

# Revealing Medieval Manuscript Treasures: Semantic Web Integration through a Polymorphic Knowledge Graph and Linked Open Data

Faria FEROOZ <sup>a,1</sup>, Monica PALMIRANI <sup>a</sup>

<sup>a</sup> *CIRSFID-ALMA-AI, University of Bologna, Italy*

## Abstract.

Medieval manuscripts represent rich and heterogeneous cultural heritage resources, offering interdisciplinary insights into historical context, textual content, physical features, and artistic elements. However, integrating such diverse data remains challenging due to inconsistencies in metadata schemas and variations in data quality. This article addresses these challenges through a semantic web-based approach applied to two distinctive medieval manuscript collections: Progetto Irnerio and Mosaico. It presents an extension of the existing Medieval Manuscript Data Integration Ontology (MMDIO), originally developed using the MeLON methodology and evaluated with the FOCA framework. This extended ontology builds upon the previously published MeMO ontology, introducing new classes and relationships designed specifically for the integration of medieval manuscript data. A key contribution is the construction of a Polymorphic Knowledge Graph that semantically integrates heterogeneous datasets from both collections, enabling faceted search, semantic browsing, and advanced visualization. Additionally, the MELIORATE Linked Open Data (LOD) platform is developed to provide unified, interoperable online access to manuscript content, significantly enhancing data accessibility and supporting interdisciplinary collaboration. This integrated approach demonstrates the potential of semantic web technologies to bridge disciplinary gaps among Digital Humanities, Legal Studies, and Computer Science, offering new methodological opportunities for cultural heritage research and digital preservation.

**Keywords.** Semantic Web, Ontology Design, Medieval Manuscripts, Polymorphic Knowledge Graph, Linked Open Data, Cultural Heritage, Digital Humanities

---

<sup>1</sup>Corresponding Author: Faria Ferooz, University of Bologna, Italy; E-mail: [faria.ferooz2@unibo.it](mailto:faria.ferooz2@unibo.it).

## 1. Introduction

The advancement of digital humanities has transformed cultural heritage research by integrating digital technologies with traditional humanistic methods. Within this evolving interdisciplinary context, medieval manuscripts hold a unique place because of their exceptional richness and complexity. These artifacts are considered precious not only because of their textual and artistic content but also from a legal and historical perspective. Features such as glosses, marginalia, layered annotations, and physical attributes provide rich insight into the production and transmission of knowledge across time [1,2].

Despite their scholarly importance, medieval manuscripts pose substantial challenges for digital representation. Their diversity in both physical composition and scholarly interpretation makes it difficult to develop standardized metadata models. Manuscripts are studied by scholars from diverse fields, each bringing a different perspective and way of interpreting the material. Paleographers and codicologists focus on physical and structural aspects, such as scripts, bindings, and foliation. Textual scholars analyze variations and editorial histories, while art historians examine iconography and illumination. Legal historians trace the evolution of legal thought and jurisprudence embedded in manuscript sources from Roman law through medieval commentaries [3,4,5,6,7]. This disciplinary fragmentation underscores the need for unified semantic frameworks that can bridge metadata heterogeneity and support cross-domain interoperability [8,9,10].

Although digitization initiatives have significantly improved access to manuscript collections and preserved fragile materials, they have also exposed critical limitations in data modeling and interoperability [11,12]. Metadata schemas, annotation strategies, and descriptive practices vary significantly between platforms, leading to inconsistencies that hinder semantic integration and collaborative discovery. Projects such as Progetto Imerio and Mosaico offer valuable manuscript collections. Still, their different data structures and lack of harmonized metadata highlight the need for ontology-driven approaches to ensure scholarly interoperability [13,14].

To address these challenges, semantic web technologies have emerged as powerful tools for structuring, linking, and querying manuscript metadata. By applying ontology-based modeling, complex relationships within and between manuscripts, including their textual, historical, and physical dimensions, can be represented. However, current ontologies often fall short of supporting the polymorphic, layered nature of manuscripts as they evolve across time and collections.

Building on prior work that developed the Medieval Manuscript Data Integration Ontology (MMDIO), this study proposes an extension of the framework by introducing new classes and object properties designed to capture manuscript-specific structures drawn from real-world use cases. In particular, the study focuses on harmonizing data from two significant manuscript collections, Mosaico and Progetto Imerio, through RDF transformation and the construction of a polymorphic knowledge graph. The goal is to support flexible SPARQL querying, semantic browsing, and scholarly integration across disciplines. This effort is operationalized through MELIORATE, a Linked Open Data platform, which supports the exploration of legal, textual, and historical dimensions of medieval manuscripts.

To guide this work, the study pursues the following objectives:

- Extend the MMDIO ontology by introducing new classes and object properties derived from the manuscript-specific use case in the Mosaico platform, enabling more granular and semantically rich modeling.
- Harmonize and transform heterogeneous metadata from Mosaico and Progetto Irnerio into RDF using the enhanced MMDIO framework, ensuring interoperability and semantic consistency across collections
- Construct the polymorphic nature of the knowledge graph and provide different views and facets of the data. It will support various types of queries and analyses, such as faceted search, semantic browsing, and visualization
- Development of the Linked Open Data portal to explore legal, textual, and historical aspects of manuscripts and to encourage collaboration and knowledge sharing among medieval manuscript scholars, researchers.

In summary, this study contributes to the evolving field of digital manuscript scholarship by advancing semantic modeling practices. Through the extension of the MMDIO ontology and the integration of heterogeneous manuscript data into a polymorphic knowledge graph, it offers a scalable and semantically consistent framework for the representation of complex textual and codicological structures. By aligning these efforts with Linked Open Data principles and implementing them in the MELIORATE platform, the study not only enhances interoperability across disparate collections but also fosters interdisciplinary engagement. This work provides a foundation for deeper scholarly investigation, flexible exploration of manuscript data, and collaborative knowledge-building in medieval studies and related disciplines.

The article is organized to guide the reader through the conceptual and technical advancements introduced. Section 2 reviews prior work on the semantic integration of medieval manuscripts, with emphasis on existing ontologies and data modeling challenges. Section 3 outlines the methodology, including the MeLON-driven development of the expanded MMDIO ontology and its evaluation using the FOCA methodology. Section 4 introduces the Polymorphic Knowledge Graph, which supports multi-dimensional representation of manuscript data. Section 5 presents the MELIORATE Linked Open Data platform, which operationalizes the ontology and PKG to deliver an integrated digital environment. Section 6 discusses the practical implications of this work for digital humanities and legal history research. Finally, Section 7 concludes the article with key findings and directions for future research.

## 2. Literature Review

The growing digitization of medieval manuscript collections has led to a wide range of semantic web and ontological frameworks aimed at improving accessibility, interoperability, and scholarly use. However, the complexity of these manuscripts, which is based on their varied historical, textual, and material characteristics, continues to pose challenges for existing data models. This literature review examines major projects and ontological approaches that have shaped the current landscape, identifying their contributions and limitations in the context of integrating heterogeneous medieval manuscript data.

### 2.1. *Europeana and the DM2E Model*

The [Europeana](#) digital library aggregates and provides unified access to cultural heritage resources from across Europe. To address the challenge of heterogeneous metadata schemas across contributing institutions, the Europeana Data Model (EDM) was developed to ensure metadata interoperability. However, EDM's general-purpose structure lacked the expressiveness required for rich manuscript metadata, particularly annotations, provenance, and foliation-specific details, leading to the development of the more specialized [DM2E](#) model.

The DM2E model extended EDM to support the specific needs of manuscript data integration. It incorporated elements from established ontologies such as Dublin Core, BIBO, FaBiO, and OAI-ORE, and introduced new properties and classes (e.g., `dm2e:scopeNote`, `dm2e:hasAnnotatableContents8`) to represent detailed manuscript structures and annotations. It also employed named graphs for provenance tracking and maintained schema-instance separation via distinct namespaces. Metadata from a variety of institutions, including libraries and archives using TEI, MARC21, METS/MODS, and custom schemas, was analyzed and harmonized. Manual mapping and iterative validation ensured semantic consistency. Tools like the D2R Server were used to convert legacy formats into RDF, supporting the creation of a Linked Open Data (LOD) knowledge graph. This graph was integrated into the Europeana platform, enabling advanced search, SPARQL queries, and semantic annotation [15]. While DM2E improved the representation of manuscript metadata, limitations remain in fully capturing historical, textual, and physical complexities.

### 2.2. *Mapping Manuscript Migrations (MMM)*

The [Mapping Manuscript Migrations \(MMM\)](#) project was designed to unify and harmonize diverse pre-modern manuscript databases by developing a comprehensive semantic data model based on CIDOC CRM and FRBRoo. The primary goal was to overcome the fragmentation of manuscript data across heterogeneous systems, such as Bibale, the Schoenberg Database of Manuscripts (SDBM), and the Medieval Manuscripts in Oxford Libraries (MMOL), which used incompatible metadata schemas and descriptive conventions. The metadata of the selected sources was analyzed, converted into RDF triples, and reconciled with or aligned to authority-linked open knowledge vocabularies, including the Virtual International Authority File (VIAF), and the Getty Thesaurus of Geographic Names (TGN). It yielded a harmonized knowledge graph of more than 20 million RDF triples, encapsulating interconnected information about manuscripts, persons, organizations, locations, and events.

The MMM infrastructure was built following the Sampo model, which supports collaborative data publishing and semantic integration. A dedicated Linked Data Service and a user-friendly web portal were developed to facilitate access to the knowledge graph. These tools enabled faceted browsing, complex SPARQL querying, and visual exploration of manuscript data without requiring technical expertise, thereby enhancing both scholarly research and public engagement. Although MMM significantly advanced the integration of provenance and ownership data, it exhibited certain limitations in its representation of textual, artistic, and codicological features of medieval manuscripts [16,17]. These aspects, which are crucial for a holistic understanding of manuscript heritage, were not the central focus of the project.

### 2.3. *Biblissima*

**Biblissima** is a major digital humanities initiative that provides unified access to distributed databases containing metadata on manuscripts and early printed books from the 8<sup>th</sup> to the 18<sup>th</sup> century. The project aggregates data across various disciplines such as codicology, iconography, and palaeography, but faces challenges in metadata harmonization due to the heterogeneity of partner database systems and schemas [18]. To overcome these barriers, Biblissima developed a unified ontology aligned with **CIDOC CRM** and **FRBRoo**, supporting semantic consistency and facilitating future integration with other CIDOC-compliant projects. A complementary XML model was also implemented to capture equivalent relationships across different partner datasets. Furthermore, metadata entities were linked to external authority files such as **VIAF**, **DBpedia**, and **GeoNames**, enhancing interoperability through Linked Open Data. To address image interoperability, the project adopted the International Image Interoperability Framework (IIIF) and integrated the Mirador viewer to enable high-resolution image comparison, annotation, and sharing within a collaborative research workspace.

Despite its achievements, Biblissima continues to face difficulties in fully harmonizing metadata due to significant variation in vocabularies, standards, and structural depth across collections. These limitations hinder the creation of a fully unified, semantically rich representation of manuscript content, especially in terms of artistic annotations, marginalia, and physical features.

### 2.4. *CIDOC CRM*

**CIDOC CRM** is an ISO standard ontology designed for cultural heritage documentation. It aims to standardize and integrate the documentation of cultural heritage information across museums, archives, and libraries. As an ISO standard (ISO 21127:2006), CIDOC CRM serves as a semantic framework to promote interoperability among heterogeneous datasets. It defines a rich semantic structure with 80 classes and 132 properties, covering temporal entities, persistent items, actors, and spatial-temporal relationships. Its extensible structure accommodates specialized modules such as **CRMdig**, **CRMarchaeo**, and **FRBRoo**, enabling adaptation to domain-specific needs [19].

One practical application of CIDOC CRM was demonstrated in the *COURAGE* project, which documented cultural opposition in former socialist countries. The *COURAGE* project applied CRM to model metadata from Cold War cultural heritage artifacts using automated RDF transformations and SHACL shape validation [20].

Despite its robustness, CIDOC CRM does not meet the diverse metadata requirements of medieval manuscript studies. Specifically, it lacks specific frameworks for capturing the textual, artistic, and contextual complexities of medieval codices. These constraints indicate that, while CIDOC CRM provides a solid framework for cultural heritage integration, adjustments or expansions may be required to satisfy the unique needs of manuscript scholarship.

### 2.5. *Hebrew Manuscripts*

This study presents an ontological data model specifically developed to organize and improve access to post-medieval Hebrew manuscripts, estimated to number between 70,000 and 80,000. These manuscripts are invaluable cultural records, yet their current digital representations are limited to basic library catalogs with constrained search and semantic capabilities. To address this, the authors propose an event-based ontology that treats each manuscript as a dynamic entity with a biographical trajectory encompassing creation, copying, and acquisition. The model was constructed through manual analysis of catalog metadata from institutions such as the National Library of Israel and the Séminaire Israélite de France. Key entities and relationships such as manuscript agents, historical figures, events, places, and life-cycle stages were extracted and encoded using RDF. The ontology reuses and aligns with standards such as CIDOC CRM, DM2E, BIBO, FOAF, BIO, CITO, and Geonames to ensure semantic interoperability.

The resulting ontology enables advanced semantic queries and automated reasoning, and supports integration with authority files via tools like `owl:sameAs`. Validation was performed through expert review, internal consistency checks, and external comparisons. Although the model succeeds in representing complex relationships and manuscript biographies, it remains limited in handling medieval manuscripts' diverse features, such as textual structure, materiality, and visual annotations [21]. Future work involves enhancing scalability, automating data extraction, and incorporating paleographic and codicological details.

### 2.6. *King Henry III*

This paper describes the development of the FRH3 ontology for the [Henry III Fine Roll Project](#), a collaborative effort between King's College London and the UK National Archives. The fine rolls are historical records documenting monetary transactions and favors granted by King Henry III of England from 1216 to 1248. The goal was to digitally represent these complex documents for web access. Initially marked up using TEI-compliant XML to reflect their physical, spatial, and temporal features, the data was transformed into RDF/OWL. The FRH3 ontology reused established ontologies such as [CIDOC CRM](#), [Dublin Core](#), SKOS, Geo, and Time. The key entities modeled included Authority, Factoid, and Gender.

Transformation from XML to RDF was accomplished using XSLT scripts, enabling the integration of structured historical data. Validation was carried out using SHACL and SPARQL to ensure consistency with ontology requirements. This allowed users to infer information and perform advanced queries beyond the original textual markup. Despite its success, the project encountered challenges, including difficulty synchronizing updates between XML and RDF, and the limited reusability of the ontology outside the

context of the fine rolls [22]. Nonetheless, FRH3 enabled accurate modeling of historical legal documents using semantic web standards.

### 2.7. *Historical Malay Manuscripts*

This case study presents the development of an event-based ontology to model the content of the historical Malay manuscript *Sulalatus Salatin* (Sejarah Melayu), published by the University of Malaya Digital Library. Due to its rich historical content and availability in digitized and transliterated forms, the manuscript served as a representative example for constructing a structured semantic model capturing events, actors, places, times, and objects. The ontology development followed the 101 Ontology Development Methodology, which included phases such as domain scoping, manual data extraction, ontology creation in OWL using Protégé, and enrichment through external linking to [DBpedia](#) via owl:sameAs. This allowed entities such as historical figures (e.g., Tun Perak) to be contextualized with structured external data. The model was validated using SPARQL queries aligned with competency questions drawn from the manuscript.

The primary objective was to surface hidden historical insights through structured event modeling. The ontology addressed challenges in metadata expressiveness that are common with standards like Dublin Core, while supporting semantic querying and reuse. Nonetheless, issues such as named entity disambiguation, historical language complexity, and poor manuscript legibility remain open problems. This work contributes to building a reusable knowledge base for Malay manuscripts and offers a semantic foundation for studying Malay history and culture [23]. Nonetheless, the event-based ontology significantly improves access to and understanding of Malay cultural heritage and supports ongoing refinement for broader applicability.

### 2.8. *Saussure's Manuscripts*

This case study presents a semantic framework developed for the study and digital publication of the historical scientific manuscripts of Ferdinand de Saussure. Given the heterogeneity of content, spanning various time periods and terminologies, the project addressed the need for explicit semantic representation of text meanings and reasoning over temporal information. Among the approximately 50,000 written pages, only 5,000 had been transcribed, highlighting the scale and complexity of the task.

A comprehensive analysis conducted with domain experts identified four core needs: manuscript access, content understanding, date inference, and publication. This led to the development of a multi-faceted ontology that integrates manuscript representations, transcriptions, contextual and terminological knowledge, and bibliographic metadata. Existing models such as CIDOC-CRM and SEM were deemed insufficient, prompting the creation of a new Semantic Model for Historical Scientific Manuscripts, which reuses components from CIDOC-CRM, SEM, SKOS, and OWL.

