Toward a Semantic Framework for Classical Chinese poetry: Multilingual and Decentralized Integration of East Asian Literary Heritage

Huang Yanjie[1,2], Jiang Hui[1],Chen Tao[2,3,*], Sha Tianming[4]

Abstract

Classical Chinese poetry, a classical form composed in literary Chinese, originated in China and spread across East Asia, including Japan, Korea, and Vietnam. It has played a key role in elite education, cultural diplomacy, and political communication. Despite its cultural significance, research on its semantic modeling remains limited, and digital representations often face issues like fragmentation and structural inconsistencies. This paper proposes a decentralized semantic infrastructure for Classical Chinese poetry as a multilingual resource for cultural heritage. The system is organized into three layers: (1) multilingual SKOS vocabularies to represent poetic concepts; (2) OWL ontologies that reuse BIBFRAME for bibliographic structure and FOAF for authorship representation; and (3) a distributed RDF resource layer supporting SPARQL 1.1 federated queries across poetic corpora. The infrastructure implements TEI standards for structuring poetic materials, including metadata and digitized text, while reserving interfaces for potential IIIF integration. Future work will focus on expanding multilingual vocabularies and integrating with cultural heritage platforms.

Keywords

Classical Chinese poetry, Semantic Web, Multilingual Semantic Modeling, Decentralized Ontology Modeling, Digital Humanities, East Asian Literary Heritage

1 Introduction

1.1 Classical Chinese poetry as a Cross-Cultural System

Classical Chinese poetry exemplifies the tension between formal unity and semantic diversity. Composed in Literary Chinese and originating in early imperial China, this poetic tradition spread widely across East Asia, including Japan, Korea, Vietnam, and the Ryukyu Islands. In Japanese scholarship it is commonly referred to as kanshi, while Korean and Vietnamese contexts similarly recognize poetry written in Classical Chinese as part of their literary heritage. Despite regional variations in cultural interpretation, poets across these regions adhered to shared conventions of meter, rhyme, and rhetorical form. Over time, however, local imagery, emotional registers, and philosophical perspectives were incorporated into this common framework, resulting in a culturally layered yet structurally coherent transregional poetic system.

Historically, Classical Chinese poetry served not only aesthetic purposes but also functioned as a medium for elite education, moral instruction, and the legitimation of political authority. In China,

[1]Foreign Language School, Sun Yat-sen University
[2]Sun Yat-sen University Digital Humanities Lab
[3]School of Information Management, Sun Yat-sen University
[4]School of Computer Science and Engineering, Sun Yat-sen University

Corresponding author:
Chen Tao
Sun Yat-sen University Digital Humanities Lab
School of Information Management, Sun Yat-sen University

Email: chent283@mail.sysu.edu.cn

poetry was a central component of the imperial examination system; in neighboring countries influenced by Chinese bureaucratic and Confucian models, poetic composition similarly became a marker of scholarly identity and cultural participation. Thus, Classical Chinese poetry operated as both a literary form and a vehicle for the transmission of ethical values and cosmological worldviews.

Despite its textual stability, the interpretation of Classical Chinese poetry has evolved into a multilayered, multilingual phenomenon. From the medieval period onward, local traditions developed practices of glossing, annotating, and interpreting poems using vernacular languages—such as Japanese kanbun kundoku, Korean idu notations, and Vietnamese script. In the modern era, while the primary texts remain in Literary Chinese, associated metadata, scholarly literature, and digital platforms now frequently appear in modern Chinese, Japanese, Korean, and English. This dynamic renders Classical Chinese poetry an ideal subject for multilingual semantic modeling, where surface linguistic uniformity masks underlying interpretative heterogeneity.

1.2   Challenges in Current Poetic Knowledge Systems

Despite its central role in East Asian literary traditions, the digital representation of Classical Chinese poetry remains highly uneven across different regions. Existing digital resources face significant challenges in semantic modeling, system interoperability, and infrastructure integration. Following the DELOS Digital Library Reference Model proposed by Candela et al. [1], which characterizes digital resources in terms of content organization, functionality, and architectural infrastructure, this study evaluates the digital maturity and semantic interoperability potential of existing Classical Chinese poetry resources through a qualitative survey of representative digital projects and infrastructures. Based on this analysis, current resources can be broadly categorized into four levels, representing a gradual progression from non-digitized materials

to semantically interoperable open data infrastructures.Each level reflects not only the technical status of the data but also the strategic differences in digital infrastructure development among various countries.

