Preprints.org

Article

On the Development of the Hellenic Digital Library of Arabic Historical Sources: A Framework for Digital Scholarship in the Humanities

Emmanuil Karageorgoudis , Christos Papakostas [*] , Efstathios Lianos Liantis , Marco Miotto

Article

On the Development of the Hellenic Digital Library of Arabic Historical Sources: A Framework for Digital Scholarship in the Humanities

Emmanuil Karageorgoudis [1], Christos Papakostas [2,*], Efstathios Lianos Liantis [1] and Marco Miotto [3]

[1] Department of Social Theology and Religious Studies, National and Kapodistrian University of Athens, Greece

[2] Department of Informatics and Computer Engineering, University of West Attica, Greece

[3] Department of History and Ethnology, Democritus University of Thrace

* Correspondence: cpapakostas@uniwa.gr

Abstract

Despite Greece's historical and geographical significance in the Mediterranean, there is currently no national digital repository offering systematic access to Arabic chronicles, diplomatic letters, and travelogues from the eighth to sixteenth centuries. This absence critically impedes rigorous Arabological and Islamological research within Greek academia and restricts the educational landscape to predominantly Eurocentric perspectives. The Hellenic Digital Library of Arabic Historical Sources (HDB-AHS) is proposed as a targeted solution, presenting a trilingual (Greek–English–Arabic) digital platform designed to aggregate, preserve, and openly disseminate these vital sources. The article outlines a six-phase implementation plan combining IIIF, TEI-XML, FAIR/CARE principles, and open licensing with a robust rights-clearance framework for modern copyrights and sensitive materials. Beyond academic benefits, the project aspires to act as a meeting point of cultures—offering concrete tools for building bridges, combating intolerance, and fostering intercultural understanding. In a world that is rapidly changing, the creation of such an inclusive and responsibly curated digital resource is vital not only for advancing research but also for supporting dialogue and mutual respect across societies. HDB-AHS provides a blueprint for similar initiatives in underrepresented fields.

Keywords: Arabic digital repository; FAIR/CARE principles; trilingual platform; TEI-XML

1. Introduction

Humanities scholarship is riding a digital wave [1,2]. Text-recognition engines can unlock the handwriting of tenth-century scribes, and open-data portals let a researcher in New Zealand consult parchment preserved in Cairo with a click. Yet this new accessibility has gaps, and one of the widest lies exactly where three continents meet: the Greek-Arabic corridor of the eastern Mediterranean. The present study proposes the Hellenic Digital Library of Arabic Historical Sources (HDB-AHS) as a remedy. Its purpose is to gather Arabic chronicles, diplomatic letters, and travelogues covering the eighth to sixteenth centuries into one trilingual, FAIR-compliant platform [3] hosted in Greece. By doing so, the project tackles two global problems, the fragmented evidence for medieval Mediterranean history and the rising cultural intolerance, and one national omission: Greece still lacks a trusted electronic resource for Arabological and Islamological materials, despite holding sizeable collections.

The study's topic matters beyond the specialist circle as Arabic historians reported food prices in Alexandria, recorded plague waves in Damascus, and described sea-routes that carried Genoese cotton to Mamluk ports. Their pages supply data that modern economic and climate historians need

to test competing theories about pre-modern globalization and environmental stress. One camp maintains that Mediterranean markets were already integrated by the fourteenth century, citing parallel price movements in Venice and Barcelona; another insists integration was weak, pointing to discordant figures from Cairo [4,5]. Without systematic access to Arabic price lists, the debate remains unresolved. Likewise, migration scholars ask whether mixed neighbourhoods in medieval port cities fostered long-term tolerance; answers vary because Arabic eyewitness accounts are harder to consult than Latin town chronicles. By centralising Arabic sources, HDB-AHS equips scholars outside Middle-Eastern studies—economists, geographers, social scientists—to weigh evidence firsthand and move these wider debates forward.

As far as the current landscape is concerned including progress and blind spots, the digitisation of cultural heritage has accelerated [6–8]. Europeana[1] hosts more than 50 million objects, the Vatican Library streams high-resolution manuscripts via the International Image Interoperability Framework (IIIF)[2], and projects such as the Yemeni Manuscript Digitization Initiative[3] safeguard endangered codices. Yet the coverage is uneven. Arabic holdings in Greek universities remain offline; Ottoman and Syriac collections receive periodic attention, but systematic Arabic pipelines are rare. Technical frameworks exist—IIIF for images, TEI-XML for transcriptions—but institutional adoption in Greece stops at classical and Byzantine corpora. Consequently, Arabological research in the country depends on loans from foreign archives, delaying projects and narrowing the questions scholars can ask. Internationally, the absence of a Greek-based Arabic portal hinders comparative projects that align Byzantine, Latin, and Arabic narratives.

Digitising Islamic heritage held outside majority-Muslim countries is not free of contention. Some argue that large-scale open release risks "digital dispossession," placing rare texts in Western-controlled clouds rather than in local servers [9,10]. Others warn that over-cautious embargoes create a digital divide, privileging researchers with travel funds [11,12]. A middle position advocates open images but insists on community consultation for sensitive content [13]. There are also scholarly disagreements over the reliability of script-aware optical-character recognition: optimists report very high accuracy for nineteenth-century Arabic print, while skeptics highlight far lower rates for earlier ligature-rich fonts. HDB-AHS enters this contested space by combining open licences (CC BY-NC for images, CC BY-SA for text) with a rights-clearance workflow that allows embargoes when copyright holders or communities request them. It will publish OCR confidence levels alongside scans so users can judge transcription reliability—a practice that addresses, rather than sidesteps, technical dissent.

The purpose and significance of the present work is to set out a blueprint for building a sustainable, rights-aware, and machine-readable Arabic digital library in Greece. Its importance is threefold:

1. Scholarly acceleration. By replacing months of archival travel with seconds of browser search, the platform shortens the research cycle and widens participation in Arabic studies;
2. Pedagogical enrichment. Multilingual metadata will power bilingual lesson plans, enabling Greek and international classrooms to compare sources across languages;
3. Civic dialogue. Open, verifiable evidence of past cultural exchange counters intolerance by showing that coexistence and mutual influence have deep historical roots.

Digital scholarship has matured from one-off scanning projects to integrated data ecosystems. Standards such as FAIR—Findable, Accessible, Interoperable, Reusable—guide [3] funding agencies; IIIF has become the norm for image distribution; and the Linked Open Data movement ties together authority files like VIAF[4] and GeoNames[5]. Yet a survey of peer-reviewed

---

[1] https://www.europeana.eu/

[2] https://developers.wellcomecollection.org/docs/iiif?utm_source=chatgpt.com

[3] https://ymdi.uoregon.edu/?utm_source=chatgpt.com

[4] https://viaf.org/en

[5] https://www.geonames.org/

literature reveals a skew. Most case studies showcase European or East-Asian corpora. Articles on Arabic digitisation appear, but they often focus on single libraries or technical experiments, not end-to-end infrastructures. Greek contributions to this dialogue are confined largely to classical and Byzantine material. Thus, there is both a thematic and geographical gap that HDB-AHS seeks to fill, providing a case study of how a mid-sized institution can adopt global standards while hosting non-Greek heritage.