The model facilitated the semantic linking of transcriptions to bibliographic references and domain-specific concepts, enabling enhanced indexing and retrieval. Contextual analysis and semantic scoring of terms were used to resolve terminological variations over time. Temporal reasoning mechanisms were implemented to infer dates and relationships using SPARQL and rule-based constructs, supporting chronology-sensitive manuscript interpretation. Despite these advances, the project faced challenges in seman-

tic indexing automation, user-friendly temporal rule formulation, and system scalability [24]. Nonetheless, the model provides a robust infrastructure for digital humanities research and lays the groundwork for integrating rich scholarly metadata into semantic systems.

### 2.9. *Progetto IRNERIO*

Progetto IRNERIO is a large-scale digital humanities initiative aimed at cataloguing and providing access to digitized medieval legal manuscripts, primarily focusing on Roman and Canon law produced in Bologna between the X-XI centuries to the beginning of the XVI century [25]. Led by the University of Bologna and CLUEB publishing, the project has made over 138,000 digitized folios available online via the CIRSIFID research center. The platform supports structured access to codices with rich metadata, gloss annotations, and critical text, enabling high-resolution exploration and scholarly study.

To model these resources, the [Medieval Manuscript Ontology \(MeMO\)](#) was developed using the SAMOD methodology. The ontology integrates existing vocabularies such as CiTO, DCTerms, FaBiO, FOAF, FRBRcore, and TVC, and is aligned with the FRBR framework and FaBiO to distinguish between textual expressions and physical manifestations. Manuscripts, glosses, and texts are modeled as `fabio:Expression`, while codices and folios are treated as `fabio:Manifestation`. MeMO also supports structural metadata, including foliation, citations, and authorship, while enabling multi-perspective analysis of manuscripts and glosses. The ontology's modular structure evolved over six SAMOD iterations, modeling core dimensions such as gloss annotations, foliation, citations, and codex metadata. It provides the semantic infrastructure for navigating scholarly narratives, supporting traditional manuscript studies and digital analysis alike.

Despite its contributions, several modeling challenges remain. These include inconsistent temporal-spatial metadata, unstructured bibliographic references, ambiguous dimension data, disambiguation of author and glossator identities, and the need to support author name variants. Additionally, a robust RDF transformation from SQL databases must ensure metadata integrity and semantic richness. MeMO demonstrates how domain-specific ontologies can transform manuscript accessibility and scholarly interaction. With further refinement, the model holds potential as a reusable semantic infrastructure for other manuscript collections and legal-historical research.

### 2.10. *Mosaico Platform*

The [Mosaico Platform](#) is a digital library project aimed at aggregating, describing, and providing access to a vast corpus of medieval legal manuscripts [26]. It enables high-resolution viewing of codices and offers enriched scholarly data, including bibliographies, transcriptions, manuscript descriptions, and biographical materials. The platform not only digitizes manuscripts but also traces their academic reception and historiography from the 19th century onwards.

Mosaico comprises two main repositories: one for XML-based manuscript descriptions enriched with TEI P5 metadata, and another for storing manuscript images. Contributions from renowned legal historians, including Bellomo, Bertram, and others, add depth and scholarly commentary to the content. The portal supports advanced naviga-

tion through structured catalogs, allowing users to browse, search, and access manuscript materials interactively.

Despite its robust digital infrastructure, Mosaico currently lacks a dedicated ontology for semantic modeling. This limits the ability of the platform to process complex queries, cross-referencing, and integration with external resources. Compared to the Innerio platform, which is founded on the MeMO ontology for formally represented metadata, Mosaico relies on general TEI-based descriptions that are not necessarily capable of reflecting medieval manuscripts' complex relationships and characteristics. Filling this gap by the development or utilization of a formal ontology would enlarge the platform's semantic scope, enabling richer scholarly analysis, improved data integration, and additional potential for digital research in medieval legal history.

### *2.11. Medieval Manuscript Data Integration Ontology (MMDIO)*

The Medieval Manuscript Data Integration Ontology ([MMDIO](#)), initially developed to address the integration of heterogeneous metadata from platforms like Progetto Innerio and Mosaico, provides a unified semantic model for representing diverse legal manuscript collections. Built upon the MeMO framework and aligned with FRBR and FaBiO principles, MMDIO offers a modular and extensible architecture designed to capture key manuscript components such as folio structures, gloss annotations, codicological features, citation networks, and textual hierarchies. Its layered design enables the mapping of both physical and intellectual aspects of manuscripts, allowing for fine-grained modeling of content across historical, textual, and legal domains.

The ontology incorporates existing vocabularies such as DCTerms, CiTO, and FOAF to ensure compatibility with Linked Open Data (LOD) ecosystems and to promote reuse across cultural heritage institutions. One of MMDIO's distinguishing strengths lies in its capacity to model multiple levels of bibliographic description from physical manifestations like folios and bindings to abstract conceptual works, while also supporting the representation of gloss annotations and manuscript-specific scholarly practices [14]. Although the initial implementation effectively demonstrated the feasibility of ontology-based harmonization across collections, it was deliberately designed to be extensible and evolvable. This allows for the incorporation of additional classes, object properties, and context-aware semantic constraints to support emerging research needs.

### *2.12. Current Research Direction*

To address these gaps, the current study extends the Medieval Manuscript Data Integration Ontology ([MMDIO](#)). The core MMDIO structure, introduced in previous work, provides class hierarchies and properties to model textual, artistic, and codicological dimensions based on data from both the Innerio and Mosaico platforms. Building upon this foundation, the present study further enhances the ontology by incorporating an additional use case and extending the ontology with new classes and properties for broader application.

The framework supports the creation of polymorphic knowledge graphs, enabling multiple semantic views of the same data and robust querying through SPARQL. This forms the foundation for the user-friendly web platform for semantic browsing, scholarly collaboration, and knowledge sharing. Building upon previous efforts and introduc-

ing targeted extensions, this research aims to provide a scalable, semantically rich, and interoperable framework suitable for diverse medieval manuscript collections.

### 3. Methodology

The methodological framework for this study builds on previous work that involved the integration of heterogeneous metadata from the Progetto Irnerio and Mosaico platforms into the Medieval Manuscript Data Integration Ontology (MMDIO). This ontology was originally developed using the MeLON methodology. The previous work focused on harmonizing metadata related to the Irnerio platform and the Authenticum collection from the Mosaico portal.

In the current phase of research, the methodology is extended to incorporate an additional use case from other collections within the Mosaico portal. This required enhancements to the existing ontology to support new data structures and complex relationships found in these expanded datasets.

#### 3.1. Content Review of Mosaico Collections

A detailed review of the Mosaico digital library was conducted to understand the structure and availability of metadata across its collections. Mosaico, while rich in content, lacks a formalized ontology and presents information inconsistently, which complicates semantic modeling and metadata integration.

The portal includes several digitized collections, notably: *Arbores*, *Archiginnasio*, *Goffredo*, *Monaco*, *Montecassino*, *Olomouc*, *Roffredo Beneventano*, and *Satana*. This study excluded the previously modeled *Authenticum* and focused on analyzing the remaining collections to identify metadata elements and structural organization.

Each collection presents metadata in different forms. For instance:

- **Montecassino** (edited by Martin Bartram) includes five manuscripts, each corresponding to specific folios. It features bibliographic references and two signatures from the author Goffredus Tranensis cited in the manuscript facsimiles and in the “Premessa” section. It also includes references to specific folios and correction sections with transcriptions.
- **Olomouc** (edited by Lucia Sorrenti) is a digital edition of *Tractatus quaestionum*, attributed to Julianus de Sessa. It contains descriptive sections such as “Introduzione e descrizione” and “Contenuto e struttura dell’opera,” which refer to individual folios, cite external works, and present 12 manuscripts with detailed structural metadata.
- **Roffredo Beneventano** includes comprehensive codex metadata like size, material, century, glosses, and annotations. The manuscript XVII.A.10, held in the National Museum of Prague, is a key example, described in detail including dimensions, foliation, and physical characteristics.
- **Satana** lists 43 codices in the “Manoscritti” section. Some include data such as incipit, explicit, origin, material, size, and bibliographic references, while others provide only minimal details.
- **Arbores** is structured around thematic divisions and outlines multiple versions of the *Arbor actionum* tradition. However, internal segmentation and manuscript-level data are described in prose rather than semantically modeled.

There is no further information available regarding the books *Archiginnasio*, *Goffredo*, and *Monaco* on the portal.

This content review confirmed the presence of rich but uneven metadata, reinforcing the need for structured ontology-based integration to ensure interoperability and semantic precision across collections.

### 3.1.1. Data Analysis for Identifying Modeling Gaps

Building on the content review, a detailed semantic analysis was conducted to identify gaps in the existing MMDIO ontology when applied to the new Mosaico collections. While MMDIO successfully handled the *Authenticum* use case, the additional collections introduced challenges related to structural segmentation, physical foliation, and authorship attribution.

Key modeling deficiencies included:

- **Montecassino:** The collection features digital representations of author signatures (e.g., Goffredus Tranensis) cited in bibliographic sections but could not be formally linked in the ontology. A property for authorship attribution was missing.
- **Roffredo Beneventano:** Manuscripts exhibit deeply nested internal divisions that exceeded MMDIO's original section granularity, requiring the introduction of hierarchical sub-structures.
- **Olomouc and others:** Several collections refer to specific folios within the manuscript, but the ontology lacked a semantic bridge between logical units like `doco:Section` and physical units like `memo:Folio`.
- **Arbores and Satana:** No new modeling requirements were identified for these collections, as the existing ontology already supported their structures.

These observations directly informed the design of new ontology components, including dedicated classes and properties for capturing internal manuscript structure, author attribution, and foliation references. These additions ensure better alignment between textual hierarchies and their physical or bibliographic counterparts.

While the original *Authenticum* use case focused primarily on core manuscript metadata, the new collections demanded more targeted extensions to model fine-grained structure, physical references, and scholarly provenance relationships.

### 3.2. Application of the MeLON Methodology in Ontology Extension

The ontology extension was guided by the MeLON methodology, ensuring systematic and rigorous development across all modeling phases. Key steps included goal definition, use case formulation, evaluation indicators, state-of-the-art analysis, ontology modeling, and iterative testing.

Evaluation criteria from the original phase, such as completeness, coherence, and usability, were reused to assess the quality of the extensions. The modeling process relied on OWL and UML tools such as Protégé and Graffoo, maintaining methodological continuity with the initial MMDIO development [14].

The present study builds upon the foundational MMDIO ontology developed using the MeLON (Methodology for Legal Ontology) framework. While the initial ontology focused primarily on integrating data from the Progetto Innerio and Authenticum collection in Mosaico, the current study re-applies the MeLON Methodology to guide the

extension of MMDIO for additional manuscript collections and more complex use cases [27].

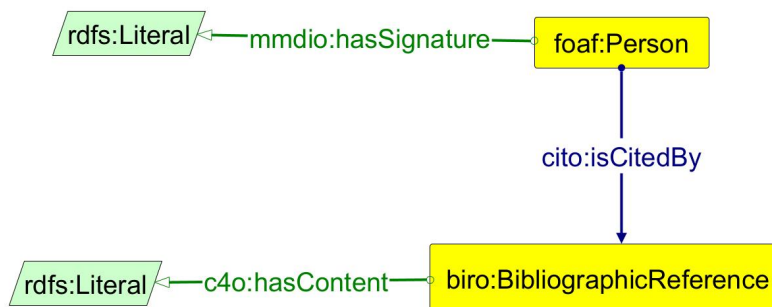
### 3.3. Ontology Goals and New Use Case

#### 3.3.1. Use Case 4: Streamlining Montecassino, Olomouc, Arbores, Roffredo Beneventano, Satana collections in the Mosaico digital library

The extension of the **MMDIO** ontology was driven by real-world requirements emerging from **Use Case 4**, which focuses on the semantic modeling of complex manuscript collections in the Mosaico digital library, specifically, Montecassino, Olomouc, Roffredo Beneventano, Satana, and Arbores. These collections required structural refinements described in the preceding ontology extension section.

In particular, the **Montecassino** collection presented several unique modeling challenges:

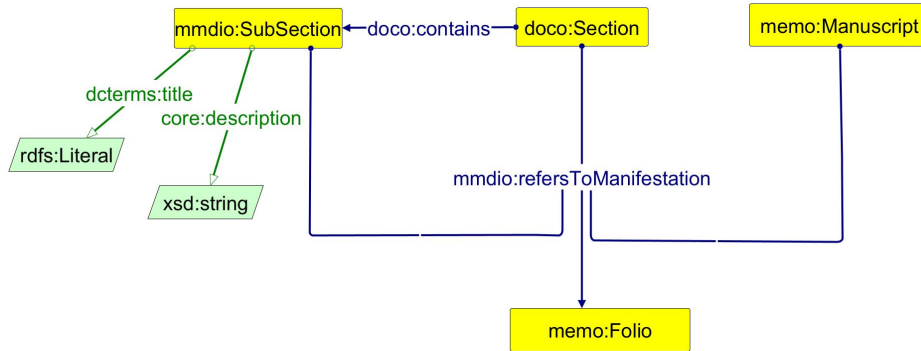
- The digital edition curated by Martin Bartram references a historical figure, *Goffredus Tranensis*, whose signature is captured in the physical manuscript. To represent this, a new object property `mmdio:hasSignature` was introduced to link `fabio:Book` to a `foaf:Person`.
- Bibliographic citations of this person are semantically linked using `cito:isCitedBy`, connecting individuals with `biro:BibliographicReference` entries to support scholarly provenance tracking as illustrated in Figure 1.



**Figure 1.** The diagram illustrates authorship attribution and scholarly citation using `mmdio:hasSignature` and `cito:isCitedBy`.

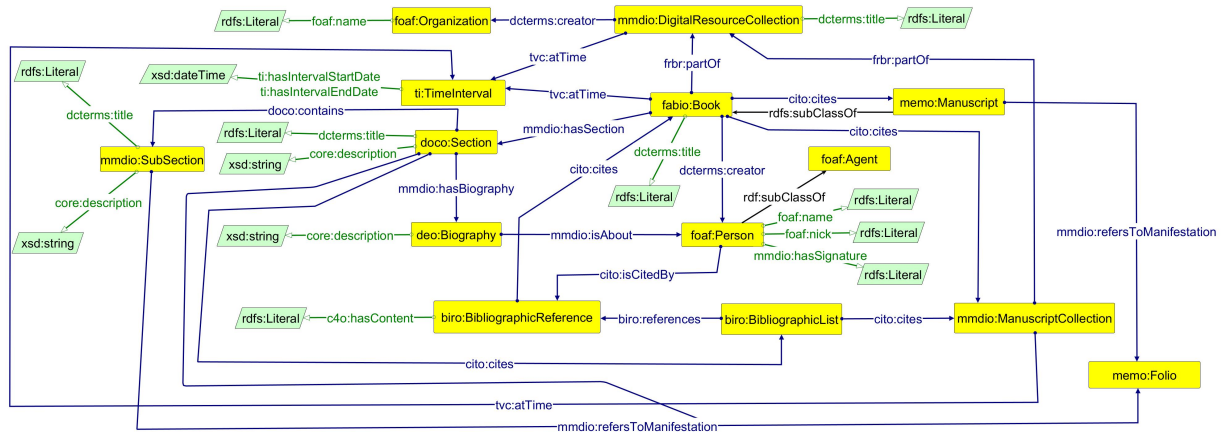
- Montecassino and other manuscripts such as Olomouc and Roffredo Beneventano exhibit internal subdivisions within sections. To handle this, the class `mmdio:SubSection` was added, and related to `doco:Section` via the existing `doco:contains` property. These subdivisions are annotated with `dcterms:title` and `core:description` to capture their semantic and descriptive content.
- Furthermore, a new object property, `mmdio:referstoManifestation`, allows `doco:Section`, `mmdio:SubSection`, and `memo:Manuscript` entities to be explicitly connected to corresponding `memo:Folio` instances. This allows accurate

alignment between logical text structures and their physical manifestations in folios, as shown in Figure 2.



**Figure 2.** The Graffoo diagram demonstrates the textual subdivisions and foliation using `mmdio:SubSection` and `mmdio:referstoManifestation`.

These additions were guided by the practical demands of the new use cases, ensuring that MMDIO remains extensible and aligned with real-world scholarly workflows. To summarize the semantic structure implemented in response to Use Case 4, Figure 3 presents a fully integrated view of the ontology extensions.



**Figure 3.** The Graffoo diagram showing the integration of structural, material, and authorship features introduced for the Mosaic collections.