(L0) Undigitized Cultural Heritage: A large portion of Classical Chinese poetry remains scattered across historical manuscripts and rare books housed in libraries and archives, yet to be digitized.

(L1) Unstructured Digital Resources: Many resources exist only as scanned images or plain text transcriptions, lacking structured metadata and semantic annotations necessary for machine-readable access.

(L2) Structured but Closed Knowledge Systems: Some projects have developed internally complex knowledge graphs, but they often lack standard alignment and cross-platform interoperability, resulting in isolated information silos.

(L3) Standards-Compliant Open Resources: a small but growing number of resources adopt international standards, such as IIIF [2] for high-resolution image publication and authority URIs for historical figures from databases like CBDB [3] and JBDB [4], offering future potential for semantic linkage.

Overall, the digital landscape of Classical Chinese poetry is marked by structural heterogeneity, inconsistent standards, and limited accessibility. Addressing these issues calls for a unified and open semantic modeling framework capable of bridging diverse platforms and enabling cross-linguistic and cross-institutional integration.

2   Related Work

2.1   Semantic Web Approaches in Digital Humanities

Over the past two decades, Semantic Web technologies have become central to Digital Humanities (DH), providing flexible and interoperable frameworks for representing complex cultural knowledge. Core standards such as RDF [5], OWL [6], SKOS [7], and SPARQL [8] have been widely used in projects to structure, link, and query interdisciplinary, multilingual, and multi-institutional humanities data.

At the conceptual level, controlled vocabularies and Knowledge Organization Systems (KOS) [9] form the foundation for semantic modeling. SKOS, the W3C-recommended standard, enables the structured expression of multilingual concepts with hierarchical and semantic consistency. For example, the Getty Vocabulary Program's Art & Architecture Thesaurus (AAT) [10] uses SKOS and has been widely adopted in museums and archival collections, while Wikidata [11] offers a multilingual, community-maintained knowledge graph that supports global concept linking.

In addition, the Getty Linked Data infrastructure provides public SPARQL endpoints and cross-linguistic concept mappings. Although the present framework does not directly integrate these endpoints, it is designed to remain interoperable with large-scale cultural heritage vocabularies. Such compatibility may support future cross-regional and cross-linguistic semantic alignment, particularly within the broader Chinese Character Cultural Sphere.

OWL offers a more expressive semantic modeling mechanism, capable of representing complex relationships among entities such as people, texts, locations, and abstract concepts. Projects like Linked Jazz and Sharing Ancient Wisdoms (SAWS) [12] demonstrate the utility of Semantic Web standards in building structured, interlinked knowledge networks. Linked Jazz [13] employs OWL and the FOAF ontology to model the social connections among jazz musicians. In contrast, the SAWS project utilizes the FRBRoo [14] and CIDOC CRM [15] ontologies to represent intricate relationships between texts, such as authorship, transmission, and citation, in medieval manuscripts.

RDF-based data structures and SPARQL queries form the foundation of Linked Open Data (LOD) [16] architecture. The Europeana [17] project aggregates cultural metadata from Europe's institutions, while Pelagios [18] aligns geographic entities in classical texts with historical information. Finland's Sampo framework [19] showcases the use of OWL ontologies and RDF data in interactive knowledge platforms.

These projects highlight the practical value of Semantic Web technologies in DH, demonstrating that through ontology reuse, vocabulary alignment, and distributed querying, scalable multilingual knowledge infrastructures can be built. This paper applies these methods to the domain of Classical Chinese poetry, exploring its potential in symbolic expression, cross-linguistic alignment, and decentralized data connectivity.

2.2 Exploring the Integration of East Asian Cultural Heritage and Semantic Technologies

In recent years, China and Japan have begun incorporating Semantic Web technologies into the semantic modeling, knowledge publishing, and interoperability of cultural heritage. While most projects are still in early stages, there is growing regional progress, particularly in multilingual vocabulary construction, ontology design, and linked open data publication.