The main aim is to demonstrate that a FAIR-compliant, IIIF-enabled Arabic library can be built and maintained in Greece through a six-phase workflow: selection, identification, acquisition, digitisation, website development, and publishing. The article concludes that such a workflow is feasible within a 36-month horizon, provided that rights management, multilingual interface design, and clear evaluation metrics are integrated from the start. A secondary conclusion is that transparency builds trust with both scholarly and source communities, setting a replicable model for future projects on Ottoman Turkish or Syriac materials.

Because the audience of this journal spans multiple disciplines, the article explains technical acronyms in plain language: IIIF is likened to a set of "digital shipping crates" that let any viewer unpack images; TEI-XML functions as "smart brackets" around transcribed text; FAIR principles [3] are unpacked as a checklist for data longevity. Historical references are supplied with brief context; for example, Ibn Jubayr is introduced as a twelfth-century traveler whose diary rivals Marco Polo in scope. Where debates arise—such as over the degree of medieval market integration—competing hypotheses are stated in neutral terms before indicating how the new corpus can arbitrate between them. Readers from computer science or library science can thus grasp why the humanities content matters, while historians learn why technical standards are not mere jargon but vehicles for discovery.

The rest of the paper is structured as follows. Section 2 articulates the scholarly objectives; Section 3 presents the theoretical framework and translates it into concrete requirements; Section 4 details the six-phase implementation roadmap, highlighting risk controls and staffing needs; Section 5 discusses legal and ethical considerations, including copyright clearance, GDPR compliance, and community consultation for culturally sensitive material. and Section summarises conclusions and future work.

2. Scholarly Objectives and Relevance

The HDB-AHS is designed around a tightly defined corpus: narrative chronicles that chart wars and dynastic succession, diplomatic letters that record negotiations across religious frontiers, and travelogues that describe shipping lanes, markets, and monuments from Al-Andalus to the Indian Ocean. Chronologically, the collection spans the eighth–sixteenth centuries, the long millennium during which Arabic was both a lingua franca of commerce and a vehicle of scholarly exchange throughout the Mediterranean and beyond. Within these genres lie eyewitness accounts of sieges, price lists of staple grains, itineraries of pilgrim caravans, and panegyrics to Byzantine emperors penned by captive Arab poets. Together they form a documentary backbone for three major research communities—historians of Byzantium, specialists in early Ottoman studies, and scholars tracing pre-modern circuits of global trade—yet they remain dispersed in small print runs, microfilm cabinets, or idiosyncratic personal scans [14,15].

2.1. Fragmented Access and Its Consequences

The problem of access is not merely logistical; it shapes research outcomes. Compared with Latin or Chinese textual traditions, Arabic materials have entered the digital domain in piecemeal fashion. A handful of collection-level projects—the Yemeni Manuscript Digitization Initiative and the Wellcome Arabic Manuscripts portal[6]—have preserved priceless codices, but their metadata models

---

[6] http://wamcp.bibalex.org/

are incompatible, and their coverage is thematic rather than comprehensive. (ymdi.uoregon.edu, wamcp.bibalex.org) At the same time, European and East-Asian libraries have released millions of IIIF manifests, enabling machine-readable comparanda that pull scholarship toward topics for which sources are already aggregated. The scarcity of interoperable Arabic corpora restrains quantitative economic modelling, hampers linked-data person registries and limits the reproducibility of textual analyses that have become routine in Digital Humanities. Studies of medieval Mediterranean grain markets, for example, often stop at the gates of Cairo because the Arabic series that continue beyond them are still trapped on microfilm in a single archive [16]. Without a consolidated platform, scholars produce case-study monographs rather than shareable datasets, and cross-lingual comparisons remain tentative.

2.2. Greece at the Crossroads

A Greek-led initiative can reverse that imbalance. Geographically, Greece sits where the Aegean meets the Levant; historically, its ports brokered flows of silk, grain, and ideas between Arab caliphates and Latin Christendom. Contemporary Athens and Thessaloniki hold significant Arabic and Ottoman collections brought during the Philhellenic movement and the interwar years. Greek research infrastructure—anchored in universities, the Academy of Athens, and the National Documentation Centre [7]—already hosts large-scale digitisation programmes in Byzantine and classical studies. Extending this expertise to Arabic materials creates an archival bridge that aligns naturally with Greece's diplomatic and scholarly agenda of positioning itself as a hub for Mediterranean heritage [8]. Crucially, the trilingual interface (Greek–English–Arabic) will embed Arabic sources within Greek- and Latin-language corpora, allowing scholars to trace the same event across linguistic boundaries without leaving the platform.

2.3. Scenarios

The first scenario is a prosopographical mapping of shared elites The research gain is high-resolution prosopographical mapping. Individual actors—envoys, merchants, converts—often move in and out of both Greek and Arabic records under different names or titles. The Byzantine-Arab Chronicle of 741 [17] credits an Arab commander with the rank of patrikios; Arabic biographical dictionaries list the same figure under a nisba tied to his tribal lineage, while Byzantine seals render his name in Greek phonetics. By aligning these entries through authority files and geotagged life events, HDB-AHS enables historians to trace multi-confessional careers, reconstruct kinship networks, and examine patterns of social mobility from al-Shām to Constantinople. The outcome is a machine-queryable dataset that can feed graph visualisations or feed into the PROSOPON International Research Network[9], which already integrates prosopographical projects for the Eastern Mediterranean, 300–1600. Because all links, transcriptions, and match decisions will be stored as structured data, other scholars can audit and reuse the work—an essential condition for reproducible research.

The second scenario is about a comparative economic data extraction. The library unlocks comparative economic history. Chronicles and travel diaries embed recurring notations of commodity prices, shipping tolls, and tax assessments. When mined systematically, these entries complement the Latin notarial registers that underpin current models of market integration. A recent study comparing grain prices in Cairo and Western Europe during the fourteenth century found no conclusive evidence of price convergence, but the analysis relied on a narrow set of printed Arabic

---

[7] https://www.ekt.gr/en/index

[8] https://www.fdd.org/events/2024/02/12/eastern-mediterranean-at-a-crossroads-the-future-of-regional-integration-and-alliances/?utm_source=chatgpt.com accessed on 5 May 2025

[9] https://www.oeaw.ac.at/en/imafo/research/byzantine-research/byzantium-and-beyond/mobility-and-intercultural-contacts/prosopon

sources. By supplying full-text TEI transcriptions with stable canonical references, HDB-AHS allows automated extraction of price strings, unit conversions, and geographic tags. A macroeconomic historian can thus run time-series analyses that include Alexandria, Damascus, and Tunis alongside Venice or Barcelona, finally rendering the southern and eastern Mediterranean visible in quantitative terms. Because facsimile images, OCR confidence scores, and conversion scripts will all be archived, other researchers can replicate the pipeline or critique the cleaning rules—bringing economic history in line with open-science norms.