The diagrams in Figures 1 and 2 illustrate the modeling elements in detail, and supplement the overall integration view in Figure 3. These enhancements directly illustrate the extensibility of MMDIO as a robust, interoperable semantic model designed to meet evolving research demands by aligning new ontology components with the structural complexity and scholarly expectations embedded in these manuscript collections.

### 3.4. Reuse of Evaluation Criteria and Modeling Practices

The MeLON methodology also enabled the reuse of evaluation indicators (e.g., coherence, completeness, usability, and effectiveness) to assess the accuracy and performance of the extended ontology. Tools like Graffoo and Protégé have again been used for UML and OWL modelling, ensuring compatibility with the original development process.

Sample data from the Mosaico collections was extracted and mapped into RDF using the expanded ontology. The datasets were uploaded to GraphDB and tested through SPARQL queries, which reflected scholarly use cases such as getting gloss positions by folio or finding subsections concerning certain manuscript parts.

By applying the MeLON methodology to this extended phase of work, the research maintained the integrity and further scalability of the MMDIO ontology to enhance its applicability with wider cultural heritage data. The extended ontology is, therefore, a more subtle and comprehensive semantic framework capable of modeling the various manuscript structures, facilitating scholars' inquiry, and enabling cross-platform integration between Progetto Innerio and Mosaico.

### 3.5. Evaluation of the MMDIO Ontology Using FOCA Methodology

In this article, the Medieval Manuscript Data Integration Ontology (MMDIO) is evaluated using the FOCA methodology (Framework for Ontology Checking and Assessment) [28]. Unlike the initial development of MMDIO, which focused on modeling and integration, the current phase emphasizes quality assessment using FOCA, a structured framework designed to evaluate ontologies based on the *Goal-Question-Metric* (GQM) approach.

FOCA is grounded in the five roles of knowledge representation proposed by Davis in 1993: *Substitute*, *Ontological Commitments*, *Intelligent Reasoning*, *Efficient Computation*, and *Human Expression*. For each of these goals, targeted questions are answered and scored, followed by a quantitative evaluation through a statistical model.

#### 3.5.1. Ontology Type Verification

The MMDIO ontology represents a domain ontology, as it models the various aspects of medieval manuscripts, textual, historical, physical, and artistic, across platforms like *Progetto Innerio* and *Mosaico*. Supports tasks such as metadata alignment, semantic linking, and digital preservation in the context of legal history and cultural heritage.

#### 3.5.2. Goal-Question Evaluation

The following table presents the results of the FOCA evaluation, grouped by knowledge representation goals:

**Table 1.** FOCA Evaluation Results for MMDIO

Goal	Question Summary	Score (%)
<b>Substitute</b>	Competency questions defined and answered; ontology reuses existing models	100
<b>Ontological Commitments</b>	Moderate abstraction, coherent property usage	87.5
<b>Intelligent Reasoning</b>	No contradictory or redundant axioms found	100
<b>Efficient Computation</b>	No modeling errors; reasoner executes efficiently	100
<b>Human Expression</b>	Concepts well written, consistently annotated, and fully documented	100

### 3.5.3. Total Quality Calculation

The FOCA quality score is computed using a beta regression model based on the weighted sum of evaluation goals. The following formula is used:

$$\hat{\mu}_i = \frac{\exp\{-0.44+0.03(\text{CovS} \times \text{Sb})_i+0.02(\text{CovC} \times \text{Co})_i+0.01(\text{CovR} \times \text{Re})_i+0.02(\text{CovCp} \times \text{Cp})_i-0.66\text{LExp}_i-25(0.1 \times \text{NI})_i\}}{1+\exp\{-0.44+0.03(\text{CovS} \times \text{Sb})_i+0.02(\text{CovC} \times \text{Co})_i+0.01(\text{CovR} \times \text{Re})_i+0.02(\text{CovCp} \times \text{Cp})_i-0.66\text{LExp}_i-25(0.1 \times \text{NI})_i\}}$$

Where:

- **CovS, CovC, CovR, CovCp**: Mean scores for each goal
- **Sb, Co, Re, Cp**: Set to 1 for total quality score
- **LExp**: Evaluator experience (1 = experienced)
- **NI**: Unanswered question flag (0 = all answered)

Substituting the evaluation results:

$$\hat{\mu}_i = \frac{\exp\{3.235\}}{1 + \exp\{3.235\}} \approx \frac{25.43}{26.43} \approx 0.962$$

### 3.5.4. Result and Interpretation

The final FOCA quality score for MMDIO is:

0.962

This score reflects excellent overall quality, confirming that MMDIO is a well-structured, coherent, and semantically rich ontology. It supports complex data modeling, cross-platform integration, and scholarly research in the domain of legal and cultural manuscript studies.

The inclusion of FOCA in this phase of the investigation provides formal validation of the structure and usability of MMDIO. It complements the MeLON methodology by offering a quantifiable quality check, ensuring that the ontology not only aligns with domain needs but is also logically robust and computationally efficient. As MMDIO continues to evolve, this evaluation framework can be reused to guide iterative improvements.

## 4. Polymorphic Knowledge Graph Construction and Integration

Building on the foundational MMDIO ontology and earlier work involving RDF-based sample dataset extraction from the *Progetto Irnerio* and *Authenticum* collections in Mosaico, this phase introduces the complete integration of heterogeneous data into a Polymorphic Knowledge Graph (PKG). The PKG provides a dynamic and context-aware representation of medieval manuscripts, supporting complex relationships between entities across space, time, and format.

### 4.1. Concept and Purpose

The PKG enables the same entity, such as a manuscript or book, to exist in multiple roles and relationships based on its historical, spatial, or functional context. This flexibility is essential for modelling medieval legal manuscripts, which often appear in varying forms, editions, and collections across different centuries. Unlike the earlier RDF models focused on isolated use cases, the PKG supports full-scale integration of all available data from both portals and represents their connections in a unified semantic structure.

## 4.2. Design Characteristics

The PKG is designed to exhibit the following features:

- **Unified Multi-Platform Representation:** Integrates full metadata from both Innerio and Mosaico platforms under the MMDIO ontology, bridging different data models through a single, interoperable framework.
- **Polymorphic Structure:** Entities may assume multiple roles depending on query context (e.g., a manuscript as a historical artefact, a digital edition, or a cited scholarly reference).
- **Scalability and Evolution:** Supports seamless extension with new manuscripts, glosses, folios, and annotations without disrupting the graph structure.
- **Cross-Contextual Querying:** Facilitates faceted search and retrieval across time, location, codex structure, textual roles, and scholarly references.

The following components play a central role in the design and structure of the graph.

## 4.3. Dynamic Representation

In a PKG, each instance (e.g., a specific manuscript, or book) is allowed to exist in multiple contexts with different relationships. This is particularly useful for medieval manuscripts like *Authenticum*, which can belong to multiple collections over time and have different interpretations or annotations added throughout history.

For example, the same manuscript may appear in different collections, depending on its period and location:

- *Authenticum* in the 12th century is part of the Angers, BM 333 collection in Italy.
- *Authenticum* in the 13th century is part of the Bruxelles, 12084 collection in France.

The dynamic nature of the PKG allows *Authenticum* to remain a single instance while participating in multiple relationships, depending on the context. The period and location associated with each relationship define how *Authenticum* is understood in different historical contexts.

## 4.4. Multi-Contextual Role Assignment

Another important feature of a PKG is the ability to assign multiple roles to an instance depending on the query or context. In the case of medieval manuscripts, *Authenticum* can have different representations depending on whether it is viewed as:

- A manuscript in historical collections.
- A digital version created by a scholar in the 21st century.
- A physical book published in 1846.

This flexible role assignment ensures that the graph can return the appropriate version of *Authenticum* based on the specific needs of the query. For example:

- When querying for historical manuscripts, *Authenticum* will appear as a manuscript belonging to the appropriate collection.
- When querying for digital resources, *Authenticum* will appear as a digital object created in 2010.

#### 4.5. Data Extraction for Full Integration

To enable full semantic integration, comprehensive manuscript metadata were extracted from both the Mosaico and Innerio platforms using a combination of manual and automated methods.

*Mosaico Platform.* Metadata from 28 collections under *Authenticum*, along with other books such as *Montecassino*, *Olomouc*, *Satana*, *Roffredo Beneventano*, and *Arbores*, was manually gathered due to the absence of a structured export interface. In books with hierarchical organization, such as *Montecassino*, *Olomouc*, and *Roffredo Beneventano* sections and subsections referred to specific folios. Custom Python scripts were used to extract and assign unique identifiers to these folios, linking them to their respective textual subdivisions.

*Innerio Platform* A combination of manual and automated extraction techniques was applied to process all codices in the Innerio platform. Python scripts were developed to identify and classify individual folio images, assigning unique identifiers that distinguished between recto (r) and verso (v) sides. For example, a folio might be represented as `folio#page1r` for the recto side of the first page and `folio#page1v` for the verso side. These identifiers were used to ensure consistency across the RDF model and digital image references.

In addition, temporal and spatial metadata were extracted from the Century field in each codex entry. For instance, an entry such as `Century: XV-XVI (Francia)` was parsed using regular expressions and string processing to separate the temporal and spatial components. The term “Francia” was extracted as the geographic reference, while the Roman numeral centuries were converted into precise ISO-standard datetime intervals. For example, XV century was translated into a start date of `1500-01-01T00:00:00` and an end date of `1599-12-31T23:59:59`. This transformation ensures compatibility with ontology development environments such as Protégé, enabling accurate time-based reasoning and querying within the knowledge graph.

Descriptive metadata, including script style, initials, rubrication, and number of writing hands from the codex records. These elements were categorized based on available descriptions in the source platform and integrated into the knowledge graph to enhance semantic richness.

*RDF Transformation.* All cleaned and enriched metadata were first loaded into structured pandas DataFrames from Excel sources. These tabular records were then mapped to RDF triples using `rdflib`, with each row converted into subject–predicate–object structures. The resulting RDF graph was serialized into standard formats and loaded into GraphDB to support SPARQL-based querying and reasoning.

#### 4.6. GraphDB Loading and Semantic Integration

To support semantic querying across heterogeneous medieval manuscript datasets, all RDF outputs were loaded into GraphDB, a scalable triple store that supports contextual reasoning and SPARQL 1.1. GraphDB was selected for its ability to handle complex ontologies, perform inference, and merge datasets from multiple sources without data duplication.

The Polymorphic Knowledge Graph (PKG) was incrementally constructed by integrating RDF data from the Mosaico and Inderio portals. As new manuscript metadata or annotations were added, GraphDB's merging logic maintained referential consistency, allowing the graph to evolve without requiring structural modifications to the ontology.

A series of SPARQL queries was implemented to validate the knowledge graph's accuracy and flexibility. These queries were tested:

- Retrieval of manuscripts by historical period and geographical region.
- Cross-platform entity relationships (e.g., *Authenticum* in Mosaico referencing codices in Inderio).
- Semantic traceability from books to sections, subsections, and physical folios.
- Annotation-level queries (e.g., gloss placements, bibliographic references).

The query results were visualized to confirm data reliability. Each use case was thoroughly tested by iterating over key node relationships, ensuring that the PKG accurately reflected both textual and material aspects of the manuscripts. This integration illustrates the graph's capacity to support faceted search, semantic navigation, and multi-dimensional analysis of cultural heritage data.

#### 4.7. Validation and Visualization

To validate the effectiveness of the PKG, each use case previously defined was re-executed using live queries over the integrated graph. The results were visualized using knowledge graph interfaces and query-driven dashboards. Each data node (e.g., codex, folio, gloss) was inspected to confirm consistency and correctness against the original source platforms.

The Polymorphic Knowledge Graph developed in this phase transforms the earlier RDF-based pilot work into a dynamic, comprehensive, and scalable knowledge infrastructure. It supports semantic enrichment of manuscript data, cross-collection comparisons, historical reconstructions, and future-linked data integrations. By encapsulating heterogeneous data sources into a polymorphic structure, this graph serves as a foundational asset for ongoing digital humanities scholarship in the legal manuscript domain.

#### 4.8. Graph Visualization of PKG Dynamics

The visualization in Figure 4 illustrates the concept of dynamic representation within the Polymorphic Knowledge Graph (PKG). A single manuscript instance, *Authenticum*, is represented once in the graph but is connected to multiple historical contexts through different relationships.

As shown, *Authenticum* is linked to two distinct codices: Angers, BM 333 and Bruxelles, 12084. These represent different manuscript collections in different centuries and locations. Temporal relationships (`atTime`) and spatial relationships (`hasLocation`, `previousLocation`) express the manuscript's presence in the 12<sup>th</sup> century in Angers and in the 13<sup>th</sup> century in Bruxelles.



```

SELECT ?title ?creatorName
WHERE {
  ?digitalResourceCollection a mmdio:DigitalResourceCollection ;
                             dcterms:title ?title ;
                             dcterms:creator ?creator .
  ?creator a foaf:Organization ;
           foaf:name ?creatorName .
}

```

Listing 1: A SPARQL query returning digital collections and their respective creators. The result in Figure 5 reveals two collections *Mosaico* and *Irnerio*, both created by *CIRSFID, University of Bologna*, confirming that the unified RDF model supports provenance and authorship tracking across repositories.

title	creatorName
"irnerio"	"CIRSFID, University of Bologna"
"mosaico"	"CIRSFID, University of Bologna"

Figure 5. Titles of digital collections and their respective creators retrieved using a SPARQL query.

A follow-up query retrieves manuscripts from both platforms by filtering collection titles:

```

SELECT ?manuscript ?manuscriptTitle ?manuscriptCollectionTitle
WHERE {
  ?manuscript a memo:Manuscript ;
             dcterms:title ?manuscriptTitle ;
             frbr:partOf ?manuscriptCollection .
  ?manuscriptCollection a mmdio:DigitalResourceCollection ;
                       dcterms:title ?manuscriptCollectionTitle .
  FILTER (?manuscriptCollectionTitle IN ("mosaico", "irnerio"))
}

```

Listing 2: A SPARQL query that returns manuscripts from “Mosaico” and “Irnerio”.

manuscript	manuscriptTitle	manuscriptCollectionTitle
<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-11">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-11</a>	Liber XI	mosaico
<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-8">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-8</a>	Liber VIII	mosaico
<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-7">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-7</a>	Liber VII	mosaico
<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-12">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-12</a>	Liber XII	mosaico
<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-10">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-10</a>	Liber X	mosaico
<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-5">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/olomuc/manuscript-collection/main/liber-5</a>	Liber V	mosaico

Figure 6. Manuscripts retrieved from both “Mosaico” and “Irnerio” digital collections.

These queries illustrate how heterogeneous metadata is seamlessly merged and queried, facilitating scholarly exploration across institutional silos.

#### 4.11. *Authenticum* Collection Metadata Exploration

It focuses on the *Authenticum* collection within the Mosaico platform. A SPARQL query (Listing 3) was designed to retrieve all metadata associated with the book titled “Authenticum”. The output, shown in Figure 7, displays all RDF triples connected to the book instance, including its author, title, sections, manuscript references, and collection affiliations.

```
SELECT ?property ?value
WHERE {
  ?book a fabio:Book;
  dcterms:title "Authenticum";
  ?property ?value.
}
```

Listing 3: SPARQL query to retrieve all metadata related to the “Authenticum” book.

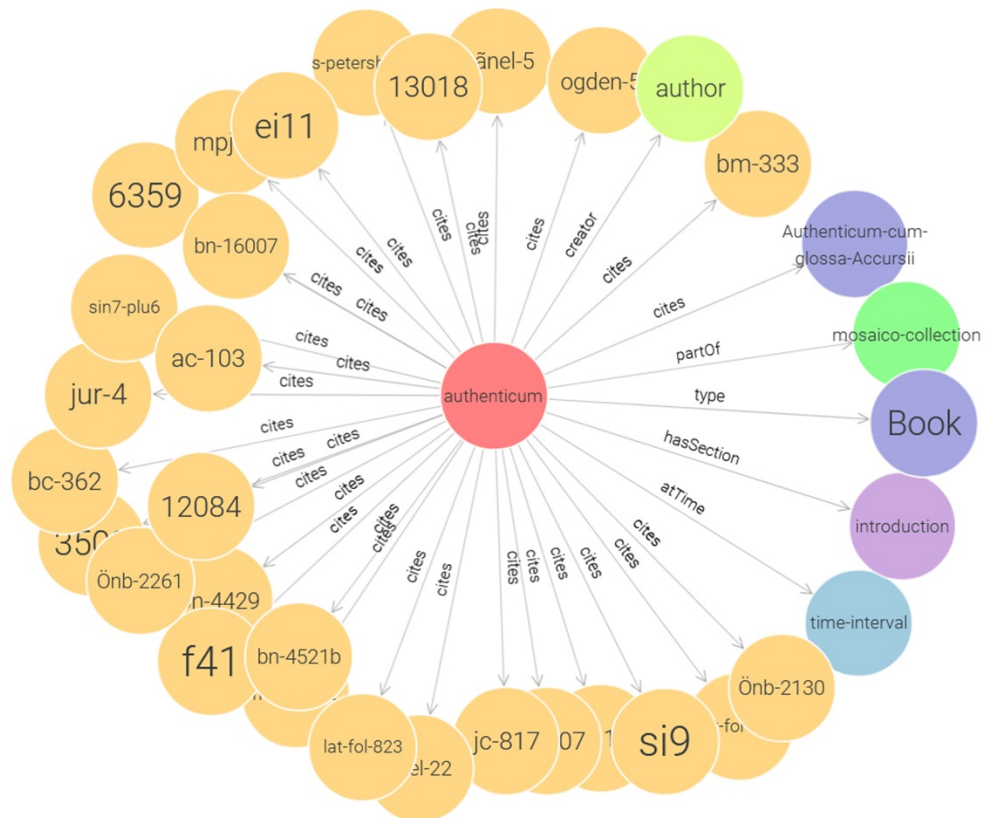


Figure 7. Graphical representation of the metadata associated with the “Authenticum” book.