In vocabulary construction and concept alignment, several Chinese projects have attempted to semanticize traditional classification systems and convert them into SKOS-compliant models. For instance, the "Chinese Archival Classification Thesaurus" [20] represents traditional hierarchical knowledge in RDF, and the "Dunhuang Narrative Thesaurus" [21] uses SKOS to describe narrative categories in Dunhuang manuscripts, published as LOD to support semantic access. Researchers from Taiwan have also developed bilingual Chinese-English vocabularies for Chinese architecture and art [22], enriching multilingual term alignment in the East Asian context.

At the ontology modeling level, various OWL-based models have been proposed in China for cultural heritage scenarios, covering general heritage documents, ancient book bibliographies, and special document types such as ancient letters. The China Biographical Database (CBDB) [3] is one of the most widely used FOAF-based biographical knowledge bases, while the Shanghai Library has developed a genealogy ontology using

LOD technology [23]. The "Semantic Description Framework for Digital Images (SDFDI)" [24] proposed by Wang Xiaoguang integrates domain-specific vocabularies for interpreting visual resources like Dunhuang murals, and the "Ontology Service Center" platform [25] advances ontology reuse mechanisms.

In Japan, many institutions have established RDF publishing systems and IIIF (International Image Interoperability Framework) [2] infrastructure. IIIF is an open standard framework that enables interoperable access, presentation, and annotation of high-resolution digital images across distributed repositories [26].The National Diet Library and other institutions publish book metadata in RDF, while the "Union Catalogue Database of Japanese Texts" [27] publishes classical images in IIIF format, achieving high-resolution image interoperability. The IIIF Curation Platform [28] by the Center for Open Data in the Humanities (CODH) supports semantic annotation and reuse of digital resources.

Simultaneously, TEI (Text Encoding Initiative) [29] has been widely adopted in Japan, combining TEI markup with IIIF to provide multimodal representations of classical texts. Notable projects like the "Digital Tale of Genji" [30] and "Digital Engishiki" [31] use the TEI + IIIF model for multimodal access. The TEI Viewer [32] platform facilitates the visualization and public access of TEI texts.

Although these projects mainly focus on bibliographic, biographical, or image data, they provide a solid foundation for the semantic development of East Asian cultural heritage in terms of modeling methods, semantic interoperability, and multilingual alignment. While still fragmented in practice, these efforts offer technical pathways and inspiration for literary types like Classical Chinese poetry, which have not yet been systematically modeled, and hold potential for broader cross-cultural semantic expression through unified frameworks.

2.3 Current Status and Limitations of Digital Resources for Classical Chinese poetry

Despite its profound influence in East Asian cultural traditions, Classical Chinese poetry's digital resources remain fragmented, structurally inconsistent, and weak in semantic modeling. These resources can be broadly categorized into the following types, each revealing significant shortcomings in semantic organization and cross-lingual interoperability, underscoring the need for a unified semantic infrastructure.

(1) Undigitized or Image-Only Resources. A large portion of Classical Chinese poetry, particularly historical collections and manuscripts in Japan, Korea, and Vietnam, remains undigitized or exists only as scanned images, lacking transcribed texts and structured metadata. These image materials are typically stored on local platforms without adopting standards like IIIF. An exception is the Union Catalogue Database of Japanese Texts project, which offers IIIF-compliant high-resolution images with basic metadata.

(2) Unstructured or Minimally Structured Text Resources. The Chinese Text Project [33] and Taiwan's "Hanquan" database [34] provide substantial Classical Chinese poetry texts, but these are generally presented in plain text or HTML format without TEI encoding or semantic annotations. As a result, they lack structured metadata for semantic querying or multi-source alignment. The GitHub-hosted chinese-poetry project [35] introduces some improvements in structuring, but still lacks semantic annotations, version control, and interlinking mechanisms.

(3) Poetry Knowledge Graph Systems with Limited Interoperability. Several Chinese scholars [36] have attempted to build poetry knowledge graphs, representing relationships among themes and poets. These systems enhance internal retrieval capabilities but rely on closed, locally maintained ontologies and lack RDF interfaces, limiting interoperability with Semantic Web standards like SKOS or OWL.

(4) Experimental Semantic Modeling of Classical Chinese poetry Using RDF/OWL. Several researchers have explored RDF/OWL-based semantic modeling of classical Chinese poetry. Zhou, Hong, and Gao [37] proposed an ontology for Tang poetry designed to support domain knowledge services. Their framework structures poets, works, historical contexts, and related literary entities, demonstrating the feasibility of ontology-driven knowledge organization within a specific literary corpus.

