and links to the scanned files. The internal storage appears to use a relational database or XML to manage the descriptions, and a file server for the images. Access is provided via keyword search and filters by archival series. Like *e-Leo*, the *Crusca digital archive* protects intellectual property: documents are accessible only within the portal without allowing downloads or copying. Overall, the Crusca system exhibits a metadata-driven approach with data normalization through archival guides and an index of units.

D. Turin 1911

The *Turin 1911: The World's Fair in Italy*[7] project introduces advanced technologies to integrate spatial and historical data. It is an interdisciplinary web application (UC San Diego and Politecnico di Torino) that reconstructs the pavilions of the 1911 Turin Fair in 3D. The site is divided into three main sections (The Project, Glossary, Explore) which organize the content by theme and topic. The sources include archival documents, historical maps, photographs, and CAD models; a GIS/3D pipeline is used to integrate these data. In terms of semantics, *Turin 1911* uses a controlled thesaurus (Getty AAT) to index domain concepts. In practice, each entity (for example, a pavilion, an object, or a place) is associated with terms from the Art & Architecture Thesaurus. This ensures consistency among the data and facilitates interoperability: for example, when the user searches for a term in the glossary, the system retrieves all records linked to the same AAT concept. The architecture also adopts open formats for geolocation (e.g. GeoJSON files or spatial databases) and XML for textual descriptions, although this is not explicitly documented. The interface includes a full-text search and spatial filters to explore the fair's data. In summary, *Turin 1911* demonstrates the use of geospatial databases and semantic vocabularies to integrate heterogeneous sources.

[6]Digital Archive of the Accademia della Crusca: https://tinyurl.com/3jzyfem6

[7]Turin 1911: https://italyworldsfairs.org/

E. Venice Time Machine

The *Venice Time Machine* (VTM) represents an even more advanced approach, aiming to build a historical "big data" archive of Venice. It is an international project (EPFL, Ca' Foscari, etc.) that digitizes over a thousand years of Venetian documents. The infrastructure relies on a high-throughput pipeline: robotic scanners acquire high-resolution images of archival records (decades of notarial deeds, censuses, maps); sophisticated machine-learning algorithms segment these images into text units and automatically identify handwritten words. Each recognized word is then converted into probabilistic text and subsequently reconciled into coherent occurrences. Finally, key terms extracted from these texts are linked into a massive knowledge graph: in practice, each entity (person, place, object) becomes a node connected to concepts, allowing navigation of historical relationships. The result is a relational archive of cultural data (implemented in an RDF triple store) that supports complex queries: for example, complete biographies or social networks can be reconstructed on the timeline of Venice. In summary, VTM combines massive data acquisition, automatic transcription (akin to HTR), and the construction of a knowledge graph.

III. Benchmarking of Existing Platforms (Technical Focus)

This section provides a technical benchmarking of existing digital platforms relevant to the EUROGLOSS project. It focuses on performance metrics, accessibility, and user experience (UX) to inform the development of the EUROGLOSS platform.

The proposed technical assessment combines both objective and subjective evaluations. More specifically, the objective evaluation is based on Google Lighthouse[8], following the methodology established in prior research such as [9], [10]. The open-source automated auditing tool produces scores from 0–100 and detailed recommendations to assess performance across four main categories:

- *Performance*: measured via First Contentful Paint (FCP) and Largest Contentful Paint (LCP), Speed Index (SI), Total Blocking Time (TBT), and Cumulative Layout Shift (CLS).
- *Accessibility*: ensuring the platform is usable by individuals with diverse needs.
- *Best Practices*: adherence to modern web development standards.
- *Search Engine Optimization (SEO)*: visibility and indexability by search engines.

In addition, Nielsen's Usability Heuristics [11] were used to subjectively assess the user interface and experience. The application of these heuristics is well-supported by the existing literature, including [12], [13]. The analysis reported in the current work specifically focused on the visibility of system status, consistency, standards, aesthetic, and minimalist design.