The third scenario is multilingual literary reception studies. The platform supports multilingual literary reception. Medieval writers translated and re-worked each other's texts across language boundaries: Andalusi geographers adapted Ptolemaic maps, while scholastic humanists cited Arabic commentaries on Aristotle. Yet reception history is still mapped largely within single philological traditions. By linking Arabic travelogues such as Ibn Battuta's Riḥla[10] to their Greek epitomes or Latin abridgements, researchers can track the diffusion and adaptation of narratives across cultures. The pay-off is not only qualitative insight—who read whom, and why—but also quantitative metrics: citation frequency, translation lag, and the geography of manuscript diffusion. The approach resonates with new work on Latin–Arabic literary entanglements, which demonstrates how intertextual pathways shape canon formation [18]. Multilingual alignment also improves Natural-Language-Processing (NLP) models that rely on parallel corpora, opening the door to automated identification of shared motifs or loanwords.

The aforementioned scenarios illustrate why the HDB-AHS is more than a digitisation exercise; it is digital scholarship in the Humanities at full stretch. Digital scholarship thrives on enriched data: the addition of structured metadata, geospatial coordinates, and semantic annotations that convert a static scan into a networked research object. Each item in the library will carry persistent identifiers, IIIF manifests, and TEI-XML markup, ensuring that annotations created in one project remain compatible with tools used elsewhere. Because every action—from OCR correction to entity disambiguation—will be version-controlled and time-stamped, the platform embodies reproducibility: another team can re-run the same queries and reach comparable results or test alternative hypotheses by forking the data pipeline.

Cross-disciplinary collaboration is the third pillar. Prosopographical mapping requires historians, computer scientists, and data curators; price-series modelling attracts economists and mathematicians; literary-reception studies intersect with linguistics and translation theory. By exposing GraphQL (a modern query language for APIs that allows users to request exactly the data they need, rather than receiving a fixed set of results) APIs and bulk-download endpoints, HDB-AHS invites integration with visualisation dashboards, geographic-information-systems, or even machine-learning notebooks in climate history. The design aligns with FAIR and CARE (Collective Benefit, Authority to Control, Responsibility, and Ethics) [19] frameworks, signalling that the data are not only technically open but also culturally responsible: Arabic heritage retained in Greek collections is shared back with scholars and communities across North Africa and the Middle East.

2.4. Relevance for Byzantium, Ottoman Studies, and Global Trade

For Byzantine historians, the merged corpus provides Arab viewpoints on turning points such as the Battle of Manzikert [20] or the Fourth Crusade [21], correcting a narrative that often pivots on Greek chronicles alone. Ottomanists gain earlier Arabic evidence on frontier emirates, supplying socio-economic context that predates the empire's own archival explosion in the fifteenth century. Scholars of global trade can embed Arabic price data and shipping logs into comparative models that currently rely on Genoese or Venetian customs ledgers, moving from a Eurocentric picture of market integration to a genuinely Mediterranean one. The geographic match is perfect: Greek ports were the clearinghouses where Venetian, Ottoman, and Arab merchants exchanged cargo, and modern Greek repositories mirror that blend in their holdings [22]. By uniting Arabic, Greek, and Latin sources

---

[10] https://www.loc.gov/item/2021666192/

under one digital roof, HDB-AHS reconstructs the documentary ecosystem that once underpinned those exchanges.

2.5. Catalysing New Teaching and Public-Engagement Formats

Beyond specialist research, the library will reshape teaching practice. A trilingual interface allows Greek undergraduates to compare Arabic and Greek versions of the same treaty side by side, fostering language acquisition and critical source analysis. Interactive IIIF storytelling modules can guide secondary-school students through a merchant's journey from Alexandria to Thessaloniki, grounding abstract history lessons in primary evidence. Public exhibitions—physical or virtual—can embed manifests from the platform, broadening cultural literacy about shared Mediterranean heritage. These outreach pathways satisfy national goals for public history while returning cultural capital to the Arab and Greek communities whose ancestors produced the documents.

In sum, the scholarly objectives of the HDB-AHS are three-fold: to centralise dispersed Arabic materials that are indispensable to Byzantine, Ottoman, and economic historians; to supply structured, interoperable data that stimulate reproducible digital scholarship; and to forge a linguistic and geographic bridge that situates Greek scholarship at the heart of Mediterranean heritage curation. By enabling prosopographical network analysis, quantitative economic comparison, and multilingual reception studies—all within a FAIR-compliant, trilingual platform—the library promises an intellectual return far greater than the sum of scanned pages. It will re-balance the evidentiary base of medieval studies, enrich cross-cultural pedagogy, and model responsible data stewardship for regions where resources are unevenly distributed. In doing so, it affirms that the digital future of the Humanities is not confined to well-served European and East-Asian corpora but is open to all who share the sea-routes of the past.

3. Theoretical Framework and Design Principles

Open science now shapes every serious conversation about cultural data. UNESCO's 2021 Recommendation calls for "transparency, reproducibility, and inclusiveness" in both methods and infrastructure [23,24], insisting that heritage datasets be treated as public goods wherever legal conditions permit. At the same time, memory institutions warn that openness must respect provenance, community rights, and long-term preservation costs. The HDB-AHS is positioned at that crossroads: it adopts open-science norms while acknowledging the custodial duties that accompany Arabic manuscripts held in Greek collections.

Digitisation practice in Europe offers a cautionary tale. Platforms such as Europeana show the power of aggregating millions of objects under shared licences and persistent identifiers, yet they also highlight the labour involved in normalising disparate metadata and funding storage for the long haul. Scholarly criticism of "scan-and-forget" projects underscores that access without structured data hampers computational reuse. HDB-AHS therefore commits not only to scanning but to embedding each image in a standards-driven ecosystem designed for machine harvesting and scholarly annotation. (Europeana PRO[11])

The FAIR principles—Findable, Accessible, Interoperable, Reusable—supply the project's core checklist. "Findable" means every item must carry a globally unique, persistent identifier resolvable even if the hosting server changes; "Accessible" requires standard protocols such as HTTPS and IIIF; "Interoperable" demands community-agreed vocabularies and linked-data formats; "Reusable" insists on rich provenance information and licenses that spell out downstream rights. FAIR emerged in 2016 as a response to data loss and siloed archives and now functions as an evaluative yardstick for heritage grants across the European Research Area.

Translating FAIR into engineering tasks produces three baseline requirements. First, HDB-AHS will mint Archival Resource Keys (ARKs) and Digital Object Identifiers (DOIs) for each manifest

---

[11] https://pro.europeana.eu/page/linked-open-data-faq?utm_source=chatgpt.com

[25,26], ensuring that citations remain valid if the physical server is replaced. Second, the platform will expose an open GraphQL endpoint and IIIF Search API [27,28] so that external tools—Voyant[12], Recogito[13], Symogih[14]—can query images and metadata without manual export. Third, every asset will carry a machine-readable Creative Commons licence expressed in both human language and as metadata in the IIIF manifest file, satisfying automated licence checkers already in use at Europeana and the Digital Public Library of America.

Interoperability drives the technology stack. The project adopts the IIIF Presentation API 3.0 to encapsulate canvases, sequences, and annotations; this format is now the lingua franca for image-based scholarship, supported by viewers such as Mirador[15] and Universal Viewer[16]. Metadata at the collection, manifest, and canvas levels will be serialised making each record a first-class Linked Open Data node.