Similarly, Cong and Takaku [38] developed an OWL ontology grounded in BIBFRAME [39], focusing primarily on bibliographic modeling and metadata interoperability. Their approach contributes to aligning classical poetry resources with library-oriented linked data standards.

While these studies provide important steps toward formal semantic modeling, their primary focus lies in resource organization and bibliographic structuring. In contrast, the present study proposes a multi-layered semantic architecture that distinguishes lexical alignment, interpretive abstraction, and poem-level relational modeling, enabling multilingual interoperability while preserving interpretive flexibility.

(5) Cross-Lingual and Interpretive Gaps Although numerous Classical Chinese poetry works have been translated into modern Chinese, Japanese, and English, there is currently no structured semantic framework that explicitly links original texts, translations, annotations, and adaptations. Relationships among these resources remain largely unmodeled, limiting multilingual retrieval, comparative literary analysis, and cross-cultural interpretation.

(6) Data-Driven Approaches without Reusable Semantic Infrastructure Recent research by Li and Chen [40] has explored the use of deep learning and large language models (LLMs) to identify symbolic and metaphorical elements in classical poetry. While these approaches contribute to pattern discovery and content-level annotation, they generally do not provide an explicit, reusable semantic structure that enables interoperability

across systems. Core poetic concepts such as imagery, metaphor, symbolism, and poetic themes remain implicit or ad hoc in existing models, and have not yet been effectively represented as machine-readable, multilingual semantic resources. This limitation poses a fundamental challenge to advanced knowledge modeling, cross-resource alignment, and semantic reasoning in the digital representation of Classical Chinese poetry.

In summary, existing digital resources and semantic modeling efforts for Classical Chinese poetry exhibit limitations across multiple dimensions: digitization, structural standardization, multilingual linkage, and interpretive semantic modeling. These gaps highlight the need for a unified, multi-layered semantic infrastructure capable of integrating lexical alignment, interpretive abstraction, and cross-cultural interoperability.

The following sections will detail the design principles and components of the proposed semantic architecture, which addresses the fragmentation issues across semantic, structural, and institutional dimensions.

3 Overview of the Semantic Infrastructure Design

In the field of Digital Humanities, scholars such as Chen, Yang, Xia, and Su [41] have noted that, in the face of the heterogeneous and dispersed nature of traditional textual resources, decentralized and distributed interconnection frameworks are considered an effective organizational model. Drawing on this insight, the semantic infrastructure for Classical Chinese poetry proposed in this paper adopts a three-layer architecture, aiming to address the issues of inconsistent standards and weak interoperability currently observed in Classical Chinese poetry resources. The framework supports cross-platform and cross-linguistic federated integration through the construction of multilingual conceptual vocabularies, ontology modeling, and a distributed resource layer.

The core of this design follows key principles of the Linked Open Data (LOD) framework as articulated by Berners-Lee [42]: using resolvable

URIs, reusing standard vocabularies, and supporting distributed semantic querying. Unlike traditional monolithic structures built around specific corpora, the proposed model serves as a generalized semantic framework, designed to accommodate heterogeneous resources and enable interoperability across platforms and languages. The main components include:

(1)Concept Layer —Multilingual conceptual vocabularies are constructed using SKOS to represent culturally sensitive poetic concepts such as imagery, themes, and rhetorical devices. Cross-linguistic alignment is achieved through explicitly defined mappings.

(2)Ontology Layer —OWL is employed to model structured relationships within the poetry domain. BIBFRAME is reused for bibliographic entity relationships, while FOAF is utilized for modeling poet and author identity information.

(3)Resource Layer —Instance data (e.g., poems, poets, poetry collections) are constructed and deployed across multiple triple stores, with cross-database connectivity enabled through SPARQL 1.1 federated querying.

The layered architecture ensures modularity, reusability, and scalability in modeling. The Concept Layer addresses the consistency of semantic vocabularies across languages; the Ontology Layer supports the structured representation of complex relationships among literary entities; and the Resource Layer enables decentralized data publication and access.

Figure 1 illustrates the overall system architecture and the relationships among the layers, demonstrating how the model supports multilingual and cross-repository semantic retrieval and access.