Figure 7 presents the query results in a graphical format. The book *Authenticum* is linked to 28 manuscript collections, such as ‘ei11’ and ‘jur-8’, which is an integral part of the Mosaico collection. In addition, the book’s author is ‘Cristina Vano’, which enhances the context and credibility of the work. It includes sections like ‘Introduction’ and is linked to a time interval which signifies the period when this digitized book was created. In addition, the book is connected to the Mosaico collection, a comprehensive digital resource collection that includes various related works. This establishes ‘Authenticum’ within a wider academic context.

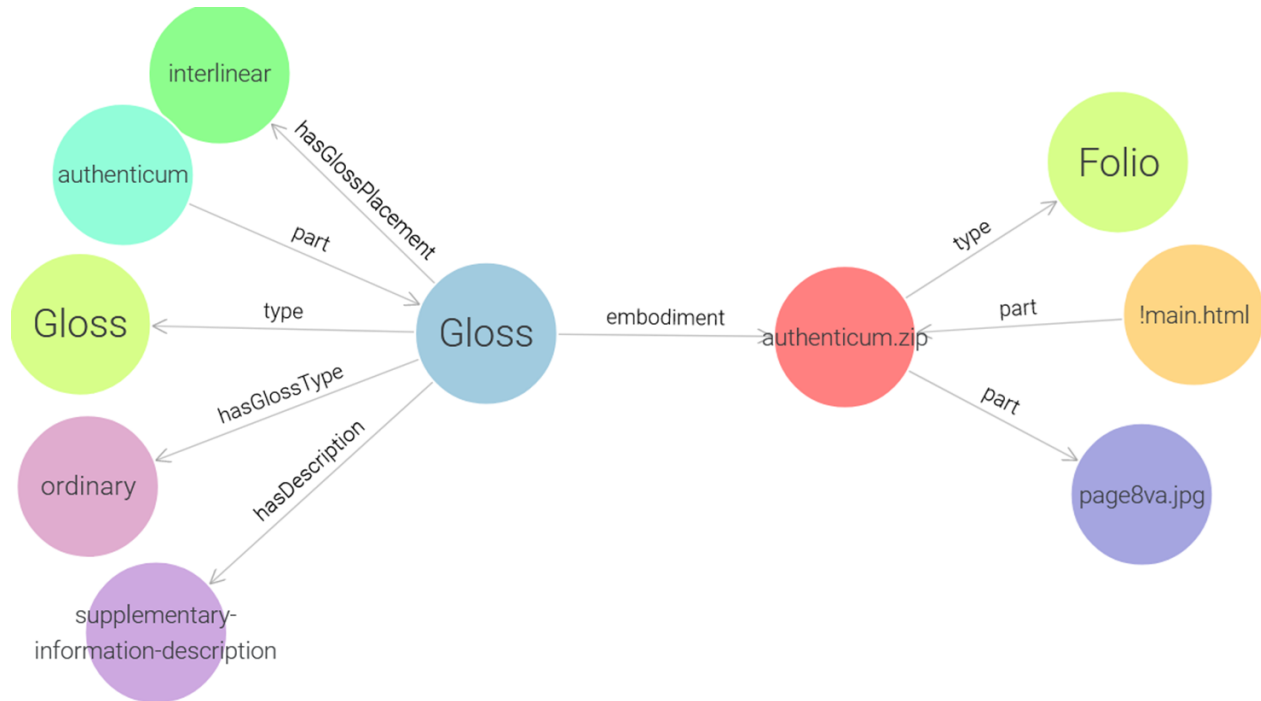
#### 4.12. Gloss Annotations Across Folios

Medieval manuscripts frequently include glosses, or scholarly annotations, which are critical for interpreting legal and philosophical texts. A query (Listing 4) was executed to retrieve glosses of type *ordinary* and *extravagant*, focusing on their placement (*inter-linear* or *marginal*) within the text.

```
SELECT ?folio ?glossTitle ?glossIdentifier ?glossType ?glossPlacement
WHERE {
  ?gloss a memo:Gloss ;
        dcterms:title ?glossTitle;
        dcterms:identifier ?glossIdentifier;
        mmdio:hasGlossType ?glossType;
        mmdio:hasGlossPlacement ?glossPlacement;
        frbr:embodiment ?folio .
  ?folio a memo:Folio .
  FILTER(?glossType IN (mmdio:ordinary, mmdio:extravagant))
  FILTER(?glossPlacement IN (mmdio:interlinear, mmdio:marginal))
}
```

Listing 4: Query to retrieve glosses along with their placement and type.

The gloss information from the ‘London, Ogden 5’ manuscript collection, particularly identified in ‘Authenticum’ manuscripts, is presented. One specific gloss retrieved through this query is titled ‘ex uno eodemque principio’, with the identifier ‘nov.6.pr’ in Figure 8. This gloss is categorized as ‘ordinary’ in type and ‘linear’ in placement. The information presented by gloss inside the text of the manuscript enables scholars to understand the interpretation, context, and meaning of the content.



**Figure 8.** Gloss types and placement across folios.

The visualized output demonstrates how each gloss is associated with its corresponding folio, and how glosses vary by type and position, offering a deeper layer of scholarly interpretation for manuscript researchers.

#### 4.13. Manuscript Provenance and Location Tracking

Provenance is a vital aspect of manuscript studies. A dedicated query (Listing 5) was run to retrieve both current and previous locations of the manuscript collection “Paris, Mazar. 452”. The output in Figure 9 indicates that the manuscript was originally kept in Paris and is now preserved in Italy.

```

SELECT ?currentLocationName ?previousLocationName
WHERE {
  ?manuscriptCollection a mmdio:ManuscriptCollection ;
    dcterms:title "Paris, Mazar. 452" ;
    core:hasLocation ?currentLocation, ?
      previousLocation.

  ?currentLocation a dcterms:Location ;
    arco:hasLocationType arco:
      CurrentPhysicalLocation ;
    mmdio:hasLocationName ?currentLocationName.

  ?previousLocation a dcterms:Location ;
    arco:hasLocationType arco:PreviousLocation ;

```

```

}
mmdio:hasLocationName ?previousLocationName .

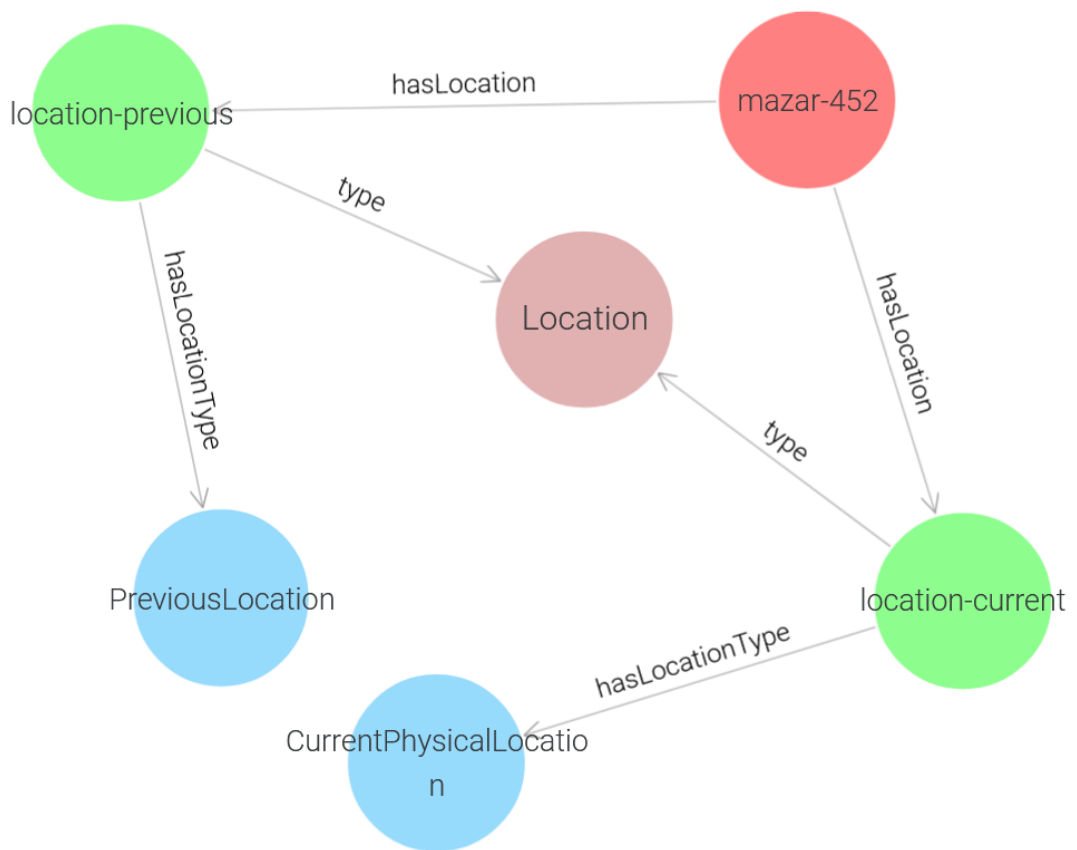
```

Listing 5: SPARQL query returning current and previous locations of a manuscript collection.

The query retrieved data on the various locations where the manuscript has been housed over time, offering valuable insights into its custodial history. This information is crucial for examining the manuscript's provenance and understanding the broader patterns of manuscript distribution.

The query returned data on the manuscript's past locations, providing valuable insight into its custodial history. This helps trace its provenance and highlights on broader patterns of manuscript circulation.

Figure 9 provides information about the previous location, 'Paris', and the current one in 'Italy'. The manuscript's journey from its last to its current location offering insight into its historical journey and transitions



**Figure 9.** The provenance chain of the "Paris, Mazar. 452" manuscript, showing its previous and current locations.

This chain of provenance is very important for research into ownership history, manuscript circulation, and preservation. The PKG allows for such relationships to be expressed semantically and queried dynamically.

#### 4.14. Physical Characteristics and Codicology

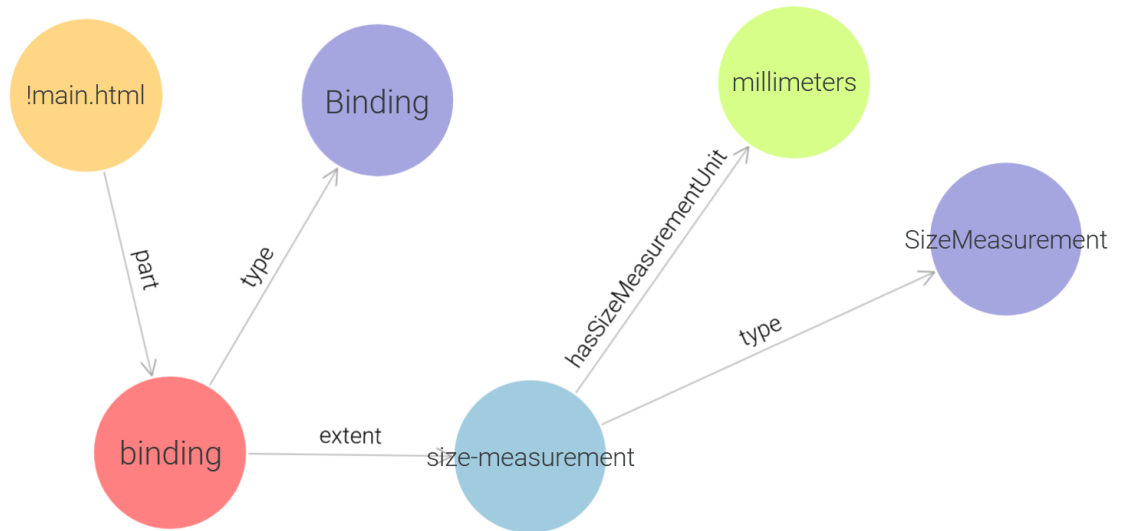
The codicological description of manuscripts includes information such as binding dimensions. A SPARQL query (Listing 6) was used to retrieve the length and width of the binding of Codex-003, along with the measurement units. The results (Figure 10) indicate that the codex measures 410 mm in length and 285 mm in width.

```
SELECT ?binding ?length ?width ?unit
WHERE {

    ?codex a memo:Codex ;
           dcterms:identifier "http://irnerio.cirsfid.unibo.it/
                               codex/003" ;
           frbr:part ?binding .
    ?binding a memo:Binding ;
             dcterms:extent ?sizeMeasurement .
    ?sizeMeasurement a memo:SizeMeasurement ;
                     memo:hasLength ?length ;
                     memo:hasWidth ?width ;
                     memo:hasSizeMeasurementUnit ?unit .
    FILTER(?unit IN (memo:millimeters, memo:centimeters, memo:
                     decimeters, memo:meters))
}
```

Listing 6: A SPARQL query to retrieve the dimensions of the binding (length and width) and the size measurement unit for the codex-003.

This query returns exact measurements of the codex's binding, including length and width in particular units. It gives an understanding of the physical composition of this particular codex, which is key information for preservation and cataloging.



**Figure 10.** The diagram representing the physical dimensions of Codex-003, detailing the binding’s length, width, and measurement unit.

Figure 10 shows the binding size of Codex-003 and is the data that has been acquired through the query. The codex measures 410 mm in length and 285 mm in width. This graphical representation is helpful for a better understanding of the scale and physical condition of the codex, which is useful for many scholarly and conservation activities.

#### 4.15. Textual Content

The query below in Listing 7 designed to extract textual roles and contents related to the manuscript “Tres libri cum glossa Accursii” from the Codex-282 manuscript collection. It aims to find the complete textual metadata that would assist in scholarly readings of the manuscript.

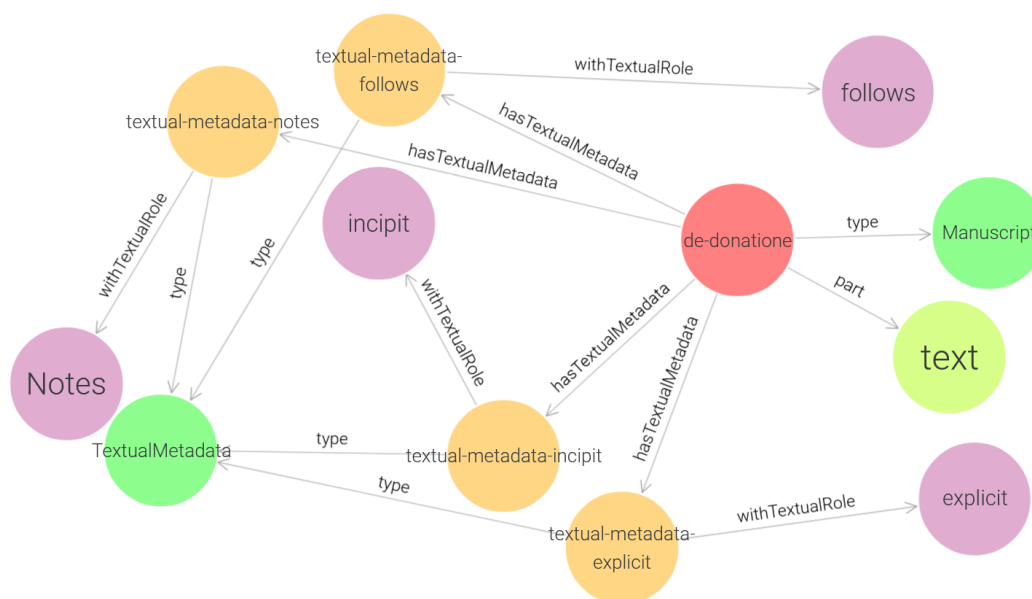
```

SELECT ?TextualRole ?hasTextualContent
WHERE {
  ?Manuscript a memo:Manuscript ;
              dcterms:title "Tres libri cum glossa Accursii" ;
              memo:hasTextualMetadata ?TextualMetadata .
  ?TextualMetadata a mmdio:TextualMetadata ;
                  memo:withTextualRole ?TextualRole ;
                  mmdio:hasTextualContent ?hasTextualContent
}

```

Listing 7: A SPARQL query which returns the textual roles and content of the textual metadata associated with the manuscript titled “Tres libri cum glossa Accursii”.