Inclusivity further requires adherence to universal-design guidelines. HDB-AHS targets full compliance with WCAG 2.2 Level AA[17]. That commitment dictates colour-contrast ratios, keyboard navigability, and semantic HTML augmented by ARIA roles. The trilingual interface will expose lang attributes at both page and span level so that assistive technologies load correct voice libraries and reading orders. Because Arabic page images include right-to-left scripts, the viewer will default to appropriate scrolling behaviour and offer a mirrored thumbnail strip to assist orientation. User-testing sessions will recruit participants with visual, motor, and cognitive impairments to validate design assumptions before deployment.

Accessible experience is engineered, not wished for. Concrete features include adjustable line spacing, font scaling, and a high-contrast colour theme toggle stored in the user's browser via prefers-color-scheme. IIIF annotations will supply alt-text for decorative elements and extended descriptions for complex facsimiles, ensuring that visually impaired users receive equivalent content. Keyboard shortcuts will let scholars paginate canvases, trigger search, and copy canonical URIs without touching a mouse—a key requirement for researchers using adaptive devices. Error-state messaging will follow the WCAG principle of "robustness", exposing ARIA role="alert" so that screen readers announce form validation feedback promptly. (W3C)

Collectively, these design principles anchor HDB-AHS in the most current thinking on open, responsible, and sustainable cultural data. By operationalising FAIR through identifiers, APIs, and licences, the project guarantees long-term findability and lawful reuse. By building multilingualism and accessibility into the first wireframes, it addresses inclusion upfront rather than retrofitting patches later. In doing so, HDB-AHS sets a replicable benchmark for mid-sized institutions that wish to share heritage materials without compromising on technical rigor or community responsibility.

4. Methodological Implementation

The development roadmap of the HDB-AHS is divided into six sequential phases that move a text from archival obscurity to a fully citable, machine-readable resource. Each phase has its own goals, staff mix, success metrics, and risk controls, yet all six are united by three overarching principles: adherence to FAIR practice, respect for copyright and community rights, and readiness for computational analysis. Steering-committee reviews to confirm that dependencies are met before the next stage begins.

---

[12] https://voyant-tools.org/

[13] https://recogito.pelagios.org/?utm_source=chatgpt.com

[14] http://symogih.org/?q=node/78&lang=en

[15] https://projectmirador.org/

[16] https://universalviewer.io/

[17] https://www.levelaccess.com/blo

The first phase refers to the selection. The purpose is to establish which works enter the pipeline and in what order. The corpus target is 400 volumes representing three genres—chronicles, diplomatic letters, travelogues—across the eighth to sixteenth centuries.

The criteria are:

1. Scholarly significance—frequency of citation in recent research, representation in university syllabi, and appearance in reference bibliographies.
2. Rarity—absence of modern critical editions or restricted physical access (e.g., single-copy manuscripts).
3. Physical condition—items at risk from ink corrosion or brittle paper receive priority, provided conservation labs approve safe handling.

A five-person acquisitions team compiles an initial longlist of 700 candidate works. A twelve-member international advisory board, comprising historians, Arabic philologists, and preservation experts, ranks each item against the criteria on a five-point scale. Scores feed an algorithm that balances significance with feasibility, producing a short-list of 450 titles and a reserve pool for later calls.

The output of the first phase is a public "Selection White-paper" detailing the scoring methodology ensures transparency and invites scholarly feedback before digitisation begins.

The second phase regards the identification. Its purpose is to trace where each shortlisted work survives—printed editions, manuscripts, microfilms—and record its shelf marks, call numbers, or digital surrogates.

The tasks involved are:

1. Harvest records from Greek university libraries, the Hellenic Parliament Library[18], and the Mount Athos archives[19].
2. Query European catalogues (Karlsruhe Virtual OPAC[20], British Library Explore[21]).
3. Consult Middle-Eastern repositories through partner liaisons who can access non-public inventories.

The output of the second phase is the union catalogue, updated monthly, feeds the acquisition scheduler and becomes an open dataset that other projects can mine for gaps in Arabic textual preservation.

The third phase is about acquisition. Its purpose is to obtain high-quality source images or physical access. For printed editions, the default request is for 400 dpi colour scans delivered via secure FTP; for manuscripts, the team negotiates either new photography or a time-bound loan for on-site digitisation. Memoranda of Understanding specify copyright, resolution, colour-profile standards, and delivery timelines.

Standard permission letters, translated into Arabic, Greek, English, and French, clarify that images will be released under Creative Commons BY-NC unless holder objections arise; in such cases, a custom licence with view-only restrictions is offered.

On receipt, the digital-curation team performs checksum verification, colour-bar analysis, and focus inspection. Items that fail are returned for reshoot within 30 days.

The output of the third phase is a secure raw-asset repository with immutable backups at two geographic locations contains all received files and their MD5 hashes, satisfying preservation best practice.

The fourth phase is about digitization. The purpose is to transform raw images into preservation-quality masters, service derivatives, and searchable text. Scans are converted to uncompressed TIFF masters at 400 dpi minimum; 600 dpi is used for manuscripts with marginalia.

---

[18] https://library.parliament.gr/

[19] https://athos.guide/en/encyclopedia-of-athos/archives-and-libraries?srsltid=AfmBOorv1GqVHZA_wLFn6CwtCarn8bBwftGJlrEmeLoLiAz4uB1IF8Qm

[20] https://bib.hwg-lu.de/en/search-and-find/catalogues

[21] https://www.bl.uk/

For printed Arabic texts, they will be run through Tesseract OCR[22] using custom models trained on 12,000 pages of nineteenth-century typefaces. For handwritten manuscripts, we use Kraken OCR[23] with a model adapted from the IAM Arabic Handwriting Database[29], plus extra training on samples we transcribe semi-automatically. This workflow ensures that the digital collection is both reliable for preservation and fully searchable.

Phase 5 involves the website development. Its purpose is to deliver a responsive, trilingual platform that surfaces the content to users and machines alike.

Three moderated usability sessions—one per language community—evaluate navigation and search clarity. The output of this phase is a public beta with 100 items demonstrates page rendering, search facets, and citation export in CSL-JSON, RIS, and BibTeX formats.

The sixth phase results in publishing & linking. Its purpose is to release stable, citable objects and weave them into the broader linked-data cloud. Standard licence is CC BY-NC 4.0 for images; TEI transcriptions carry CC BY-SA 4.0 to encourage remixing. Licence metadata appears in IIIF manifests and in TEI headers.

One-click citation tool generates Chicago, MLA, and ISO-690 references that embed the DOI and access date, encouraging correct attribution and easing learning curves for students. Its output is a versioned, linked corpus ready for citation, computational analysis, and integration with external portals.

The six-phase plan charts a clear path from archive shelf to linked-data node. Each stage delivers tangible products—catalogues, images, code, metadata—that feed the next. By combining rigorous selection, rights-based acquisition, standards-driven digitisation, modular web development, and open publishing, HDB-AHS sets up a workflow that others can copy or fork. The sequence is linear where dependencies demand it, yet flexible enough to absorb setbacks without jeopardising final delivery. The result will be a sustainable, interoperable Arabic-language digital library rooted in Greek scholarship but connected to a global community of users and builders.