3.1 Namespace Declaration, IRI Design, and Terminological Clarification

To ensure structural transparency and formal consistency in the presentation of RDF examples, all namespace prefixes used throughout this paper are declared in advance in this section. The framework adopts a layered namespace strategy that distinguishes ontology definitions, conceptual vocab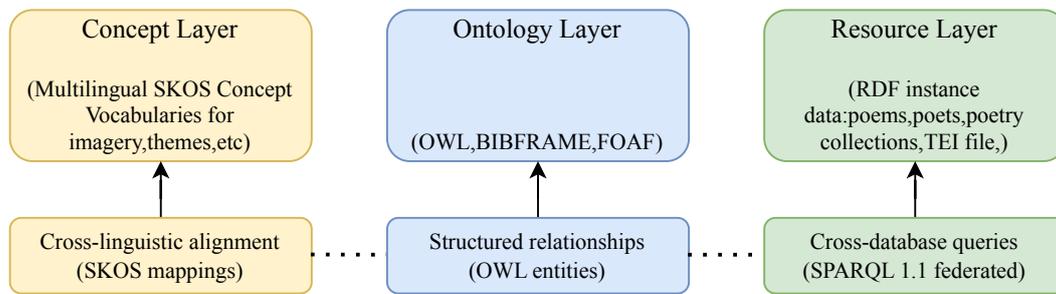ularies, and instance-level resources.For readability in the RDF examples presented in this paper, the prefix declarations adopt simplified resource namespaces.

Namespace Declarations Used in This Framework

```
1
2  @prefix skos:
   ↪    <http://www.w3.org/2004/02/skos/core#>
   ↪    .
3  @prefix bf:    <http://id.loc.gov/⌋
   ↪    ontologies/bibframe/>
   ↪    .
4  @prefix foaf:  <http://xmlns.com/foaf/0.1/>
   ↪    .
5  @prefix rdfs:
   ↪    <http://www.w3.org/2000/01/rdf-schema#>
   ↪    .
6  @prefix xsd:
   ↪    <http://www.w3.org/2001/XMLSchema#> .
7
8  @prefix hp:
   ↪    <https://w3id.org/hanpoetry/ontology#>
   ↪    .
9  @prefix concept:
   ↪    <https://w3id.org/hanpoetry/concept/> .
10
11 @prefix poem:         <https://w3id.org/⌋
   ↪    hanpoetry/resource/poem/>
   ↪    .
12 @prefix knowledge:    <https://w3id.org/⌋
   ↪    hanpoetry/resource/knowledge/>
   ↪    .
13 @prefix person:       <https://w3id.org/⌋
   ↪    hanpoetry/resource/person/>
   ↪    .
14 @prefix collection:   <https://w3id.org/⌋
   ↪    hanpoetry/resource/collection/>
   ↪    .
15 @prefix manifestation: <https://w3id.org/⌋
   ↪    hanpoetry/resource/manifestation/>
   ↪    .
16 @prefix range:        <https://w3id.org/⌋
   ↪    hanpoetry/resource/range/>
   ↪    .
17 @prefix canvas:       <https://w3id.org/⌋
   ↪    hanpoetry/resource/canvas/>
   ↪    .
18 @prefix teires:       <https://w3id.org/⌋
   ↪    hanpoetry/resource/tei/>
   ↪    .
19
```

Concept Layer	Ontology Layer	Resource Layer
(Multilingual SKOS Concept Vocabularies for imagery,themes,etc)	(OWL,BIBFRAME,FOAF)	(RDF instance data:poems,poets,poetry collections,TEI file,)
Cross-linguistic alignment (SKOS mappings)	Structured relationships (OWL entities)	Cross-database queries (SPARQL 1.1 federated)

Figure 1. Overview of the Semantic Framework for Classical Chinese poetry

In this design, three principal namespace layers are separated.

First, the ontology namespace (hp:) defines OWL classes and properties that form the structural schema of the framework, including entities such as hp:Poem, hp:Knowledge, and hp:Collection.

Second, the conceptual namespace (concept:) contains SKOS-based multilingual lexical concepts. This layer supports hierarchical organization and cross-linguistic alignment while remaining distinct from ontology-level semantic modeling.