This query effectively obtains the textual metadata associated with particular textual roles, including incipit, explicit, follows, and notes, ensuring that each element’s textual detail is appropriately captured. This structured retrieval is essential for academics who require a thorough understanding of the manuscript’s textual framework for analysis or citation.



**Figure 11.** A visual representation of the manuscript’s textual roles and metadata elements, demonstrating the richness of textual content.

The Figure 11 efficiently presents the textual metadata elements incipit, explicit, follows, and notes. It provides a clear representation of how these elements interact within the manuscript’s structure. This enables scholars to easily find and retrieve specific textual content relevant to their research interests.

#### 4.16. Cross-Repository Citation Networks

An essential feature of the PKG is the ability to detect citations across repositories. One query (Listing 8) retrieves instances in which books in the Mosaico collection cite manuscripts in the Innerio collection. The output (Figure 13) shows how a book like “Authenticum” authored by Cristina Vano cites Innerio manuscripts such as “Authenticum cum glossa Accursii”.

```

SELECT ?book ?bookTitle ?manuscript ?manuscriptTitle
WHERE {
  ?book a fabio:Book ;
    frbr:partOf ?bookCollection;
    dcterms:title ?bookTitle;
    cito:cites ?manuscript .
  ?bookCollection a mmdio:DigitalResourceCollection;

```

```

    dcterms:title "mosaico" .
?manuscript a memo:Manuscript ;
    dcterms:title ?manuscriptTitle ;
    frbr:partOf ?manuscriptCollection .
?manuscriptCollection a mmdio:DigitalResourceCollection ;
    dcterms:title "irnerio" .
}

```

Listing 8: A SPARQL query showing citations from Mosaico books to Innerio manuscripts.

The query identifies relationships where books in the Mosaico collection, such as ‘Authenticum’ by ‘Christina Vano’, cite manuscripts like “Authenticum cum glossa Accursii” from the Innerio collection. This connection exemplifies how scholarly resources can be interconnected across different digital repositories



**Figure 12.** Cross-repository citation graph: “Authenticum” from Mosaico cites a manuscript from Innerio.

Figure 12 visually illustrates the interconnections among various platforms, mapping how specific books and manuscripts are linked. This graphical representation helps to clarify the complexity of these relationships and underscores the importance of effective digital resource management in academic contexts.

The query results not only confirm the successful integration of these platforms but also highlight the strength of digital collections in facilitating scholarly research across multiple repositories. This form of inter-repository citation demonstrates the knowledge graph’s ability to model scholarly relationships across diverse sources, thereby supporting citation analysis in manuscript studies.

#### 4.17. Comprehensive Use Case Coverage: Integrated Exploration of Mosaico Collections

The PKG also supports deep exploration of specific collections, such as:

*Montecassino Collection* To evaluate the effectiveness of the extended MMDIO ontology, a set of digital manuscripts from the Mosaico platform was queried and analyzed. These include Montecassino, Olomouc, Roffredo Beneventano, Arbores, and Santana, each presenting distinct structural and semantic modeling challenges such as folio mapping, hierarchical text organization, and bibliographic linkage.

For example, the Montecassino and Roffredo Beneventano books were tested using SPARQL queries to extract detailed structural elements. Listing 9 illustrates a query used to retrieve sections, subsections, manuscript collections, and associated bibliographic references from the Montecassino book.

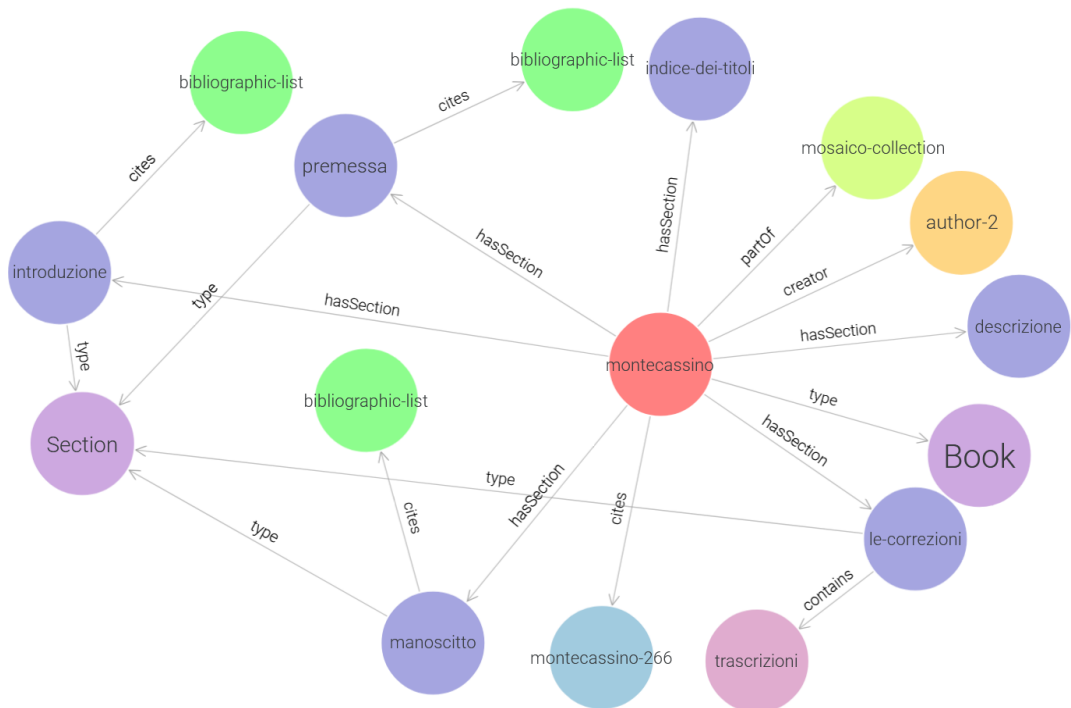
```

SELECT ?book ?section ?mauscriptCollection ?
  BibliographicList
WHERE {
  ?book a fabio:Book ;
    dcterms:title "Montecassino (APPARATUS DECRETALIUM)";
    mmdio:hasSection ?section;
    cito:cites ?mauscriptCollection.
  ?section a doco:Section;
    cito:cites ?BibliographicList.
  ?BibliographicList a biro:BibliographicList.
  ?mauscriptCollection a mmdio:ManuscriptCollection.
}

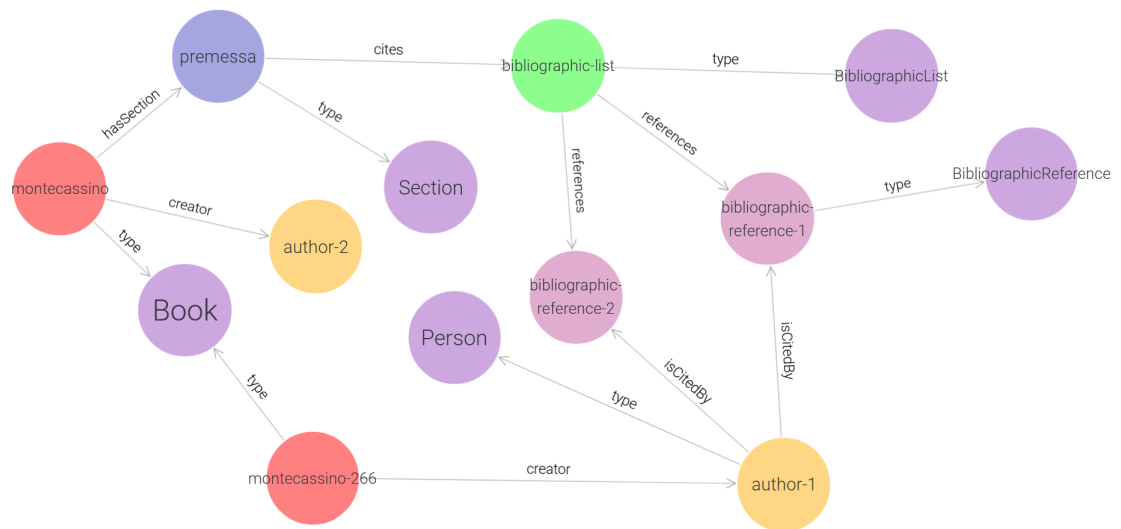
```

Listing 9: A SPARQL query to retrieve the details associated with ‘Montecassino’ book, including its manuscript collection, sections, subsections, and the bibliographic lists.

The query retrieves the structural components of the Montecassino book, including sections such as *Descrizione* and *Le correzioni*, and subsections like *Trascrizioni*, along with the associated manuscript collection. Each element is linked to specific folios, providing a clear and systematic view of how textual content is mapped onto its physical representation.



**Figure 13.** The figure shows the organized connections between sections, subsections, and manuscripts in the Montecassino book and how these relate to folio representations.



**Figure 14.** The diagram presents the author of the physical version of Montecassino cited by the bibliographic reference.

Figure 13 illustrates the logical structure of the Montecassino book, showing how sections, subsections, manuscript collections, and bibliographic lists are interlinked within its digital representation. Figure 14 focuses on the physical version of the manuscript, attributed to *Goffredus Tranensis*, whose signature is available online at [this link](#), and referenced within the *Permissa* section.

These visualizations are especially useful for scholars, as they offer a structured overview of the manuscript's organization, facilitating efficient navigation and deeper understanding of its textual and bibliographic components.

*Olomouc Collection* This query, as shown in Listing 10, is formed to get the folios, with due attention to specific associated data, constituting some facet of the manuscript known as 'Liber 1' within the 'Olomouc' book collection. This query aims to identify folio portions using recto and verso identifiers, which are useful in understanding manuscript layout from a physical standpoint.

```

SELECT ?Folio ?RectoIdentifier ?VersoIdentifier ?manuscript
WHERE {
  ?Book a fabio:Book;
        dcterms:title "Olomouc".
  ?manuscript a memo:Manuscript ;
             dcterms:title "Liber I" ;
             mmdio:refersToManifestation ?Folio .

  ?Folio a memo:Folio .
  OPTIONAL {
    ?Folio frbr:part ?recto .
    ?recto a memo:Recto ;
           dcterms:identifier ?RectoIdentifier .
  }
  OPTIONAL {
    ?Folio frbr:part ?verso .
  }
}

```

```
    ?verso a memo:Verso ;
      dcterms:identifier ?VersoIdentifier .
  }
}
```

Listing 10: A SPARQL query which returns the folios of Liber 1 manuscript in ‘Olmouc’ Book.

The query will return all the folios connected to the Liber 1, using `mmdio:refersToManifestation` to establish a relationship between a manuscript and its folios. The optional clauses are used to query for recto or verso identifiers, so that whenever such information is available, it is collected. Additionally, if some folio has exclusively recto or verso information, the data is obtained as well, preventing information loss due to insufficient specifications.

The result is a structured catalog of Liber 1 manuscript folios, providing insight into how the manuscript pages are physically assembled and referenced. By identifying both the recto (front) and verso (back) sides of each folio, the query gives a detailed picture of the folio structure.



**Figure 15.** This graphical representation of the ‘Liber 1’ manuscript folios displays the connections between folios and their respective recto and verso identifiers.

The Figure 15 visually represents the relationship between the Liber 1 manuscript and its associated folios, as retrieved by the query. The figure shows how each folio in the digital version of the ‘Olomouc’ manuscript is linked to its respective recto and verso sides. In this visualization, 11 folios are referenced, each displaying its corresponding recto and verso identifiers, which help to distinguish the front and back of each folio. This visualization is particularly useful for researchers and scholars as it provides a clear depiction of the manuscript’s structure, helping to trace the exact layout of each folio in the collection.

*Roffredo Beneventano* The query in Listing 11 is designed to retrieve detailed information from the digital version of the Roffredo Beneventano book. The query specifically targets the sections of the book and the manuscript collections that are cited within the text. By retrieving both the sections and the referenced manuscript collections, this query provides a comprehensive overview of how the book is structured and its relation to other historical documents.

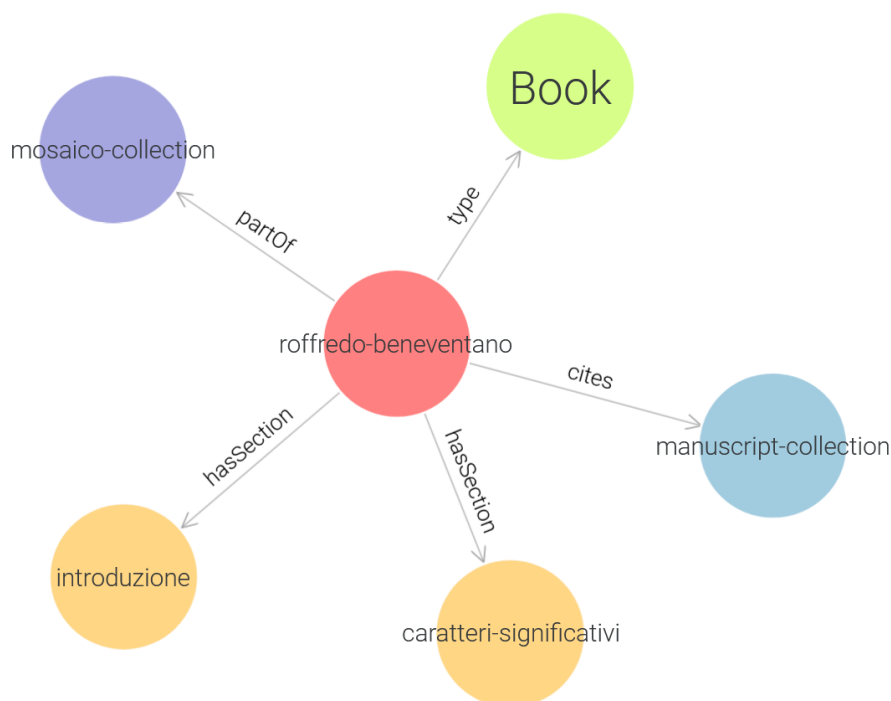
The query first retrieves the sections of the Roffredo Beneventano book, using the `mmdio:hasSection` property to link the book to its sections. It also retrieves the

manuscript collection referenced within these sections, ensuring that all cited collections are included. The query relies on the `cito:cites` property to identify the manuscript collections connected to the book.

```
SELECT ?book ?section ?mauscriptCollection ?manuscript
WHERE {
  ?book a fabio:Book ;
        dcterms:title "Roffredo Beneventano";
        mmdio:hasSection ?section;
        cito:cites ?mauscriptCollection.
  ?section a doco:Section.
  ?mauscriptCollection a mmdio:ManuscriptCollection.
}
```

Listing 11: A SPARQL query to retrieve sections and manuscript collection of in the Roffredo Beneventano book.

The output of the query includes the relevant sections and the manuscript collections cited within each section. This structured data helps scholars understand the organization of the Roffredo Beneventano book and its references to external manuscript collections.