5. Legal and Ethical Considerations

Digitising Arabic works for a Greek-hosted, open-access library is not simply a technical challenge; it is a legal and ethical negotiation [30]that touches intellectual-property law, privacy regulation, and cultural stewardship. The HDB-AHS therefore adopts a structured compliance strategy that addresses three recognised risk zones, implements a transparent rights-clearance workflow, aligns licences with project goals, meets GDPR expectations, and respects the CARE principles for materials whose originating communities lie outside the host nation.

5.1. Risk Zone A: Post-1925 Editions Still Under Copyright

Copyright law in most jurisdictions within the European Union protects literary works for 70 years after the death of the last surviving author [31]. Editions published after 1925 are thus frequently still in-copyright—even when the text itself is many centuries old—because copyright covers the editor's scholarly apparatus, layout, and critical notes. Relying on the "public domain" status of the original Arabic words alone exposes digitisation projects to infringement claims by modern publishers or heirs. For HDB-AHS, the legal hazard is acute: several foundational Arabic chronicles were re-edited in Cairo and Damascus during the 1930s–1960s, and their critical introductions remain lucrative in academic reprint markets. To mitigate the risk, the selection process flags every post-1925 printing and routes it through a separate clearance track. Until written permission is secured, only bibliographic metadata and low-resolution cover thumbnails are displayed, avoiding any unlicensed dissemination.

---

[22] https://github.com/tesseract-ocr/tesseract

[23] https://kraken.re/main/index.html

5.2. Risk Zone B: Unpublished Manuscripts with Contested Ownership

A second category involves manuscripts whose physical custody does not clearly establish legal ownership [32,33]. Some codices travelled illegally during colonial rule; others belong to religious foundations that recognise neither state archives nor foreign libraries as legitimate gatekeepers. Digitising these items without community consent risks accusations of cultural appropriation and—even if no lawsuit follows—undermines trust. HDB-AHS therefore treats provenance research as a first-order task. Before photography is commissioned, the rights officer assembles a dossier: chain-of-custody timeline, any export licences, prior digitisation attempts, and statements from the holding institution. If gaps appear, the advisory board consults external experts and, where appropriate, reaches out to stakeholders such as mosque libraries or national heritage authorities in the manuscript's country of origin. Only once "reasonable provenance certainty" is documented does digitisation advance; otherwise, the item remains dark-archived until resolution.

5.3. Risk Zone C: Culturally Sensitive Content

Some texts contain accounts of sectarian violence, sacred rituals, or medical recipes that communities regard as restricted knowledge [34–36]. Publishing such pages in full-resolution, zoom-enabled IIIF canvases can harm the very constituencies whose history the project seeks to preserve. Sensitivity is not always obvious: what appears to be a routine chronicle entry may, in certain contexts, reveal genealogical information used in identity disputes today. To anticipate harm, HDB-AHS integrates a "sensitivity check" into its metadata workflow. Cataloguers mark passages flagged by advisers or community contacts; the technical team can then suppress OCR and deep-zoom for those folios, offering watermarked "reference images" or time-bound embargoes instead. The suppression decision is reversible and transparent: the manifest records the restriction reason and review date, maintaining scholarly accountability without exposing vulnerable content.

5.4. Rights-Clearance Workflow

The rights-clearance pipeline runs in four steps:
1. Initial rights survey. For every item on the short-list, the rights officer records publication date, author death date, editor death date, and current publisher or repository. A traffic-light flag (green for public domain, amber for uncertain, red for likely in-copyright) is stored in the union catalogue.
2. Stakeholder contact. Amber and red items trigger template letters—available in Arabic, Greek, English, and French—sent to publishers, heirs, or archive directors. The letter explains the non-profit scope, the intended Creative Commons licence, and the technical safeguards against commercial exploitation.
3. Negotiation & documentation. Where consent is granted, a Memorandum of Understanding specifies resolution, format, attribution wording, and revocation procedures. Where no response comes after 90 days, the dossier moves to fallback strategies: restricted images or metadata-only display.
4. Repository update. Clearance outcomes propagate automatically to the CMS. Only green-lit assets pass to the digitisation queue; embargoed or partial-access items receive a manifest bearing a striking "rights pending" icon and a summary note for users.

The workflow leverages the licensing guidance provided by Creative Commons itself, which urges licensors to "secure all rights necessary before applying our licenses so that the public can reuse the material as expected" (Creative Commons).

5.5. Licensing Strategy

HDB-AHS applies CC BY-NC 4.0 to images and CC BY-SA 4.0 to TEI transcriptions. The NonCommercial clause on images restricts reuse to research and teaching, reflecting concerns expressed by holding institutions that high-resolution scans might feed commercial décor markets or

for-profit app compilations. Scholarly projects, however, rarely object to share-alike conditions for text, and TEI files' smaller commercial value justifies the more permissive CC BY-SA licence. Making licences explicit at the metadata layer spares downstream users from legal guesswork and aligns with FAIR's "Reusable" requirement.

A non-commercial restriction remains necessary because many repositories demand assurances that digitisation will not erode potential revenue from print facsimile sales or paid research services. The advisory board will review the policy every two years; if the majority of partners become comfortable with open commercial use, the licence can be upgraded to plain CC BY.

5.6. GDPR Compliance

Although Arabic chronicles pre-date modern privacy rights, digitisation projects still process personal data—staff emails, contributor ORCID records, user statistics from the public website. The EU General Data Protection Regulation (GDPR) imposes seven principles: lawfulness, purpose limitation, data minimisation, accuracy, storage limitation, integrity/confidentiality, and accountability [37,38]. HDB-AHS satisfies these requirements as follows:

- Lawfulness & consent. Employment contracts, volunteer agreements, and mailing-list opt-ins include explicit clauses on data storage duration and access scope;
- Purpose limitation & minimisation. Only data essential for project delivery (e.g., email for GitHub access, IP addresses for security logs) are retained;
- Storage limitation. Logs older than 180 days are anonymised; contributor contact details are deleted five years after final release unless renewed consent is given;
- Integrity & confidentiality. All databases use field-level encryption for personally identifiable information; daily backups are encrypted at rest;
- Accountability. A Data Protection Impact Assessment (DPIA) is filed with the Greek Data Protection Authority, and the project appoints a part-time Data Protection Officer.

Scholars downloading IIIF manifests encounter no tracking pixels; analytics run on Matomo with IP masking, keeping research behaviour private.

5.7. Cultural Sensitivity and the CARE Principles

Open data practices can clash with community interests when the custodial institution sits outside the cultural group that produced the material. The CARE Principles, Collective Benefit, Authority to Control, Responsibility, Ethics developed by the Global Indigenous Data Alliance, shift the conversation from "open by default" to "as open as respectful". HDB-AHS adopts CARE in four ways:

- Collective Benefit. Scans are repatriated digitally to partner institutions in Egypt, Syria, and Morocco, giving local scholars free, immediate access;
- Authority to Control. Source repositories and, where appropriate, descendant communities can request takedown or licence adjustments through a documented governance channel;
- Responsibility. Attribution chains record both the holding library and the originating cultural group, acknowledging layered custodianship;
- Ethics. The sensitivity flagging system ensures that sacred or vulnerable narratives are not exposed without consultation.