Third, instance-level resources are organized under structured resource namespaces (e.g., /poem/, /person/, /collection/). This hierarchical URI design provides stable and semantically meaningful identifiers for poems, persons, collections, canvases, and related entities.

In the deployed implementation, certain resource URIs may include intermediate dataset identifiers (e.g., "tang" or "heian") in order to distinguish independently maintained datasets. These identifiers appear as path segments within the URI structure and denote the dataset to which a resource belongs, such as Tang or Heian poetry corpora. While they provide dataset-level organization, all resources remain within the unified hanpoetry URI space rather than forming independent namespaces. The relationship between these dataset partitions and the decentralized data architecture is further explained in Section 3.4.

This separation ensures modularity, semantic clarity, and extensibility, and supports potential deployment in distributed or federated Semantic Web environments.

It is also necessary to clarify the rationale behind the namespace identifier "Hanpoetry." The term is retained as a technical label for the ontology and resource structure, and it does not imply restriction to the Han dynasty.

Across East Asia, poetry written in Literary Chinese has historically been referred to using terms derived from the character "Han" (漢). In Japanese scholarship, such poetry is commonly called kanshi (漢詩), and in Korean contexts hansi (漢詩). In Chinese academic usage, while poetry composed within China is generally described simply as "ancient poetry" or "Classical poetry," works composed outside China in Literary Chinese are often designated as "域外漢詩." Although this expression is typically rendered in English as "Chinese poetry abroad" or similar formulations rather than "Han poetry," the Chinese designation itself belongs to the same lexical family as "漢詩," reflecting the continued appearance of the character "Han" in certain scholarly naming conventions.

These usages indicate that the term "Han poetry" has formed part of the scholarly vocabulary associated with Literary Chinese poetic traditions across different regional contexts. Within this framework, "Hanpoetry" functions solely as a concise and stable technical identifier within the Semantic Web architecture.

For conceptual precision in academic discourse, the descriptive terminology used throughout the narrative text of this paper remains "Classical Chinese poetry." The namespace label serves only as a compact internal identifier within the ontology and does not alter the scholarly terminology adopted in the discussion.

3.2 Concept Layer: Building a Multilingual SKOS Conceptual Vocabulary System for Classical Chinese poetry

To support the semantic alignment across languages, multilingual interoperability, and structured organization of Classical Chinese poetry knowledge, this study constructs a multilingual SKOS-based conceptual vocabulary system. As a lightweight semantic layer, this system bridges abstract concept modeling and resource-level representation, enabling cross-linguistic concept alignment, decentralized resource management, and ontology integration.

Existing approaches to Linked Data often neglect the linguistic dimension and fail to systematically address multilingual data modeling and publishing [43]. Given the symbolically rich and concise nature of Classical Chinese poetry, where words often carry multiple meanings in different cultural contexts, this study introduces a multilingual support mechanism in the conceptual design. All concepts are modeled as skos:Concept instances and are labeled with preferred labels (skos:prefLabel) in Chinese, English, and Japanese, ensuring semantic consistency and multilingual interoperability.

3.2.1 Vocabulary Structure Design
The SKOS-based conceptual vocabulary for Classical Chinese poetry is designed to support structured representation and multilingual interoperability. Within the proposed framework, the vocabulary layer provides a lightweight semantic structure that complements the ontology layer by organizing interpretable concepts used in the analysis of poetic texts.

The vocabulary categorizes concepts into four primary types: knowledge concepts, character concepts, poem title concepts, and poetry collection concepts. In this framework, knowledge concepts refer to interpretive semantic elements commonly used in the analysis of Classical Chinese poetry, including imagery, emotions, spatial settings, temporal references, and thematic motifs. These concepts function as semantic anchors that allow poems to be associated with interpretable literary elements.

Character concepts represent poets and historically significant literary figures. These concepts enable consistent identification of individuals and support multilingual labeling across different scholarly contexts.

Poem title concepts reflect a distinctive feature of Classical Chinese poetic practice. In many literary contexts, such as poetic gatherings or court events, multiple poets may compose poems under the same title. In addition, variations in historical transmission often result in different recorded versions of poem titles across manuscripts and editions. For this reason, poem titles are represented as conceptual resources within the vocabulary layer to facilitate intertextual comparison and cross-poem indexing.