**Figure 16.** The graphical diagram showing the sections and manuscript collections in the Roffredo Beneventano book.

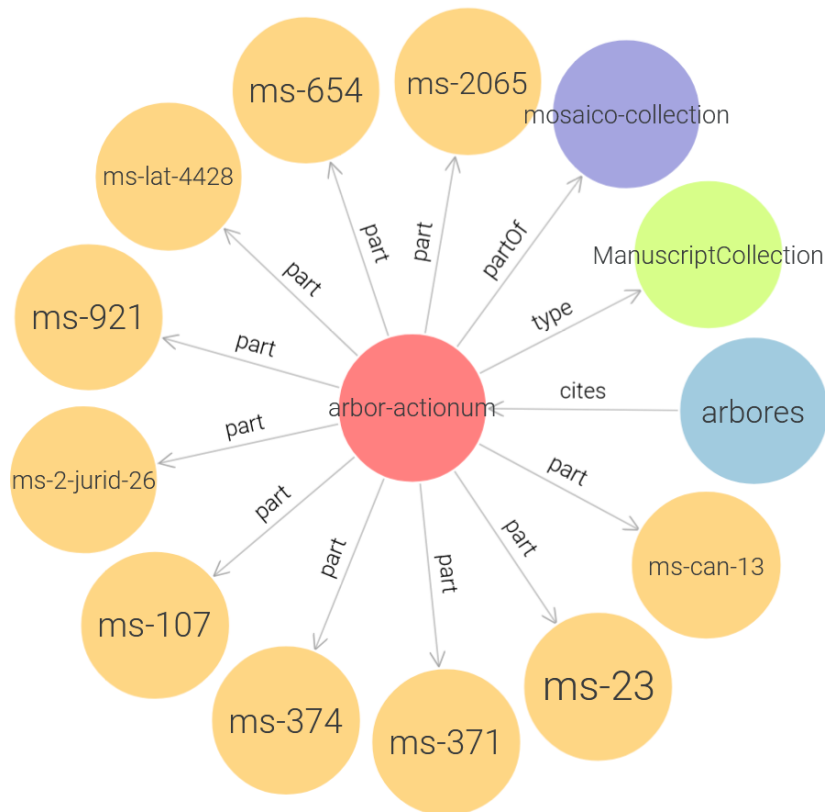
Figure 16 visually represents the relationship between the sections of the Roffredo Ben-eventano book and the manuscript collections it cites. Key sections, such as Introduzione (Introduction) and Caratteri Significativi (Significant Features), are shown about the cited manuscript collection. This visualization provides a clear depiction of how the book's sections are organized and linked to external sources, offering a better understanding of the book's content and historical significance.

*Arbores Collection* The query in Listing 12 is designed to retrieve the manuscripts that are part of the 'Arbores' collection. This query focuses on identifying all manuscripts associated with the 'Arbores' book by referencing the manuscript collections cited within it. The purpose of this query is to provide a detailed overview of the manuscripts that belong to the 'Arbores' collection, which are cataloged in various libraries across Europe. The query first retrieves the manuscript collections cited in the 'Arbores' book using the `cito:cites` property. Each manuscript collection is further explored using the `frbr:part` property to identify the individual manuscripts that belong to the collection. The result of this query is a comprehensive list of manuscripts, providing valuable data on how the 'Arbores' collection is structured and the specific manuscripts that it references.

```
SELECT    ?manuscriptCollection ?manuscript
WHERE {
  ?book a fabio:Book ;
        dcterms:title "Arbores";
        cito:cites ?manuscriptCollection .
  ?manuscriptCollection a mmdio:ManuscriptCollection;
        frbr:part ?manuscript .

  ?manuscript a memo:Manuscript;
}
```

Listing 12: A SPARQL query to retrieve the manuscript part of the 'Arbores' collection. The output of the query successfully retrieves all manuscripts that are part of the Arbores collection, offering a clear picture of the manuscript structure and their locations in different libraries.



**Figure 17.** The visual representation of manuscripts in the ‘Arbores’ collection and their library locations.

Figure 17 visually represents the manuscripts that are part of the Arbores collection, detailing their library locations.

*Satana Collection* The query designed in Listing 13 is intended to retrieve the sections from the ‘Satana’ digital version of the book and the manuscripts that belong to its manuscript collections. This query helps in organizing and understanding how the ‘Satana’ book is structured and how it references various manuscripts within its collection. The query retrieves all sections of the ‘Satana’ book using the `mmdio:hasSection` property and connects each section with the cited manuscript collection through the `cito:cites` property. It also identifies the individual manuscripts within each manuscript collection using the `mmdio:ManuscriptCollection` property, ensuring that both the sections and their related manuscripts are fully cataloged.

```

SELECT ?book ?section ?mauscriptCollection ?manuscript
WHERE {
  ?book a fabio:Book ;
    dcterms:title "Satana";
    mmdio:hasSection ?section;
    cito:cites ?mauscriptCollection.

```

```

?section a doco:Section.
?manuscriptCollection a mmdio:ManuscriptCollection;
    frbr:part ?manuscript.
?manuscript a memo:Manuscript.
}

```

Listing 13: A SPARQL query to retrieve sections and manuscripts belong to its collection.

The output of the query includes relevant sections and the manuscripts that are part of the collection. This structured data is essential for scholars who are studying the organization of the ‘Satana’ book and its associated manuscripts, providing a clearer understanding of the relationship between the book’s sections and its cited manuscript collections.



Figure 18. This figure shows the Satana manuscript sections and the manuscript collection it cites

Figure 18 represents the sections associated with the ‘Satana’ book, such as Descrizione (Description), Metodologia (Methodology), and Elenco manoscritti (List of Manuscripts). It also illustrates how these sections cite various manuscript collections. These diagrams are particularly helpful for scholars, as they visualize the complex structure of the Satana book and its manuscript collections. By providing an overview of the sections and a breakdown of the manuscripts, researchers can more easily navigate the digital version of the manuscript collection.

#### 4.18. Summary

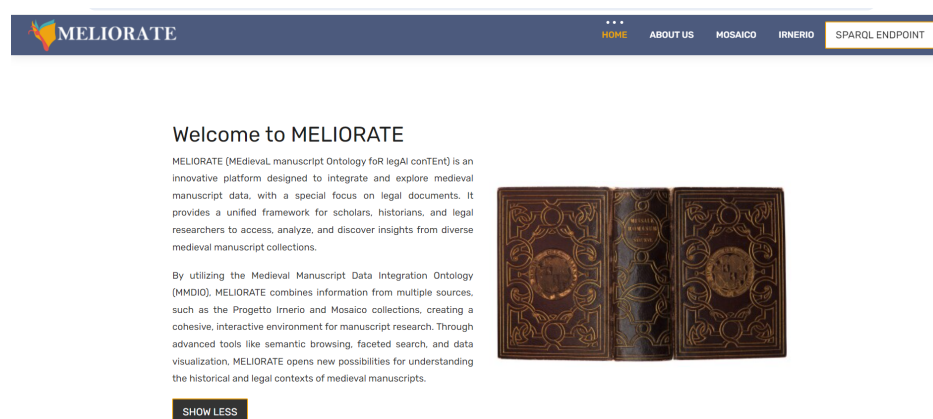
The above use cases confirm the ability of the Polymorphic Knowledge Graph to integrate, query, and visualize complex medieval manuscript metadata. By unifying collections from Mosaico and Inerio, the PKG enables scholars to seamlessly explore manuscripts across platforms and contexts. Each SPARQL query demonstrates a different facet of scholarly inquiry, from bibliographic metadata and gloss annotations to citation tracing and codicological detail. The PKG thus represents a powerful and extensible model for digital humanities research, offering scalable and semantically rich access to cultural heritage data.

### 5. MELIORATE: A Linked Open Data Portal for Medieval Manuscripts

The **ME**dieval **manuscrIpt** **Ontology** **f**o**R** **legAl** **conTEnt** (**MELIORATE**) is a Linked Open Data (LOD) platform developed to provide unified, semantic access to medieval manuscripts from heterogeneous digital sources. Built upon the **MMDIO** ontology, MELIORATE supports cross-repository research by integrating datasets from both the *Progetto Inerio* and *Mosaico* platforms. It enables the dynamic exploration of complex legal, historical, and cultural relationships via a polymorphic knowledge graph.

#### 5.1. Platform Overview and Interface

MELIORATE offers a user-centric interface for navigating digitized manuscript data through structured access points such as collections, codices, and folios. Figure 19 shows the homepage layout, which introduces users to the system and provides intuitive access to each integrated collection. Tabs direct users to individual corpora, ensuring both experts and novices can explore the underlying knowledge graph effectively.



**Figure 19.** The MELIORATE homepage interface providing entry points to key manuscript collections.

## 5.2. Mosaico Collections

MELIORATE integrates several thematic collections from Mosaico, each offering rich metadata about medieval legal texts and associated manuscripts. These include **Authenticum**, **Montecassino**, **Olomouc**, **Arbores**, **Roffredo Beneventano**, and **Satana**. Each book or collection in the Mosaico portal is semantically described and interlinked using RDF.

Figure 20 displays the digital books available under Mosaico, which serve as core references in medieval legal traditions. Users can explore associated sections, bibliographic citations, manuscript collections, folios, and annotations through a structured interface.

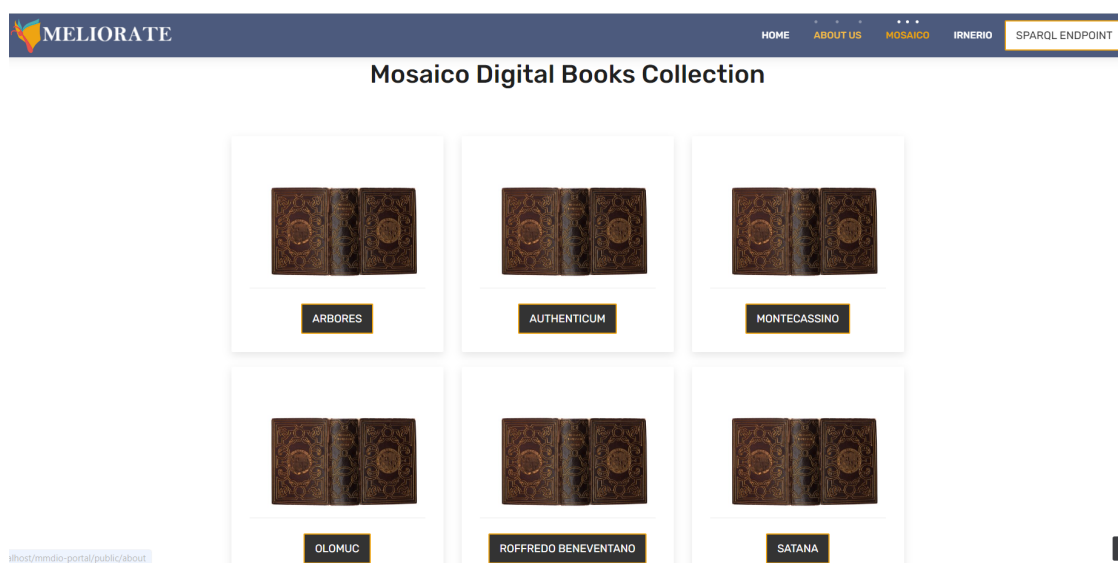


Figure 20. Digital books from the Mosaico collection in MELIORATE.

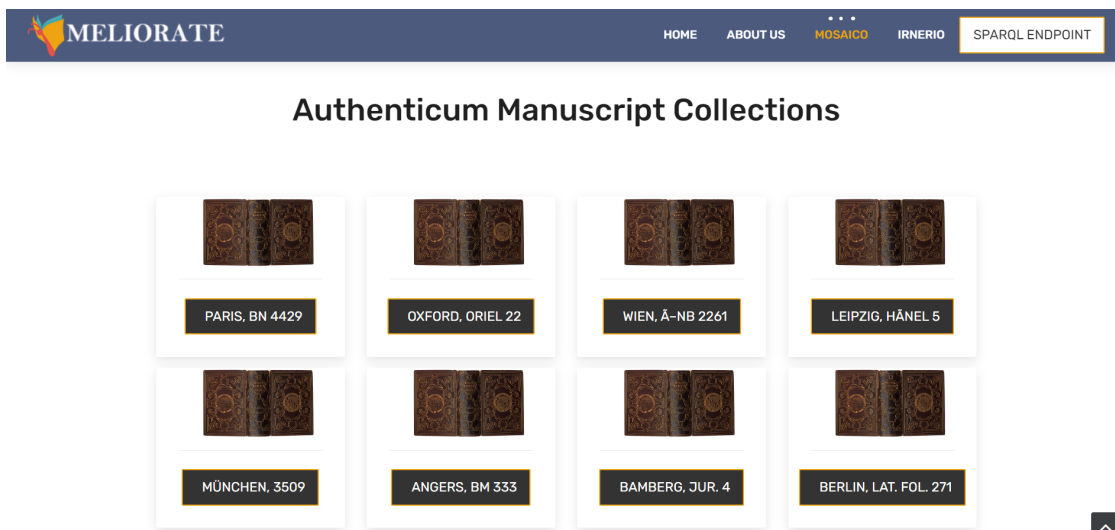
Each book has its associated manuscript collection which is further linked to metadata such as creation period, geographic origin, bibliographic references, and foliation. This enables a comprehensive understanding of each manuscript's historical trajectory and scholarly significance.

## 5.3. Authenticum Collection

The Authenticum, one of the most prominent digital collections available on the Mosaico Platform, is organized into several divisions and includes 26 manuscript collections, each with a unique viewpoint on medieval legal traditions. Users can easily explore numerous aspects of the Authenticum collection, with information provided for each of the relevant sections and manuscript collections within a user-friendly organizing system. The connected metadata is presented clearly to provide an easy portal navigation experience. Figure 21 shows the manuscript collection and sections linked to the digital Authenticum. The manuscript collection is entirely navigable, allowing the user to view information about the manuscripts held in the collections, the periods covered by

the collections, where they are held, bibliographic references, further descriptions, and embodiment in a specific codex.

Figure 21 presents the manuscript collections and sections associated with the digital version of Authenticum. Each manuscript collection is fully navigable, allowing users to view detailed information such as the manuscripts included in the collection, the period they cover, their location, bibliographic references, supplementary descriptions, and how they are embodied in a specific codex.



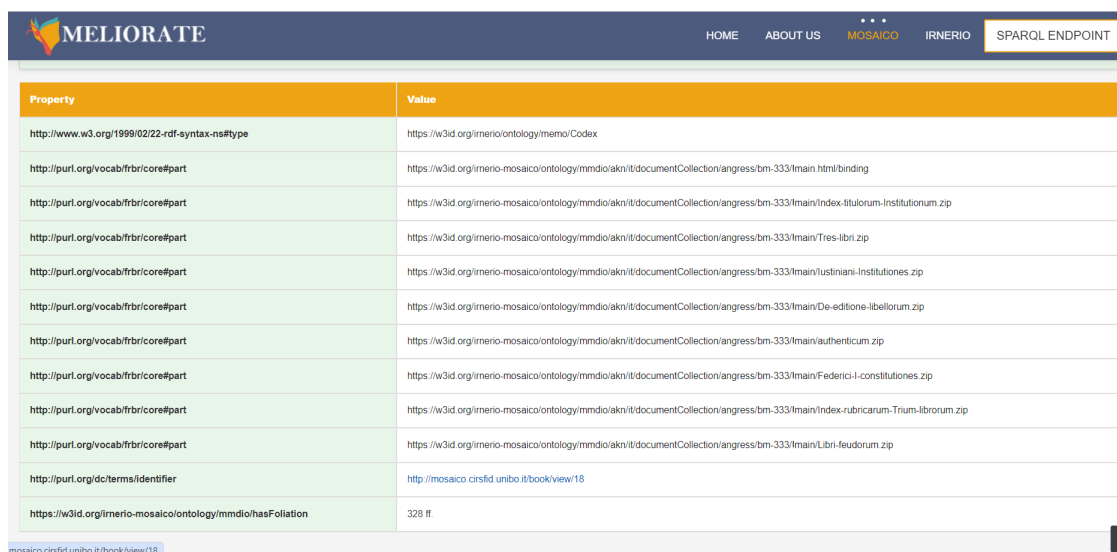
**Figure 21.** The user interface provides an organized layout of the ‘Authenticum’ collection and its sections.

Figure 22 highlights a specific collection, such as Angres, DM, 33, and displays the associated manuscripts that are part of the collection, along with the time they cover, their current and previous locations, supplementary descriptions, and how they are embodied within a particular codex.