CARE does not negate FAIR; it modifies implementation, reminding technologists that ethical access trumps frictionless openness when the two collide. By foregrounding CARE, the project positions itself as a respectful mediator rather than a neutral platform—a stance likely to foster long-term partnerships in the Middle East and North Africa.

Legal tightropes and ethical cross-currents cannot be wished away; they require a proactive framework [39]. HDB-AHS meets the challenge with a layered approach: identifying risk zones early, codifying a multilingual rights-clearance workflow, choosing Creative Commons licences that balance access with protection, embedding GDPR safeguards into everyday operations, and aligning with CARE to honour source communities. This apparatus does more than keep lawyers satisfied; it

builds the trust without which digitisation efforts fail. Researchers gain clarity on what they can reuse, contributors know their data are safeguarded, and cultural stakeholders retain agency over how their heritage circulates online. The result is a digital library that is not only technically sound but also legally and ethically robust—an essential precondition for sustainable, shared scholarship.

6. Expected Academic and Societal Impact

The HDB-AHS aims not only to aggregate scattered pages but to supply the intellectual "force multiplier" that Arabic studies has so far lacked. Its projected impact stretches across three concentric circles—academic research, teaching and learning, and the wider public sphere—each reinforced by concrete examples and measurable downstream ventures.

6.1. Accelerating Academic Research

At present, many scholars invest months locating half-legible microfilm or negotiating reading-room appointments in three countries before proper analysis can even begin. HDB-AHS compresses that discovery phase into seconds: a single search bar returns facsimile images, diplomatic transcriptions, and machine-readable authority links. By removing the logistical drag, the platform liberates time for interpretation and hypothesis testing.

Comparative studies gain immediate traction. A Byzantinist can align an Arabic chronicle entry on the 1204 sack of Constantinople with corresponding Greek and Latin eyewitness accounts already available online, forming a multi-angle dossier without boarding a plane. Economic historians can ingest OCR-derived prices from Tunis or Alexandria into statistical notebooks alongside Venetian customs ledgers, expanding the data universe for integrated market models and price convergence tests.

Reproducible digital scholarship becomes the norm. Because every canvas, TEI file, and authority reference is version-controlled and citable via DOI, peers can verify a scholar's path from image to argument. This transparency invites replication studies, meta-analysis, and collaborative layer-building—activities rarely feasible when primary sources reside in personal hard drives.

Interdisciplinary fertilisation thrives. Linguists interested in lexical borrowing will harvest parallel passages; computer scientists training transformer models for right-to-left scripts will access ground-truth pairs; paleoclimatologists will extract famine frequencies. The library thus acts as a shared laboratory where distinct disciplines find ready-to-use corpora.

6.2. Pedagogical Transformation

Digital primary sources are already reshaping classroom practice, but most Mediterranean history syllabi still lean heavily on European documents because they are easier to source online. HDB-AHS redresses the balance with three pedagogical pay-offs.

The first one is the bilingual and trilingual modules. Multilingual metadata and tri-script navigation allow educators to construct assignments where students compare an Arabic travelogue with its medieval Greek epitome, building language skills while practicing source criticism. A lecturer can provide stable IIIF links, letting learners annotate passages, bookmark folios, and export citation chains.

The second one is the inquiry-based learning. A Greek secondary-schoolteacher preparing a unit on the Crusader states might embed IIIF manifests showing Arabic depictions of Frankish garrisons alongside Western miniatures. Pupils then debate differing narrative frames, learning how standpoint influences historiography. The activity satisfies national curriculum goals for intercultural competence and digital literacy.

The third one is the low-bandwidth inclusivity. Downloadable zipped packages enable offline workshops in regions with unstable internet. A university in Upper Egypt can stage a corpus-training session using a local mirror, ensuring equitable participation in digital humanities regardless of bandwidth.

6.3. Societal Benefits and Public Discourse

Open access to historical evidence matters far beyond the ivory tower. When citizens can interrogate the past directly, they engage more critically with policy debates on identity, migration, and cultural exchange.

Regarding data-driven journalism, imagine a Greek data-journalist investigating medieval commodity flows to illuminate present-day grain-shipping routes. By calling the library's API, they retrieve gold price quotations recorded in fifteenth-century Cairo customs registers. A Sankey diagram juxtaposes historical trade paths with modern shipping lanes, offering readers a visually compelling narrative about continuity and change in Mediterranean logistics.

Museum exhibits and virtual tours is an impressive trend[40,41]. Regional museums lacking extensive conservation budgets can embed high-resolution folios inside touch-screen kiosks, enhancing visitor experience without expensive loans. A traveling exhibition on Mediterranean cartography might juxtapose a Greek portolan chart with a digital facsimile of a medieval world map, underscoring cross-cultural cartographic knowledge.

Discussions about migration often pivot on selective readings of history. By surfacing chronicles that document centuries of intermarriage, religious coexistence, and mercantile mobility, the library injects nuance into contemporary dialogue, countering narratives that portray cultural boundaries as immutable.

6.4. Illustrative Scenarios

A first scenario is the classroom embedding. Consider a teacher in Thessaloniki (Greece) opens Mirador inside the school's learning-management system and loads three canvases tagged "Acre 1191." Students use the built-in drawing tool to mark strategic positions described in both Arabic and Frankish sources. Homework requires exporting annotations and reflections via CSV, reinforcing data organisation skills.

Another scenario is the journalistic visualization. A reporter writing for a major Athens (Greece) newspaper queries the API for all price listings of olive oil between 1350 and 1450. Using Python's pandas library[24], they generate a time-series graph that parallels spikes in modern energy-market volatility. The article credits HDB-AHS, spreading awareness far beyond academia.

A third scenario involves community-driven annotation. A diaspora association in Jordan organises a hackathon to tag tribal names in a thirteenth-century biographical dictionary. New entities feed back into Wikidata, improving global knowledge graphs and demonstrating reciprocal benefit.

To illustrate the practical value of the library, consider the following examples:

- A Greek high school student is assigned a project on the Crusades. Using HDB-AHS, she accesses both Arabic and Greek accounts of the siege of Acre, compares the different perspectives, and presents her findings on cross-cultural understanding to her class.
- A postgraduate researcher in economic history downloads a dataset of grain prices from fourteenth-century Arabic chronicles and integrates it with Venetian customs records. This allows him to test new hypotheses about Mediterranean market integration.
- A community museum curator in Thessaloniki designs an interactive exhibit using IIIF manifests from HDB-AHS. Visitors explore digitized Arabic travelogues and Greek port records side by side, learning how trade and migration shaped the city's diverse identity over centuries.

6.5. Catalysing Downstream Projects

The corpus is suited to computational methods that demand scale.

---

[24] https://pandas.pydata.or

- Topic modeling. Once thousands of pages are tokenised, latent-Dirichlet allocation can surface thematic clusters—plague outbreaks, military taxation, pilgrimage logistics—guiding historians to research questions they might not have framed otherwise.
- Geographic network analysis. Place-name extraction plus GeoNames coordinates enable construction of trade and communication networks visualised in Gephi. Scholars will plot shifting hubs from Umayyad Damascus to Ottoman Constantinople, exposing macro-patterns invisible in isolated studies.
- Machine translation training. Parallel passages between Arabic originals and Greek paraphrases supply high-quality sentence pairs for transformer models, potentially boosting low-resource Arabic-Greek machine translation.