[8]Google Lighthouse: https://developer.chrome.com/docs/lighthouse

The selection of platforms analyzed in this section is based on a criterion-based purposive sampling approach [14], designed to cover the main functional dimensions relevant to the EUROGLOSS project. An initial pool of more than 20 candidate platforms was identified through web-based exploration of digital heritage repositories and terminology databases. Each candidate was then evaluated against five predefined functional criteria derived from the core EUROGLOSS requirements: (1) glossary-style terminological structure, (2) presence of visual documentation, (3) map-based or spatial visualization, (4) user interaction features, and (5) support for filtered or faceted search. Platforms that met at least three of the five criteria were retained. From this subset, the six platforms with the closest thematic alignment to the EUROGLOSS scope were selected through expert consensus among the research team.

A. Atlas des techniques de la Construction Romaine

The *Atlas des techniques de la Construction Romaine* (ACoR)[9] provides a structured repository of Roman construction techniques, organized through thematic categories and descriptive entries [15]. The platform adopts a hierarchical taxonomy rather than a fully semantic ontology. Terminological entries are presented as independent records with limited relational linking between concepts. From an IT perspective, the architecture relies on a traditional relational database with static content pages. Metadata structures are minimal, and cross-referencing between entries is manually curated. The absence of RDF or ontology-based modeling restricts semantic interoperability and machine-readable linking. Data presentation is text-centered, accompanied by images and explanatory diagrams. Navigation is primarily hierarchical, with limited faceted search capabilities. The system does not support spatial visualization or dynamic filtering.

This platform highlights the need to implement features like i) semantic relationships between terms, ii) dynamic filtering mechanisms, iii) persistent identifiers for entries, and iv) automated cross-linking between concepts.

B. WHC UNESCO

The *World Heritage Centre (WHC) UNESCO*[10] platform integrates structured heritage information with interactive map-based exploration. The system combines a relational backend with geospatial visualization components, allowing users to browse sites by region, typology, and thematic criteria.

The data model is metadata-driven, based on standardized heritage descriptors. However, semantic linking between lexical terms and heritage entities is limited. Terminology is embedded within descriptive texts rather than managed as structured vocabulary. From an interface perspective, the platform emphasizes map-based navigation, faceted search filters, and layered information panels. While effective for browsing, usability evaluations indicate limited user control over visualization layers and filtering persistence. The architecture demonstrates the importance of: GIS-based exploration, faceted metadata filtering, integration of textual and spatial data, scalable map visualization.

This platform suggests adopting a modular geospatial component integrated with semantic lexical data.

C. Historic England

Historic England[11] provides a national heritage database combining textual metadata with interactive cartographic visualization. The platform is built around a large-scale relational database supporting spatial queries and advanced filtering. The IT infrastructure prioritizes data retrieval performance and large dataset management. Spatial indexing enables geographic search, while metadata fields allow filtering by period, typology, and designation status. However, the system lacks explicit semantic modeling of terminology. The user interface presents dense clusters of map markers, which may reduce readability when exploring large datasets. This limitation highlights the importance of scalable visualization strategies such as clustering and progressive loading.

The platform demonstrates key features like i) spatial indexing for large datasets, ii) scalable map marker clustering, iii) advanced faceted search, and iv) integration of terminology with geographic entities

D. Catalogo Generale dei Beni Culturali

The *Catalogo Generale dei Beni Culturali*[12] represents a comprehensive national-level cultural heritage database. The system integrates multiple datasets and provides advanced search capabilities based on structured metadata.

The architecture relies on centralized data aggregation with a standardized cataloging schemas. Records include controlled vocabulary descriptors, geographic coordinates, and hierarchical classification. Visualization tools include map-based browsing and interactive exploration interfaces.

A notable feature is the integration of AI-assisted guidance for search and navigation, which supports users in discovering relevant content. This suggests the potential role of intelligent assistance in EUROGLOSS for navigating complex lexicographical datasets.

From a technical perspective, the platform emphasizes: i) standardized metadata schemas, ii) centralized aggregation of heterogeneous datasets, iii) AI-assisted search and navigation, interactive visualization tools.