Poetry collection concepts represent historically recognized poetry compilations, such as Quan Tang Shi (Complete Tang Poems), which serve as important organizational structures for large poetic corpora.

These concept categories are hierarchically structured using SKOS properties such as skos:broader and skos:narrower to establish semantic relations between concepts.

To support collaborative and distributed data management, the vocabulary system follows a modular design that distinguishes between a shared core vocabulary and dataset-specific local vocabularies. The core vocabulary contains domain-level concepts that form the stable semantic backbone of the framework, including widely used interpretive categories for Classical Chinese poetry.

In addition to this shared conceptual layer, individual datasets may maintain their own local vocabularies that introduce corpus-specific concepts or specialized annotations required for particular research contexts. For example, datasets corresponding to different poetic corpora may extend the vocabulary with additional imagery terms, thematic motifs, or historical references that are specific to their textual collections.

This modular vocabulary structure is consistent with the decentralized dataset architecture described in Section 3.4. While datasets may maintain independent local vocabularies, all vocabularies follow

the same SKOS modeling principles and remain compatible with the shared ontology and conceptual framework. As a result, vocabulary modules from different datasets can be combined or reused without compromising semantic interoperability.

The SKOS vocabularies used in this study are publicly available in RDF format through the project repository.

Unlike traditional SKOS vocabularies that emphasize rich semantic mappings such as skos:related or skos:exactMatch, this system adopts a minimalist approach in which semantic integration is achieved through explicit links between poems and knowledge concepts. Each poem is modeled as an RDF instance of the class hp:Poem defined in the ontology layer and is associated with relevant interpretive concepts through the property hp:hasKnowledge. In contrast, poem titles in the SKOS vocabulary are represented as conceptual labels rather than as poem entities. The detailed modeling approach and integration mechanism are discussed in Section 4.2.

3.2.2 System Validation and Visualization

To demonstrate the SKOS conceptual vocabulary system and validate its effectiveness, we utilized a simple SKOS Tree Viewer designed to visually inspect the constructed vocabulary file. The viewer is used not only to verify data integrity but also to showcase how the SKOS-based system facilitates multilingual semantic representation. The SKOS Tree Viewer is available open-source, allowing researchers to access and use it for vocabulary validation and concept exploration.

As a representative example, we conducted a keyword search for "Moon" to illustrate the system's multilingual labeling mechanism and conceptual organization. The concept hpc:Moon is categorized under celestial objects and is labeled as "月亮" in Chinese, "Moon" in English, and "つき" in Japanese within the SKOS framework. Its semantic relations include broader terms related to natural imagery, demonstrating the hierarchical structuring supported by SKOS.

Although these labels refer to the same celestial object, literary interpretations of the moon motif vary across cultural contexts. In Chinese classical poetry, the moon frequently evokes homesickness and reflections on temporal transience. In Korean traditions, it is linked to agricultural abundance, feminine symbolism, and seasonal ritual practices [44]. In Japanese contexts, both religious–mythological interpretations—associating the moon with immortality beliefs, celestial mythology, and the conception of the "other world" (tokoyo) [45]—and Heian literary traditions, where the moon appears in relation to impermanence, death, and emotional reflection [46], demonstrate distinct interpretive emphases.

Although many of these symbolic resources were historically transmitted through the Sinitic cultural sphere, their thematic configurations were locally reinterpreted within different literary and religious frameworks. Accordingly, the SKOS layer in our framework functions strictly as a lexical alignment mechanism. It ensures cross-linguistic referential consistency without encoding culturally specific symbolic meanings as intrinsic properties of the concept.

Interpretive variation is therefore not fixed at the conceptual level but is reflected through poem-level associations discussed in Section 4.2. This design allows multilingual interoperability while avoiding semantic homogenization.

As shown in Figure 2, the SKOS Tree Viewer presents the concept URI, multilingual preferred labels, rdfs:comment, and broader semantic relationships. The visualization clearly highlights how multilingual labels are consistently linked to the same concept, reflecting the system's cross-linguistic consistency. The SKOS Tree Viewer is available as a supporting tool on the open data platform, enabling researchers to explore, validate, and utilize the vocabulary system for concept retrieval and analysis.