Property	Value
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	https://w3id.org/irnerio-mosaico/ontology/mmdio/ManuscriptCollection
http://purl.org/vocab/fr/core#part	https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angers/bm-333/main/instituti-ones
http://purl.org/vocab/fr/core#part	https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angers/bm-333/main/index-ibellorum-institutionum
http://purl.org/vocab/fr/core#part	https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angers/bm-333/main/de-editione-ibellorum
http://purl.org/vocab/fr/core#part	https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angers/bm-333/main/authenticum
http://purl.org/vocab/fr/core#part	https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angers/bm-333/main/federici-i-constitutiones
http://purl.org/vocab/fr/core#part	https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angers/bm-333/main/index-rubicarum-trium-ibronum
http://purl.org/vocab/fr/core#part	https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angers/bm-333/main/ibri-audorum
http://purl.org/vocab/fr/core#part	https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angers/bm-333/main/ibri-ibri
http://purl.org/vocab/fr/core#embodiment	https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angers/bm-333/main.html
http://purl.org/dc/terms/title	Angers, BM 333

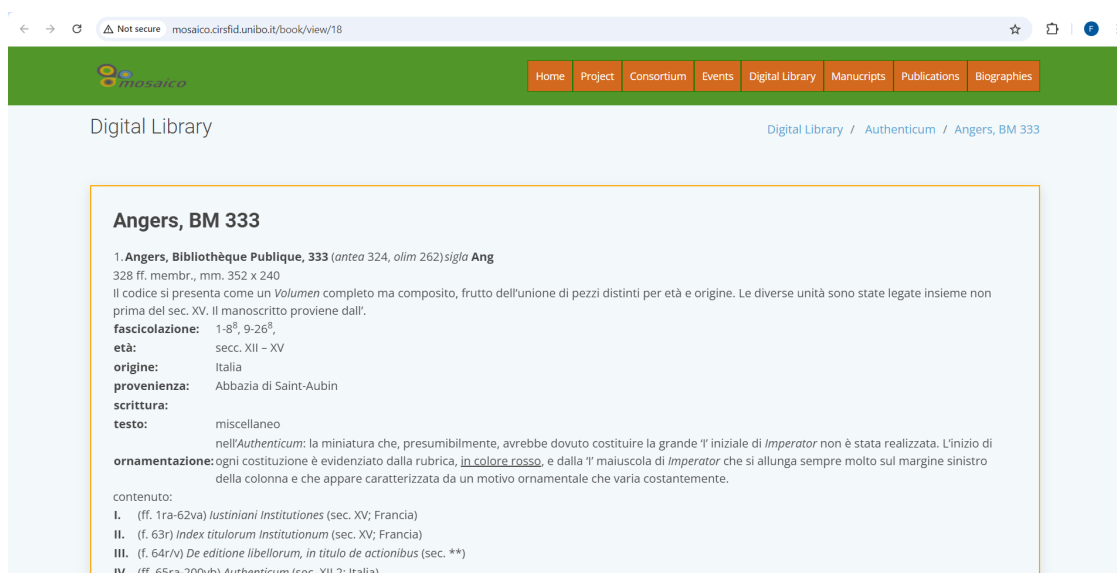
**Figure 22.** The interface showcases detailed information about the Angres collection.

Figure 23 provides details about the codex in which the Angres, DM, 33 manuscript collection is embodied. It includes information about the foliation, specific folios that are part of it, and their unique identifiers. The identifiers provide direct links to the Mosaico portal where this codex is hosted. Figure 24 shows the interface of the Mosaico portal displaying the Angres, DM, 33 codex.



Property	Value
<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	<a href="https://w3id.org/irnerio/ontology/memo/Codex">https://w3id.org/irnerio/ontology/memo/Codex</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main.html/binding">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main.html/binding</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Index-titulorum-Institutionum.zip">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Index-titulorum-Institutionum.zip</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Tres-libri.zip">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Tres-libri.zip</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Justiniani-Institutiones.zip">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Justiniani-Institutiones.zip</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/De-editione-libellorum.zip">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/De-editione-libellorum.zip</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/authenticum.zip">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/authenticum.zip</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Federici-I-constitutiones.zip">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Federici-I-constitutiones.zip</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Index-rubricarum-Trium-librorum.zip">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Index-rubricarum-Trium-librorum.zip</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Libri-feudorum.zip">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/angres/bm-333/main/Libri-feudorum.zip</a>
<a href="http://purl.org/dc/terms/identifier">http://purl.org/dc/terms/identifier</a>	<a href="http://mosaico.cirsfid.unibo.it/book/view/18">http://mosaico.cirsfid.unibo.it/book/view/18</a>
<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/hasFoliation">https://w3id.org/irnerio-mosaico/ontology/mmdio/hasFoliation</a>	328 ff.

Figure 23. The interface provides the information about the codex.



mosaico.cirsfid.unibo.it/book/view/18

Home Project Consortium Events Digital Library Manuscripts Publications Biographies

Digital Library Digital Library / Authenticum / Angres, BM 333

### Angres, BM 333

1. Angres, Bibliothèque Publique, 333 (antea 324, olim 262) *sigla Ang*  
328 ff. membr., mm. 352 x 240  
Il codice si presenta come un *Volumen* completo ma composito, frutto dell'unione di pezzi distinti per età e origine. Le diverse unità sono state legate insieme non prima del sec. XV. Il manoscritto proviene dall'.

**fascicolazione:** 1-8<sup>8</sup>, 9-26<sup>8</sup>,  
**età:** secc. XII – XV  
**origine:** Italia  
**provenienza:** Abbazia di Saint-Aubin  
**scrittura:**  
**testo:** miscellaneo  
nell'*Authenticum*: la miniatura che, presumibilmente, avrebbe dovuto costituire la grande "I" iniziale di *Imperator* non è stata realizzata. L'inizio di  
**ornamentazione:** ogni costituzione è evidenziata dalla rubrica, in colore rosso, e dalla "I" maiuscola di *Imperator* che si allunga sempre molto sul margine sinistro della colonna e che appare caratterizzata da un motivo ornamentale che varia costantemente.

contenuto:  
I. (ff. 1ra-62va) *Justiniani Institutiones* (sec. XV; Francia)  
II. (f. 63r) *Index titulorum Institutionum* (sec. XV; Francia)  
III. (f. 64r/v) *De editione libellorum, in titulo de actionibus* (sec. \*\*)  
IV. (ff. 65ra-200vb) *Authenticum* (sec. XII,2; Italia)

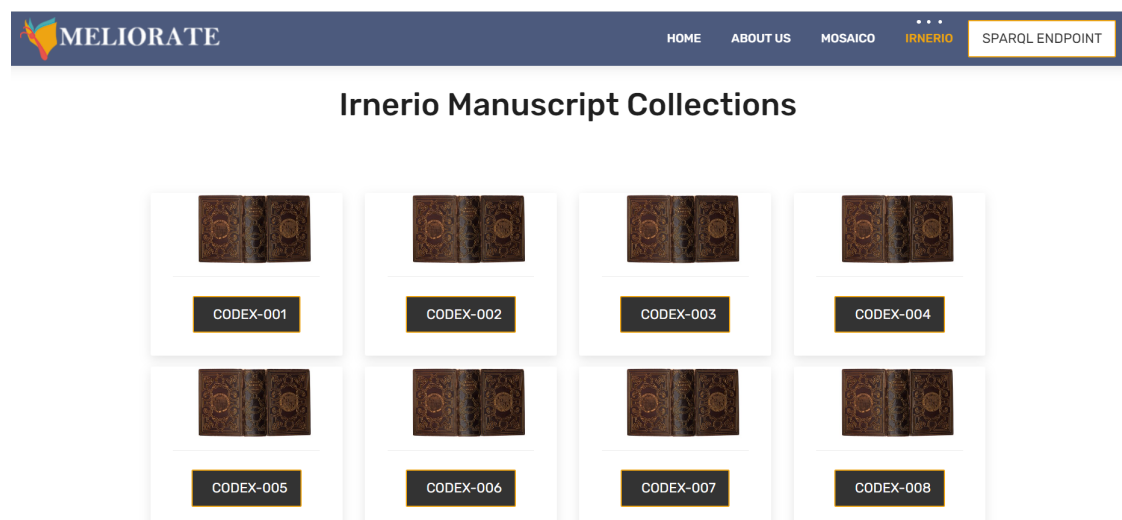
Figure 24. The original interface of the Mosaico portal for the Angres, DM, 33 collection.

This approach transforms the information from the original portal into a more structured format. It allows users to easily explore the ‘Authenticum’ collection. The organized interface simplifies navigation through the manuscript collections. This helps users gain deeper insights into medieval legal traditions and codex structures.

#### 5.4. *Irnerio Platform Integration*

The Irnerio section in MELIORATE comprises 298 codices enriched with bibliographic, historical, and artistic metadata. The Irnerio collection is named after the 12th-century jurist Irnerius, who restored Roman law. The collection allows users to browse the collections, read descriptions, and retrieve the associated folios.

Each manuscript collection includes detailed metadata, such as bibliographic references, geographical and chronological details, and artistic elements. Figure 25 presents an overview of the entire collection, covering codices from codex-001 to codex-286, as well as other collections labeled ‘A’, ‘B’, ‘C’, ‘F’, ‘G’, ‘I’, ‘J’, ‘K’, and ‘L’. The structured navigation accurately captures and organizes all relevant information from these collections.



**Figure 25.** Overview of Irnerio codices integrated into MELIORATE.

Figure 26 provides the information associated with the codex-001, such as the manuscripts that are part of its collection, the collection title, the time interval, the location, and its embodiment of the codex.

IRI: <https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001>

Property	Value
<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/ManuscriptCollection">https://w3id.org/irnerio-mosaico/ontology/mmdio/ManuscriptCollection</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main/biblia-vulgata">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main/biblia-vulgata</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main/interpretationes-hebraicorum-nominum">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main/interpretationes-hebraicorum-nominum</a>
<a href="http://purl.org/vocab/frbr/core#embodiment">http://purl.org/vocab/frbr/core#embodiment</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main.html">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main.html</a>
<a href="http://purl.org/dc/terms/title">http://purl.org/dc/terms/title</a>	Codex-001
<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/hasDescription">https://w3id.org/irnerio-mosaico/ontology/mmdio/hasDescription</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/supplementary-information-description">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/supplementary-information-description</a>
<a href="http://purl.org/spar/tvc/atTime">http://purl.org/spar/tvc/atTime</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/time-interval">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/time-interval</a>
<a href="https://w3id.org/arco/ontology/core/hasLocation">https://w3id.org/arco/ontology/core/hasLocation</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/location-current">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/location-current</a>
<a href="http://purl.org/vocab/frbr/core#partOf">http://purl.org/vocab/frbr/core#partOf</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/irnerio-collection">https://w3id.org/irnerio-mosaico/ontology/mmdio/irnerio-collection</a>

Figure 26. The portal provides the details about the particular collection Codex-001

Figure 27 offers more information about Codex-001, including the folios contained within, details about its foliation, script style, and the unique identifier of the codex, which links to the source.

IRI: <https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001>

Property	Value
<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/ManuscriptCollection">https://w3id.org/irnerio-mosaico/ontology/mmdio/ManuscriptCollection</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main/biblia-vulgata">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main/biblia-vulgata</a>
<a href="http://purl.org/vocab/frbr/core#part">http://purl.org/vocab/frbr/core#part</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main/interpretationes-hebraicorum-nominum">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main/interpretationes-hebraicorum-nominum</a>
<a href="http://purl.org/vocab/frbr/core#embodiment">http://purl.org/vocab/frbr/core#embodiment</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main.html">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/main.html</a>
<a href="http://purl.org/dc/terms/title">http://purl.org/dc/terms/title</a>	Codex-001
<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/hasDescription">https://w3id.org/irnerio-mosaico/ontology/mmdio/hasDescription</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/supplementary-information-description">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/supplementary-information-description</a>
<a href="http://purl.org/spar/tvc/atTime">http://purl.org/spar/tvc/atTime</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/time-interval">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/time-interval</a>
<a href="https://w3id.org/arco/ontology/core/hasLocation">https://w3id.org/arco/ontology/core/hasLocation</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/location-current">https://w3id.org/irnerio-mosaico/ontology/mmdio/akn/it/documentCollection/codex-001/location-current</a>
<a href="http://purl.org/vocab/frbr/core#partOf">http://purl.org/vocab/frbr/core#partOf</a>	<a href="https://w3id.org/irnerio-mosaico/ontology/mmdio/irnerio-collection">https://w3id.org/irnerio-mosaico/ontology/mmdio/irnerio-collection</a>

Back

Figure 27. The interface provides detailed information about the embodiment of Codex-001.

← → ↻ Not secure innerio.cirsfid.unibo.it/codex/001

## CIRSFID Innerio

<p><b>Project</b></p> <ul style="list-style-type: none"> <li>Homepage</li> <li>About the project</li> <li>Partners</li> <li>Help</li> </ul> <p><b>Codices</b></p> <ul style="list-style-type: none"> <li>Codices</li> <li>Authors</li> <li>Search</li> </ul> <p><b>Change language</b></p> <ul style="list-style-type: none"> <li>Italiano</li> </ul>	<p>CODEX: 001</p> <hr/> <p><b>Century:</b> XIII/2  <b>Material:</b> membranaceo  <b>Size (mm):</b> 333x240 (210x155)  <b>Foliation:</b> I+504+II  <b>Columns:</b> 2</p> <hr/> <p><b>Description</b></p> <p>Altre segnature: IV-5; B-II-5. Scrittura gotica libraria italiana. Mani diverse. Fascicoli di ff. 6 con richiami. Colonne di linee 46. Due follaioni, ambedue antiche, ma posteriori al codice: abbiamo seguito la pi<sup>a</sup> evidente.</p> <hr/> <p><b>Works</b></p> <ol style="list-style-type: none"> <li>Biblia vulgata (con i Prologi di S. Girolamo o a lui attribuiti, prima di ogni libro o di ogni gruppo di libri)</li> <li>Interpretationes hebraicorum nominum</li> </ol> <hr/> <p><b>Bibliography</b></p> <p>A. Mart<sup>á</sup>nez de Pons, Bibliotheca Albornotiana. Continet manuscriptos codices qui adservantur in Regali ac Majori Collegio Hispanorum gratia ab eximio Card. Albornotio erecto..., ad cod. [MS nella Biblioteca del Collegio. Le descrizioni relative ai codici civilistici sono state parzialmente edite a stampa da M.J. Pel<sup>á</sup>iez, <i>Cat<sup>á</sup>logo de la documentaci<sup>3</sup>n</i> p. 485-548, specialm. 504-528]; A. Garc<sup>á</sup>a y Garc<sup>á</sup>a e C. Piana, <i>Los manuscritos filos<sup>3</sup>fico-teol<sup>3</sup>gicos, hist<sup>3</sup>ricos y cient<sup>3</sup>ficos del Real Colegio de Espa<sup>3</sup>a de Bolonia</i>, Salmanticensis 14 (1967), p. 84 nu. 23 [Inventario del 1453], 87, 92, 94; A. Garc<sup>á</sup>a y Garc<sup>á</sup>a, <i>El legado de libros del cardenal Gil de Albornoz al Colegio de Espa<sup>3</sup>a de Bolonia</i>, Studi Senesi 84 (1972), p. 38 nt. 57</p>	<p><b>Unnumbered pages</b></p> <ul style="list-style-type: none"> <li>Copertina anteriore</li> <li>Copertina posteriore</li> <li>Foglio di guardia anteriore - recto</li> <li>Foglio di guardia anteriore - verso</li> <li>Foglio di guardia posteriore - recto</li> <li>Foglio di guardia posteriore - verso</li> <li>Piatto anteriore</li> <li>Piatto posteriore</li> </ul>
---	--	---

Figure 28. The interface represent the associated link to the Innerio Platform

By navigating through the identifiers, users can access the Innerio portal, allowing researchers to explore structured information related to specific codices in a more organized manner. Figure 28 demonstrates how the Innerio platform interfaces with the broader Mosaico system, making it easy to study and examine the manuscripts.

The linked open data platform allows scholars to deeply study the manuscripts and explore the structured material on the Innerio platform in detail. This organization of information enhances the accessibility and usability of the manuscripts, providing a comprehensive resource for legal, theological, and historical research.

### 5.5. SPARQL Endpoint and Querying

A robust SPARQL query interface is integrated within MELIORATE (Figure 29), offering access to predefined queries and allowing users to write custom ones. The endpoint is designed to support diverse querying needs from retrieving metadata for individual folios to analyzing gloss distribution and bibliographic references.

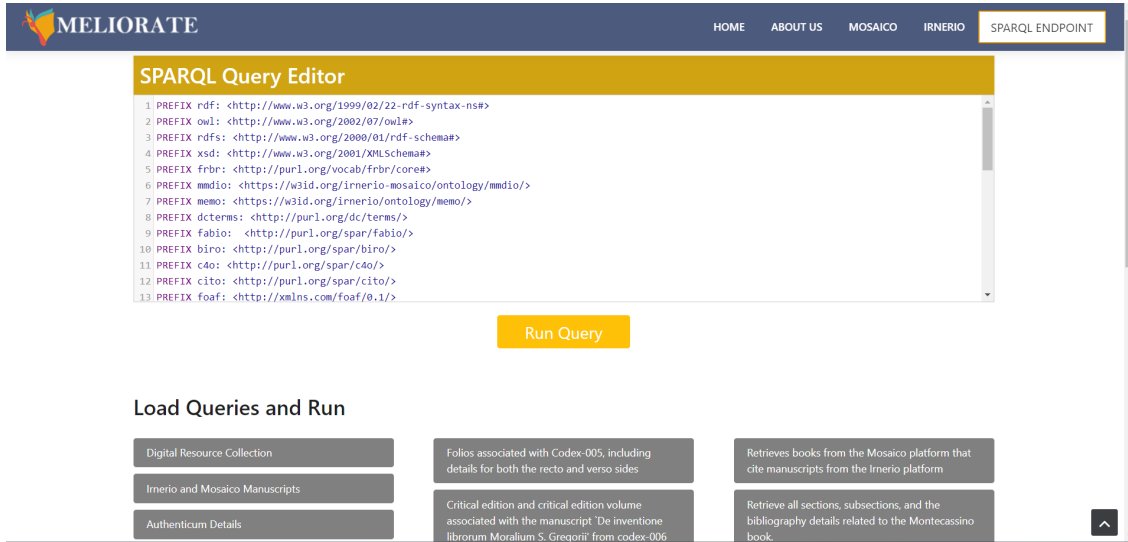


Figure 29. SPARQL editor with support for custom and predefined queries.