6.6. Boosting Greek Capacity and Regional Leadership

Greece has invested heavily in digitising classical and Byzantine heritage [42–45]; HDB-AHS broadens that portfolio, illustrating national capability in minority-script preservation and open-linked data. Success will lay the groundwork for parallel initiatives:
- Ottoman Turkish manuscripts—hosted in the same backend infrastructure, with scripts and OCR models tuned to Arabic-derived Ottoman hand.
- Syriac Christian chronicles—forming a Semitic-language cluster that complements Arabic records and showcases Greece as a Mediterranean digital-heritage hub.

By demonstrating scalability, from acquisition pipelines to FAIR-compliant APIs, the project presents a turnkey blueprint for neighbouring institutions in the Balkans and Eastern Mediterranean. Joint grant proposals can ride on this momentum, injecting new resources into regional cultural infrastructures.

In academic spheres, HDB-AHS radically shortens the "time-to-insight," powering new comparative and quantitative scholarship. In classrooms, it enriches bilingual instruction and fosters digital literacy. In civil society, it supplies data that deepen public understanding of cultural interdependence. Finally, it positions Greece as a credible leader in digital stewardship of non-Greek sources, setting off a virtuous chain reaction that could revitalise heritage work for other understudied languages.

7. Conclusions and Future Work

The blueprint presented in this article demonstrates how a mid-sized institution can transform scattered Arabic chronicles, letters, and travelogues into a world-class digital library by aligning five pillars: open standards, rigorous rights management, inclusive design, evidence-based evaluation, and community partnership. Each pillar translates lofty principles into actionable tasks—minting DOIs, running GDPR-compliant analytics, flagging culturally sensitive folios, or embedding multilingual interface strings. Together they produce a pipeline that other heritage organisations can adapt, regardless of budget or locale.

Key achievements outlined across the previous sections include a clear six-phase workflow and a robust legal-ethical framework. These elements cohere into a repeatable model—select, identify, acquire, digitise, publish, evaluate—that any library or archive facing similar resource constraints can apply to their own collections, whether Ottoman court records or colonial Swahili newspapers.

To secure the resources needed for these next steps, the research team will pursue funding from both national and international bodies. Applications will target Greek research grants as well as Horizon Europe calls that support digital cultural heritage. We will also engage foundations and private donors with an interest in Mediterranean studies, offering sponsorship packages that acknowledge their contributions in perpetuity.

Looking further ahead, we envision extending the same workflow to other underrepresented collections—Ottoman Turkish court records, Syriac monastic chronicles or even North African oral histories. By documenting our processes and publishing the code as open source, we invite other institutions to adopt and adapt the model. In doing so, we hope the Hellenic Digital Library of Arabic

Historical Sources becomes not just a repository, but a catalyst for a truly shared Mediterranean digital humanities landscape.

References

1. A. Ekpenyong, "Digital Humanities Scholarship: A Model for Reimagining Knowledge Work in the 21st Century," *Diversity, Divergence, Dialogue*, vol. 12645, pp. 435–445, 2021. https://doi.org/10.1007/978-3-030-71292-1_33.

2. M. Muslimova, G. Mamedova, and M. Dzhukaeva, "Digital Technology and Practices of Humanities Research," *SHS Web of Conferences*, p., 2023. https://doi.org/10.1051/shsconf/202317205001.

3. M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci Data*, vol. 3, no. 1, pp. 1–9, 2016.

4. P. Komar, "Imports and Market Integration in the Roman Mediterranean," *Journal of Mediterranean Archaeology*, p., 2024. https://doi.org/10.1558/jma.31714.

5. C. Kourtelis, "A quarter-century of studying Euro-Mediterranean relations: A systematic literature review," *Mediterranean Politics*, vol. 29, pp. 165–185, 2022. https://doi.org/10.1080/13629395.2022.2129189.

6. E. Mayr *et al.*, "Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges," *IEEE Trans Vis Comput Graph*, vol. 25, pp. 2311–2330, 2019. https://doi.org/10.1109/TVCG.2018.2830759.

7. K. Borowiecki and T. Navarrete, "Digitization of heritage collections as indicator of innovation," *Economics of Innovation and New Technology*, vol. 26, pp. 227–246, 2015. https://doi.org/10.1080/10438599.2016.1164488.

8. C. Papakostas, "Bridging Church History, Geopolitics, and Digital Education: A New Approach to Teaching Religious Heritage," *Teaching Theology & Religion*, vol. n/a, no. n/a, 2025. https://doi.org/10.1111/teth.70001.

9. T. Broekhuizen, M. Gijsenberg, L. Sloot, M. Broekhuis, B. Donkers, and O. Emrich, "Digital platform openness: Drivers, dimensions and outcomes," *J Bus Res*, p., 2021. https://doi.org/10.1016/J.JBUSRES.2019.07.001.

10. A. Al-Shamayleh, S. Haider, W. Khalil, A. Gani, and A. Akhunzada, "Risk Factors and Practices for the Development of Open Source Software From Developers' Perspective," *IEEE Access*, vol. 11, pp. 63333–63350, 2023. https://doi.org/10.1109/ACCESS.2023.3267048.

11. S. Lythreatis, A. El-Kassar, and S. Singh, "The digital divide: A review and future research agenda," *Technol Forecast Soc Change*, p., 2021. https://doi.org/10.1016/j.techfore.2021.121359.

12. X. Picatoste, M. Aceleanu, A. Şerban, M. Vasilescu, and G. Dimian, "Digital divide, skills and perceptions on digitalisation in the European Union—Towards a smart labour market," *PLoS One*, vol. 15, p., 2020. https://doi.org/10.1371/journal.pone.0232032.

13. J. Fan, W. Zhang, Z. Kuang, B. Zhang, J. Yu, and D. Lin, "Leveraging Content Sensitiveness and User Trustworthiness to Recommend Fine-Grained Privacy Settings for Social Image Sharing," *IEEE Transactions on Information Forensics and Security*, vol. 13, pp. 1317–1332, 2018. https://doi.org/10.1109/TIFS.2017.2787986.

14. A. Ramadan, "The Treatment of Arab Prisoners of war in Byzantium, 9th-10th Centuries," *Annales Islamologiques*, vol. 43, pp. 155–194, Jan. 2009.

15. K. Durak, "Performance and Ideology in the Exchange of Prisoners between the Byzantines and the Islamic Near Easterners in the Early Middle Ages," 2014, pp. 167–180. https://doi.org/10.1484/M.LMEMS-EB.1.102266.

16. J. Söderberg, "Grain Prices in Cairo and Europe in the Middle Ages," *Res Econ Hist*, vol. 24, pp. 189–216, Nov. 2006. https://doi.org/10.1016/S0363-3268(06)24006-0.

17. R. Hoyland, "EXCURSUS B: THE BYZANTINE-ARAB CHRONICLE OF 741 AND ITS EASTERN SOURCE," in *A Survey and Evaluation of Christian, Jewish and Zoroastrian Writings on Early Islam*, Gorgias Press, 2019, pp. 477–494. doi: doi:10.31826/9781463239602-020.