E. Atlas Obscura

Atlas Obscura[13] is a platform combining cultural heritage data with user-generated contributions. The architecture supports storytelling, multimedia integration, and social interaction. The system uses a flexible content model allowing user submissions, editorial review, and multimedia enrichment.

[9]ACoR: https://acor.huma-num.fr/search/

[10]WHC UNESCO:https://whc.unesco.org/en/list

[11]Historic England: https://historicengland.org.uk/listing/the-list/

[12]Catalogo Gen. Beni Cult..: https://catalogo.beniculturali.it/

[13]Atlas Obscura: https://www.atlasobscura.com/

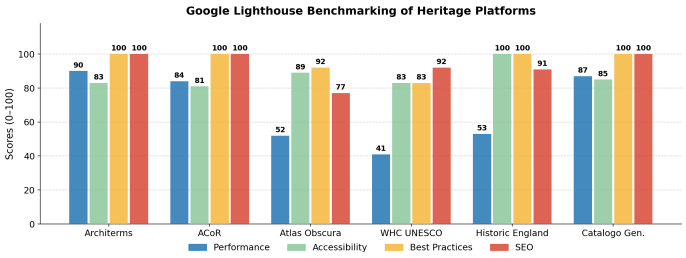


Fig. 2. Comparative analysis of the selected cultural heritage platforms based on Google Lighthouse performance indices. The benchmarking highlights the trade-offs between rich multimedia content and technical performance.

While effective for engagement, the platform prioritizes narrative content over structured terminology. Search and discovery mechanisms rely on tagging and thematic categorization rather than formal semantic modelling. This limits interoperability but enhances exploratory browsing.

The platform demonstrates both opportunities and risks. Among the advantages it is possible to mention: i) rich multimedia integration, ii) narrative contextualization, and iii) user engagement mechanisms. Limitations of the platform are: i) lack of controlled vocabulary, ii) inconsistent metadata structure, and iii) limited semantic interoperability

F. Architerms

Architerms[14] is a lightweight terminology platform focused on architectural vocabulary. The system provides concise definitions organized alphabetically and by thematic categories. The IT infrastructure appears minimal, prioritizing performance and simplicity. Entries are static and lack relational linking, spatial context, or semantic annotation. Search functionality is limited to keyword lookup. Despite its simplicity, the platform demonstrates usability advantages, including fast loading times and clear navigation. However, the absence of advanced data structures restricts scalability.

The analysis indicates the following features: i) high-performance interface design, ii) structured semantic relationships, iii) advanced search capabilities, and iv) integration with spatial and archival data.

IV. Discussion and Requirements for EUROGLOSS

A. Methodological Focus

The analyzed case studies reveal key technological requirements for EUROGLOSS:

Data acquisition and transcription: massive scanning of historical sources and automatic text recognition (OCR/HTR). Semantic indexing: use of controlled vocabularies (e.g., Getty AAT) and the creation of knowledge graphs to link terms, improving consistency and semantic search.

Open standards and formats: flexible encoding of content (XML/TEI for texts, RDF/OWL – e.g. CIDOC-CRM – for cultural metadata) to ensure international interoperability

[14]Architerms:https://www.architerms.com/

TABLE I
Performance indices evaluated through the Google Lighthouse: First Contentful Paint (FCP), Largest Contentful Paint (LCP), Speed Index (SI), Total Blocking Time (TBT), and Cumulative Layout Shift (CLS).

Platform	FCP	LCP	SI	TBT	CLS
Architerms	0.8s (0.06s)	1.0s (0.00s)	1.4s (1.15s)	31.8ms (20.52ms)	0 (0.00)
ACoR	1.66s (0.05s)	1.86s (0.05s)	3.22s (0.31s)	2.00ms (4.00ms)	0.01 (0.01)
Atlas Obscura	1.12s (0.12s)	2.02s (0.32s)	2.04s (0.19s)	50ms (26.08ms)	0.00 (0.00)
WHC UNESCO	1.30s (0.06s)	1.48s (0.10s)	2.44s (0.19s)	1378ms (1179.14ms)	0.07 (0.02)
Historic England	0.70s (0.00s)	3.30s (0.75s)	1.68s (0.98s)	534.00ms (210.77ms)	0.37 (0.01)
Catalogo Gen. Beni Cult.	0.50s (0.00s)	5.28s (2.40s)	0.92s (0.04s)	134.00ms (64.68ms)	0.00 (0.00)

(though this is not always explicitly mentioned in the examined projects).