Query outputs are visualized in a tabular format with interactive links to deeper data (Figure 30), streamlining access to relevant information for advanced research.

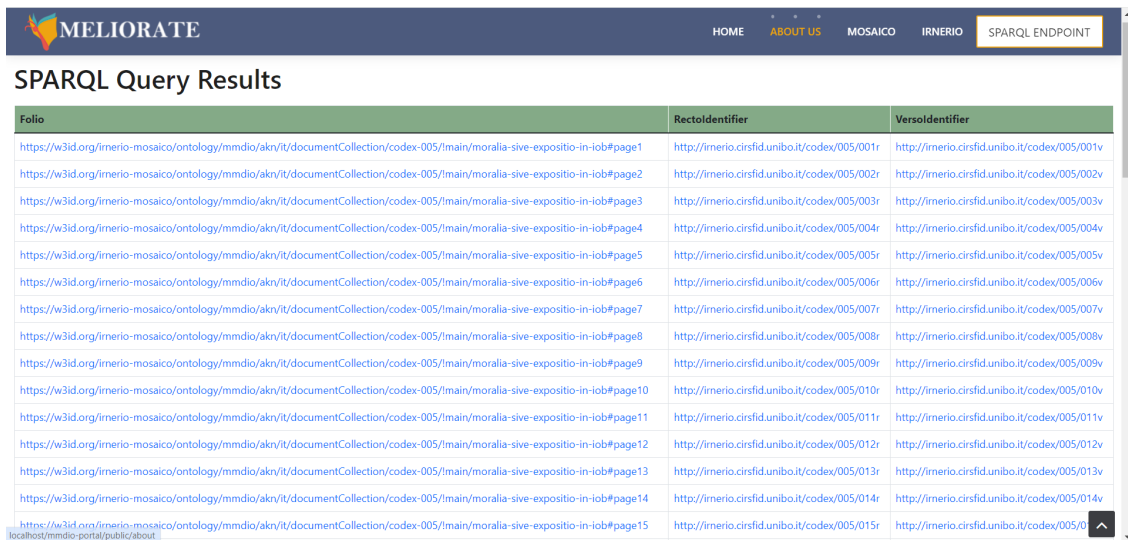


Figure 30. SPARQL query results showing structured manuscript metadata.

## 5.6. Technical Infrastructure and Architecture

Several tools and technologies were employed to build the MELIORATE portal, ensuring a smooth user experience and seamless interaction with the medieval manuscript knowledge graph. The backend was developed using the PHP Laravel framework, chosen for its

robust MVC architecture that supports rapid development and a clear separation of concerns. Laravel facilitated efficient handling of routes, controllers, views, user inputs, and query executions. For data storage, GraphDB served as the RDF triplestore backend, storing the manuscript data in RDF format. Laravel communicates with GraphDB through its SPARQL endpoint, allowing dynamic execution of queries against the manuscript knowledge graph. On the client side, the front end was built using HTML5, CSS3, and Bootstrap, providing a modern and responsive design compatible with multiple devices. Additionally, JavaScript and AJAX were integrated to enable asynchronous data loading, giving users the ability to submit SPARQL queries and view results interactively and efficiently.

### 5.7. Evaluation Against LOD Principles

MELIORATE is evaluated against key Linked Open Data (LOD) principles to assess its adherence to standards of openness, interoperability, and accessibility. The platform utilizes established semantic web vocabularies and ontologies, including FaBiO for bibliographic data, FRBR for modeling manuscript relationships, and MeMO (extended via MMDIO) for medieval manuscript metadata. It also integrates FOAF, Dublin Core, DoCO, CiTO, and BIRO to enhance metadata structure and citation management, ensuring a semantically rich and interoperable research environment.

A core LOD principle is the interlinking of related datasets. MELIORATE connects manuscripts from the Mosaico and Progetto Imerio platforms, facilitating cross-platform exploration. It uses persistent URIs for manuscript records, enabling reliable referencing. Bibliographic data links to historical editions such as Heimbach's *Authenticum*. However, integration with external LOD repositories like Europeana, Wikidata, and VIAF is currently limited. Expanding these links would improve interoperability and enrich the platform's metadata.

The platform features an open SPARQL endpoint supporting both predefined and custom queries. This allows users to access detailed manuscript metadata, folio-level information, and bibliographic references. However, SPARQL complexity remains a barrier for non-technical users, and the absence of a REST API limits broader application integration. Additionally, reliance on external sources (e.g., Mosaico and Imerio) may affect data availability and query performance.

A preliminary usability study involving legal historians and medieval scholars indicated strong appreciation for MELIORATE's structured data and discovery features. However, users reported difficulties with advanced SPARQL queries and noted some incomplete manuscript metadata, suggesting a need for external enrichment.

To enhance compliance with LOD principles, the following improvements are recommended:

- Integrate MELIORATE with external LOD repositories such as Wikidata, Europeana, and VIAF to expand metadata and connectivity.
- Develop a graphical query builder to improve accessibility for non-technical users.
- Provide a REST API to enable programmatic access and integration with third-party applications.
- Apply AI tools like Named Entity Recognition (NER) and automated annotation to enrich and complete metadata records.

The MELIORATE is a significant development in the field of medieval studies. It provides academics and scholars with integrated access to a diverse range of medieval manuscript materials, including historical context, physical features, textual data, and artistic elements. The platform follows the linked open data principles to improve accessibility and ensure that valuable historical pieces are easily discovered and linked. The platform's well-organized collections, such as Mosaico and Irnerio, allow researchers to examine diverse medieval texts. Thus, the platform serves as the foundation for research into medieval history, law, religion, and culture. Currently, this platform has revolutionized the medieval manual method of studying manuscripts while also opening up new avenues for future research, discovery, and cultural heritage preservation.

## 6. Discussion

This study describes the expanded methodology used to integrate disparate data about medieval manuscripts from the Progetto Irnerio and Mosaico systems. It expands on the previously developed Medieval Manuscript Digital Integration Ontology (MMDIO), which established the basic ontology representation of the historical, physical, textual, and artistic features of medieval manuscripts; the current study introduces new classes and object properties to meet the needs of supporting a polymorphic knowledge graph design. Rather than reinventing the ontological model, this work focuses on adapting and enriching the MMDIO structure to better handle evolving data structures, facilitate dynamic querying, and represent multi-contextual relationships between manuscripts, codices, and bibliographic resources.

One of the key contributions of this research is the implementation of a polymorphic knowledge graph that unifies and contextualizes manuscript data across both platforms. By converting and mapping diverse resources into RDF and aligning them with the extended MMDIO framework, the graph supports a more nuanced, flexible model in which manuscripts can appear in different historical, spatial, and scholarly roles without duplication. This approach enables researchers to explore interlinked data, track manuscripts across time and collections, and analyze glosses, foliation, annotations, and bibliographic references in depth. The polymorphic structure is particularly effective for legal and historical manuscripts, where a single entity can carry multiple interpretations across centuries.

To operationalize this semantic framework, the MELIORATE platform was developed as a public-facing Linked Open Data portal. The platform integrates the enhanced knowledge graph, offering researchers and scholars direct access to structured manuscript data through a user-friendly interface. MELIORATE supports both custom and predefined SPARQL queries, enabling users to perform detailed, cross-platform searches and retrieve results spanning different dimensions: textual, artistic, physical, and historical. The portal is designed to serve both technical and non-technical audiences, supporting interdisciplinary collaboration while also enhancing accessibility and reuse of manuscript data through open standards.

This article therefore, contributes a practical extension to the original MMDIO ontology, introducing a polymorphic knowledge infrastructure for handling complex, heterogeneous manuscript collections. It demonstrates how data integration can be achieved without restructuring entire ontologies, instead layering semantic extensions that reflect the evolving needs of manuscript representation and digital scholarship.

## 7. Conclusion and Future Work

This study advances the semantic modeling of medieval manuscripts by extending the previously developed Medieval Manuscript Data Integration Ontology (MMDIO). The ontology was enriched with new classes and properties to address a manuscript-specific use case from the Mosaico platform, allowing for the accommodation of more complex and heterogeneous metadata structures. Using the enhanced MMDIO, data from both the Mosaico and Progetto Irierio collections were transformed into RDF, ensuring semantic alignment and interoperability across diverse sources.

A key contribution of this research is the construction of a polymorphic knowledge graph, which enables multiple semantic perspectives and supports faceted exploration of manuscript metadata. This integration of heterogeneous data provides richer contextual information and facilitates deeper historical interpretation. To operationalize the model, the MELIORATE Linked Open Data platform was developed, offering semantic browsing, SPARQL querying, and structured access to support interdisciplinary research and collaboration.

The unified framework provides a flexible, scalable, and reusable infrastructure for integrating manuscript metadata across repositories. By adopting Linked Open Data standards and ontology-driven methods, the knowledge graph ensures consistency, long-term sustainability, and cross-disciplinary applicability. This work contributes to digital preservation, improves scholarly interoperability, and bridges the domains of cultural heritage, digital humanities, computer science, and legal history.

While the approach offers clear benefits, it also presents challenges. Manual metadata extraction remains time-consuming and limits scalability. Although the PKG structure is flexible, maintaining ontological consistency across new datasets will require careful schema evolution and robust alignment strategies. Furthermore, the reliance on expert modeling makes full automation difficult, especially for semantic distinctions that require human interpretation.

To address these limitations, future work will focus on developing automated pipelines for data extraction, entity resolution, and ontology alignment. Scheduled Extract, Transform, and Load (ETL) scripts could significantly reduce manual workload, while Large Language Models (LLMs) present promising opportunities to enrich metadata, uncover implicit semantic relationships, and support more intuitive natural language querying. Integration of MELIORATE into infrastructures such as the European Data Space for Cultural Heritage (Europeana) will further enhance the platform's visibility, interoperability, and long-term sustainability.

Overall, this research contributes a reusable semantic infrastructure for manuscript data integration, bridging cultural heritage, digital humanities, legal history, and computer science. By modeling manuscripts as evolving, multi-contextual objects within a polymorphic graph, this work supports new forms of interpretive analysis that respect the textual, historical, and physical complexity of manuscript traditions and opens new directions for digital preservation, scholarly engagement, and collaborative research.

## References

- [1] Amy JW Slowik. Collecting medieval manuscripts. *Collection building*, 28(3):126–130, 2009.
- [2] Lena Rohrbach. The power of book: Medial approaches to medieval nordic legal manuscripts. *Journal Name is Missing*, 2014.
- [3] Christopher De Hamel. The european medieval book. *The Book: A Global History*, pages 59–79, 2013.
- [4] Katrin Kogman-Appel et al. *A mahzor from Worms: art and religion in a medieval Jewish community*. Harvard University Press, 2012.
- [5] Emily Steiner and Candace Barrington. *The letter of the law: legal practice and literary production in medieval England*. Cornell University Press, 2002.
- [6] Sarah Bromberg. The social life of illumination: Manuscripts, images, and communities in the late middle ages ed. by joyce coleman, mark cruse, and kathryn a. smith. *The Catholic Historical Review*, 100(4):813–814, 2014.
- [7] Bryan C Keene. *Toward a Global Middle Ages: Encountering the World through Illuminated Manuscripts*. Getty Publications, 2019.
- [8] Elsa De Luca, Erika Loic, and Alicia Miguélez Caveró. Intermediality in medieval iberian manuscript cultures: methodological reflections on ongoing and future research. *Journal of Medieval Iberian Studies*, 14(1):1–14, 2022.
- [9] Ana Tourais, Conceição Casanova, and Catarina Fernandes Barreira. Filling the gap: New approaches to medieval bookbinding studies. *Journal of Medieval Iberian Studies*, 14(1):109–126, 2022.
- [10] Ricardo Córdoba de la Llave. Interdisciplinary exploration of medieval technical manuscripts from the iberian peninsula. *Journal of Medieval Iberian Studies*, 14(1):96–108, 2022.
- [11] Andrea Winkler. Digitized medieval manuscripts in the classroom: a project in progress. *The History Teacher*, 35(2):201–223, 2002.
- [12] Liv Ingeborg Lied. Digitization and manuscripts as visual objects: Reflections from a media studies perspective. *Digital Biblical Studies*, page 15, 2019.
- [13] John Palfrey. Cornerstones of law libraries for an era of digital-plus. *Law Libr. J.*, 102:171, 2010.
- [14] Faria Ferooz and Monica Palmirani. An ontological framework for integrating the heterogeneous medieval manuscript resources: A case study of progetto irnerio and mosaico. In *Knowledge Graphs in the Age of Language Models and Neuro-Symbolic AI*, pages 403–419. IOS Press, 2024.
- [15] Evelyn Dröge, Julia Iwanowa, and Steffen Hennicke. A specialisation of the europeana data model for the representation of manuscripts: The dm2e model. *LIBRARIES IN THE DIGITAL AGE*, page 41, 2014.
- [16] Eero Hyvönen, Esko Ikkala, Mikko Koho, Jouni Tuominen, Toby Burrows, Lynn Ransom, and Hanno Wijsman. Mapping manuscript migrations on the semantic web: A semantic portal and linked open data service for premodern manuscript research. In *The Semantic Web—ISWC 2021: 20th International Semantic Web Conference, ISWC 2021, Virtual Event, October 24–28, 2021, Proceedings 20*, pages 615–630. Springer, 2021.
- [17] Mikko Koho, Toby Burrows, Eero Hyvönen, Esko Ikkala, Kevin Page, Lynn Ransom, Jouni Tuominen, Doug Emery, Mitch Fraas, Benjamin Heller, et al. Harmonizing and publishing heterogeneous premodern manuscript metadata as linked open data. *Journal of the Association for Information Science and Technology*, 73(2):240–257, 2022.
- [18] Eduard Frunzeanu, Régis Robineau, and Elizabeth MacDonald. Bibliissima’s choices of tools and methodology for interoperability purposes= bibliissima: selección de herramientas y de metodología para fomentar la interoperabilidad. *CIAN-Revista de Historia de las Universidades*, 19:115–132, 2016.
- [19] Dominic Oldman and CRM Labs. The cidoc conceptual reference model (cidoc-crm): primer. *CRM Labs*, 5, 2014.
- [20] Ghazal Faraj and Andrés Micsik. Representing and validating cultural heritage knowledge graphs in cidoc-crm ontology. *Future Internet*, 13(11):277, 2021.
- [21] Maayan Zhitomirsky-Geffet and Gila Prebor. Toward an ontopedia for historical hebrew manuscripts. *Frontiers in Digital Humanities*, 3:3, 2016.
- [22] Jose Miguel Vieira and Arianna Ciula. Implementing an rdf/owl ontology on henry the iii fine rolls. In *OWLED*, 2007.
- [23] MN Zahila, Abdullah Noorhidawati, and MK Yanti Idaya Aspura. Content extraction of historical malay manuscripts based on event ontology framework. *Applied Ontology*, 16(3):249–275, 2021.
- [24] Sahar Aljalbout and Gilles Falquet. A semantic model for historical manuscripts. *arXiv preprint arXiv:1802.00295*, 2018.

- [25] Sebastian Barzaghi, Monica Palmirani, and Silvio Peroni. Development of an ontology for modelling medieval manuscripts: The case of progetto innerio. *Umanistica Digitale*, (9):117–140, 2020.
- [26] Monica Palmirani and Luca Cervone. A multi-layer digital library for mediaeval legal manuscripts. In *Digital Libraries and Archives: 8th Italian Research Conference, IRCDL 2012, Bari, Italy, February 9-10, 2012, Revised Selected Papers 8*, pages 81–92. Springer, 2013.
- [27] Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, Livio Robaldo, et al. Pronto: Privacy ontology for legal compliance. In *Proc. 18th Eur. Conf. Digital Government (ECDG)*, pages 142–151, 2018.
- [28] Judson Bandeira, Ig Ibert Bittencourt, Patricia Espinheira, and Seiji Isotani. Foca: A methodology for ontology evaluation. *arXiv preprint arXiv:1612.03353*, 2016.