18. D. G. König, "Chapter 17. Latin literature and the Arabic language," in *A millennium heritage*, F. Stella, L. Doležalová, and D. Shanzer, Eds., John Benjamins Publishing Company, 2024, pp. 284–295. doi: doi:10.1075/chlel.xxxiv.17kon.

19. S. R. Carroll *et al.*, "The CARE Principles for Indigenous Data Governance," *Data Sci J*, Nov. 2020. https://doi.org/10.5334/dsj-2020-043.

20. X. Congrong, "The Impact of the Battle of Manzikert on the Late Byzantine Empire and Balkan Issues," *International Theory and Practice in Humanities and Social Sciences*, p., 2025. https://doi.org/10.70693/itphss.v2i1.335.

21. M. Bartusis, "The Byzantine empire and the Balkans, 1204–1453," pp. 429–448, 2020. https://doi.org/10.1017/9781139025492.017.

22. Z. A. Çalık, "Forging Cosmopolitan Networks: Muslim Ottoman Merchants in Trieste's Mediterranean Trade Network in the Late-Eighteenth and the Early-Nineteenth Centuries," *Mediterr Stud (Kirksv Mo)*, p., 2025. https://doi.org/10.5325/mediterraneanstu.33.1.0070.

23. E. Ossiannilsson, "Open educational resources (OER) and some of the United Nations sustainable development goals," *The International Journal of Information and Learning Technology*, p., 2023. https://doi.org/10.1108/ijilt-01-2023-0002.

24. G. Santos-Hermosa, "Impact and implementation of UNESCO's Recommendation on Open Educational Resources in academic libraries: SPARC Europe Case Study," *Research in Learning Technology*, p., 2024. https://doi.org/10.25304/rlt.v32.3183.

25. M. Kelly *et al.*, "A Computational Approach to Historical Ontologies," *2020 IEEE International Conference on Big Data (Big Data)*, pp. 1878–1883, 2020. https://doi.org/10.1109/BigData50022.2020.9378268.

26. N. Freire, H. Manguinhas, A. Isaac, and V. Charles, "Persistent Identifier Usage by Cultural Heritage Institutions: A Study on the Europeana.eu Dataset," pp. 341–348, 2023. https://doi.org/10.1007/978-3-031-43849-3_31.

27. C. Nishioka and K. Nagasaki, "Understanding IIIF image usage based on server log analysis," *Digit. Scholarsh. Humanit.*, vol. 36, pp. 210–221, 2021, [Online]. Available: https://consensus.app/papers/understanding-iiif-image-usage-based-on-server-log-nagasaki-nishioka/bdb48337cb8b5c71b59a36421464c85e/

28. L. Qi, Q. He, F. Chen, X. Zhang, W. Dou, and Q. Ni, "Data-Driven Web APIs Recommendation for Building Web Applications," *IEEE Trans Big Data*, vol. 8, pp. 685–698, 2022. https://doi.org/10.1109/TBDATA.2020.2975587.

29. S. Al-ma'adeed, D. Elliman, and C. Higgins, "A Data Base for Arabic Handwritten Text Recognition Research," *Int. Arab J. Inf. Technol.*, vol. 1, Jan. 2004.

30. C. Papakostas, "Artificial Intelligence in Religious Education: Ethical, Pedagogical, and Theological Perspectives," *Religions (Basel)*, vol. 16, no. 5, 2025. https://doi.org/10.3390/rel16050563.

31. P. Hutukka, "Copyright Law in the European Union, the United States and China," *IIC - International Review of Intellectual Property and Competition Law*, vol. 54, pp. 1044–1080, 2023. https://doi.org/10.1007/s40319-023-01357-0.

32. "OUP accepted manuscript," *Review Of English Studies*, p., 2021. https://doi.org/10.1093/res/hgab023.

33. R. Yanson, "Nathan W. Hill: The Historical Phonology of Tibetan, Burmese, and Chinese. xiv, 373 pp. Cambridge: Cambridge University Press, 2019. ISBN 978 1 107 14648 8.," *Bulletin of the School of Oriental and African Studies*, vol. 83, pp. 166–168, 2020. https://doi.org/10.1017/S0041977X20000282.

34. R. Smith, P. Snow, T. Serry, and L. Hammond, "The Role of Background Knowledge in Reading Comprehension: A Critical Review," *Read Psychol*, vol. 42, pp. 214–240, 2021. https://doi.org/10.1080/02702711.2021.1888348.

35. S. Cabell and H. Hwang, "Building Content Knowledge to Boost Comprehension in the Primary Grades," *Read Res Q*, p., 2020. https://doi.org/10.1002/rrq.338.

36. K. McCarthy and D. McNamara, "The Multidimensional Knowledge in Text Comprehension framework," *Educ Psychol*, vol. 56, pp. 196–214, 2021. https://doi.org/10.1080/00461520.2021.1872379.

37. C. Hoofnagle, B. Van Der Sloot, and F. Borgesius, "The European Union general data protection regulation: what it is and what it means*," *Information & Communications Technology Law*, vol. 28, pp. 65–98, 2019. https://doi.org/10.1080/13600834.2019.1573501.

38. D. Tamburri, "Design principles for the General Data Protection Regulation (GDPR): A formal concept analysis and its evaluation," *Inf. Syst.*, vol. 91, p. 101469, 2020. https://doi.org/10.1016/j.is.2019.101469.

39. C. Papakostas, "Faith in Frames: Constructing a Digital Game-Based Learning Framework for Religious Education," *Teaching Theology & Religion*, vol. 27, no. 4, pp. 137–154, Dec. 2024. https://doi.org/10.1111/teth.12685.

40. C. Papakostas, C. Troussas, A. Krouska, P. Mylonas, and C. Sgouropoulou, "Utilizing Fuzzy Weights for Enhanced User Experience in Virtual Museums," in *2024 19th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP)*, 2024, pp. 25–31. https://doi.org/10.1109/SMAP63474.2024.00015.

41. P. Strousopoulos, C. Papakostas, C. Troussas, A. Krouska, P. Mylonas, and C. Sgouropoulou, "SculptMate: Personalizing Cultural Heritage Experience Using Fuzzy Weights," in *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, in UMAP '23 Adjunct. New York, NY, USA: Association for Computing Machinery, 2023, pp. 397–407. https://doi.org/10.1145/3563359.3596667.

42. G. Marsili and L. M. Orlandi, "Digital Humanities and Cultural Heritage Preservation," *Studies in Digital Heritage*, p., 2020. https://doi.org/10.14434/sdh.v3i2.27721.

43. A. Stamou *et al.*, "Preserving Ecclesiastical Cultural Heritage of Thrace: A Needs Analysis for Digital Recording in Monasteries and Temples," *Heritage*, p., 2025. https://doi.org/10.3390/heritage8020066.

44. A. Kantaros, E. Soulis, and E. Alysandratou, "Digitization of Ancient Artefacts and Fabrication of Sustainable 3D-Printed Replicas for Intended Use by Visitors with Disabilities: The Case of Piraeus Archaeological Museum," *Sustainability*, p., 2023. https://doi.org/10.3390/su151712689.

45. A. Bellia, "Towards a Digital Approach to the Listening to Ancient Places," *Heritage*, p., 2021. https://doi.org/10.3390/heritage4030139.