Advanced user interface: full-text search engines and thematic/spatial exploration tools (filters by period, location, or category), as in the Explore section of Turin 1911.

Multilingual support: the ability to handle lemmas in multiple languages (e.g. bilingual glossaries Italian/Latin) and to integrate diverse lexicographic references.

Rights and access management: protection of digital content with controlled-access mechanisms (similar to *Crusca digital archive* and *e-Leo*, which prevent downloads and protect copyright).

B. Technical Focus

Objective results achieved by running the Google Lighthouse test are reported in Figure 2 and Table I.

It is worth noticing that *Architerms* serves as the benchmark for technical performance, achieving a score of 90/100 with an FCP of 0.8s. This efficiency is closely followed by *Catalogo Generale dei Beni Culturali*, which achieved a performance score of 87/100. In contrast, more complex platforms like *Historic England, Atlas Obscura* and *WHC UNESCO* show significantly lower performance indices, probably due to the overhead of heavier multimedia content and external scripts.

Architerms, *ACoR*, and *Catalogo Generale dei Beni Culturali* also achieved perfect scores (100/100) in the Best Practices and SEO categories, demonstrating the effectiveness of their technical optimization and the use of structured metadata in making heritage glossaries discoverable. While *Historic England* leads in Accessibility (100/100), *Architerms* and *ACoR* maintain high standards in this area with scores of 83/100 and 81/100.

The platforms were further analyzed using Nielsen's Usability Heuristics to assess the quality of the user interface (UI) and interaction design.

Visibility of system status: *Historic England* provides excellent accessibility (100/100) but suffers from "map clutter". When zooming in, individual elements overlap, obscuring the view. This finding identifies a critical need for "bubbling" or clustering techniques in EUROGLOSS to manage high data density across European sites.

Consistency and standards: *Atlas Obscura* was highlighted for its successful implementation of community-driven features ("Login/Upload"). However, it lacks visibility for its "Search Nearby" feature, which is often hidden from the user, violating the principle of "Recognition rather than Recall".

Specialized search functionality: *ACoR* provides a robust multi-criteria search that follows the logical steps of a construction site (foundations, elevations, etc.). However, its UI was noted as potentially confusing due to a "one at a time" navigation style that can hinder the comparative analysis of techniques.

C. Scalability and Data Management

A critical challenge identified across the benchmarked platforms concerns the capacity to handle increasing data volumes without degrading usability or semantic coherence. As observed in the case of *Historic England*, the accumulation of geospatial records leads to visual clutter when rendered on interactive maps, requiring clustering and progressive loading strategies. For EUROGLOSS, which is expected to grow continuously as the European network expands and new archival corpora are integrated, scalability must be addressed at both the data and interface levels.

At the data layer, the modular ontology design enables incremental extension without structural redesign [16]. The adoption of RDF triple stores — as demonstrated by the *Venice Time Machine* — provides horizontal scalability for semantic queries, while AI-assisted term clustering can partially automate the integration of new terminological entries, reducing manual curation overhead as the dataset grows.

At the interface level, dynamic marker clustering and lazy-loaded result sets will be essential for maintaining responsiveness at scale. Future evaluations should re-apply the Lighthouse benchmarking protocol reported here to monitor performance under increasing data loads, establishing a longitudinal quality baseline for the platform.

V. Conclusion

In conclusion, this paper establishes the foundational requirements for EUROGLOSS, an interoperable knowledge infrastructure designed to bridge the gap between physical architectural assets and their multifaceted digital representations. By employing a dual-track analysis, this work has demonstrated that the next generation of heritage platforms must excel not only in semantic depth but also in technical performance to support modern IR applications.

The methodological review of established projects, such as the *Venice Time Machine* and the *Glossario dell'Edilizia Romana*, highlights the necessity of adopting Semantic Web ontologies (e.g., CIDOC CRM) and encoding standards (XML/TEI) to manage complex historical narratives. Simultaneously, the above technical benchmarking via Google Lighthouse and Nielsen's Usability Heuristics identified critical performance bottlenecks in current systems, particularly regarding geospatial data density and mobile accessibility. These findings provide a clear roadmap for the EUROGLOSS architecture: a system that integrates high-fidelity historical data with high-performance, machine-readable assets.

The ultimate contribution of this work lies in defining a scalable framework for Knowledge Digital Twins. By transforming fragmented archival sources into standardized, semantically rich infrastructures, EUROGLOSS enables advanced human-machine collaboration. Future work will focus on the integration of conversational AI assistants and real-time XR visualizations, ensuring that the technical and linguistic layers of European construction history are preserved and accessible through authoritative, intelligent services.

References

[1] F. Meschini, *Oltre il libro: forme di testualità e digital humanities*. Editrice Bibliografica, 2020.

[2] F. Tomasi *et al.*, *Organizzare la conoscenza: Digital Humanities e Web semantico. Un percorso tra archivi, biblioteche e musei*. Editrice Bibliografica, 2022.

[3] L. Pandolfo, L. Cardone, L. Cutzu, R. Perna, B. Seligardi, and G. Simi, "The wow project: Bridging ai and cultural heritage for actress writings," in *Proc. Workshop on Artificial Intelligence for Cultural Heritage*, vol. 3536, 2023, pp. 34–41.

[4] A. Cannavo, G. Arrigo, A. Visconti, F. De Lorenzis, and F. Lamberti, "Designing social immersive virtual environments for the metaverse: The case study of metalibrary," *Virtual Reality & Intelligent Hardware*, vol. 7, no. 3, pp. 279–298, 2025.

[5] COST Association, "Memorandum of understanding for the implementation of the cost action "a glossary of technical construction vocabulary in 17th-18th century european court residences" (eurogloss) ca24102," 2025, brussels, 19 May 2025.

[6] P. Martens and K. De Jonge, "Palatium: Court residences as places of exchange in late medieval and early modern europe, 1400-1700," *European Architectural History Network Newsletter*, vol. 6, no. 2, pp. 16–25, 2012.

[7] V. Burgassi, "La struttura burocratica nei cantieri di corte sabaudi tra xvii e xviii secolo. organizzazione amministrativa per un progetto dinastico unitario," *ArcHistoR*, no. 20, pp. 64–95, 2023.

[8] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne *et al.*, "The fair guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[9] S. S. Muna, N. Nurdin, and T. Taufiq, "Tokopedia and shopee marketplace performance analysis using metrix google light-house," *International Journal of Engineering, Science and Information Technology*, vol. 2, no. 3, pp. 106–110, 2022.

[10] A. Lipiński and B. Pańczyk, "Performance optimization of web applications using qwik," *Journal of Computer Sciences Institute*, vol. 28, pp. 197–203, 2023.

[11] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proc. SIGCHI Conference on Human Factors in Computing Systems*, 1990, pp. 249–256.

[12] H. Y. P. Uly and D. B. Setyohadi, "A usability testing approach on a boarding house search website in kupang city using the nielsen heuristic method," in *Proc. Int. Conf. on Electrical, Electronics and Information Engineering*. IEEE, 2021, pp. 1–5.

[13] Ł. Krawiec and H. Dudycz, "Identification of heuristics for assessing the usability of websites of public administration units," in *Proc. Federated Conference on Computer Science and Information Systems*. IEEE, 2019, pp. 651–657.

[14] M. Q. Patton, *Qualitative Research & Evaluation Methods*, 4th ed. SAGE Publications, 2015.

[15] J.-P. Garric and S. Frommel, "Il glossario dell'edilizia romana: un progetto tra storia dell'architettura e digital humanities," *Studi di storia dell'arte*, 2020.

[16] N. F. Noy and D. L. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory, 2001.