3.3 Ontology Layer: Design Principles and Functional Structure

The ontology layer proposed in this paper aims to provide a semantic framework and multimodal
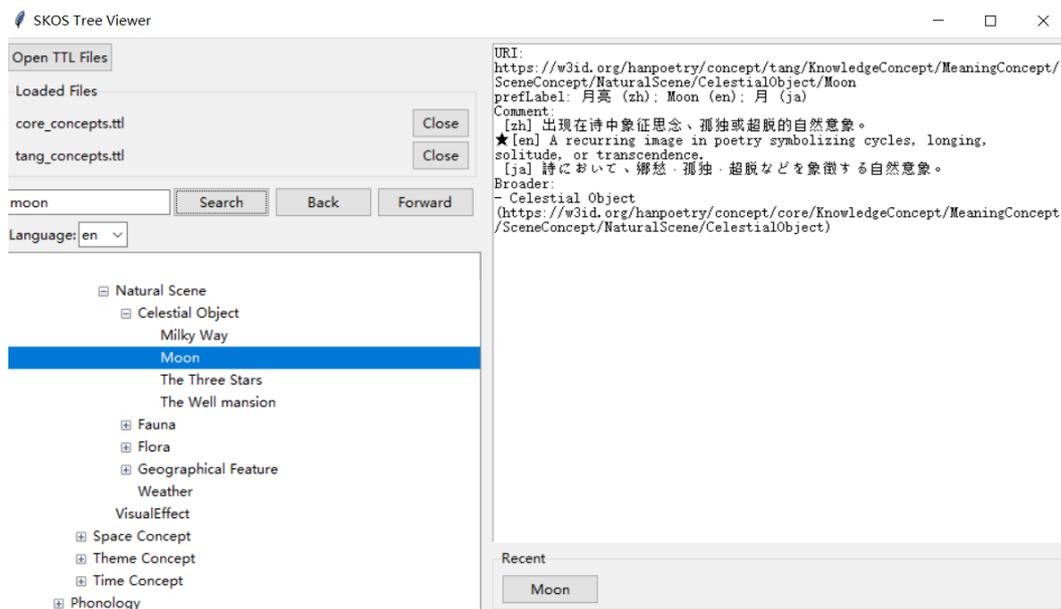
Figure 2. Visual representation of the SKOS conceptual vocabulary for Classical Chinese poetry: The SKOS Tree Viewer displays the concept "Moon" with its URI, multilingual preferred labels, rdfs:comment, and broader semantic relationships.

integration for Classical Chinese poetry resources. The model is built around four core entities: Knowledge, Poem, Person, and Collection, as illustrated in Figure 3. Among these, the Person class is aligned with the FOAF (Friend of a Friend) vocabulary, which provides a widely used model for representing persons in Linked Data environments. In our framework, FOAF is referenced at the vocabulary level to maintain compatibility with existing semantic web standards while keeping the ontology lightweight. The Person class links poems to related individuals such as poets and critics.

In representing Chinese personal names, this framework follows ISO 7098:2015 (Information and documentation —Romanization of Chinese)[47], which retains the conventional surname–given name order in Romanized form. For example, in the case of Li Bai, "Li" corresponds to the family name and "Bai" to the given name.

Meanwhile, the Knowledge and Poem classes form the core of the ontology representation, while the Collection class is responsible for managing the physical representations of these resources. In this framework, poem instances are linked to knowledge entities through semantic properties, while knowledge entities are aligned with corresponding SKOS concepts introduced in Section

3.2. The detailed linking mechanism between these layers is described in Section 4.2.

3.3.1 Poem Class: Independent Bibliographic Representation In this model, the Poem class is conceptually aligned with the bf:Work concept defined in the Bibframe model, treating each poem as an abstract, bibliographically independent entity. Although poems are often included in poetry collections for editing or historical archiving, they retain their independent status in literary studies and academic citations. Therefore, the Poem class is defined as a subclass of bf:Work. Individual poem instances are assigned unique identifiers, allowing metadata such as title, author, genre, and background information to be represented in RDF, together with related abstract knowledge such as themes, imagery, emotions, and rhetorical devices.

The Poem class itself does not directly contain text or visual content but serves as a semantic anchor linking to multiple physical manifestations, including translated versions, annotated editions, and digital copies. The Poem class can also be linked to multiple collections, appearing in different versions or translations, thus maintaining semantic consistency across variants while preserving the poem's integrity as a work.
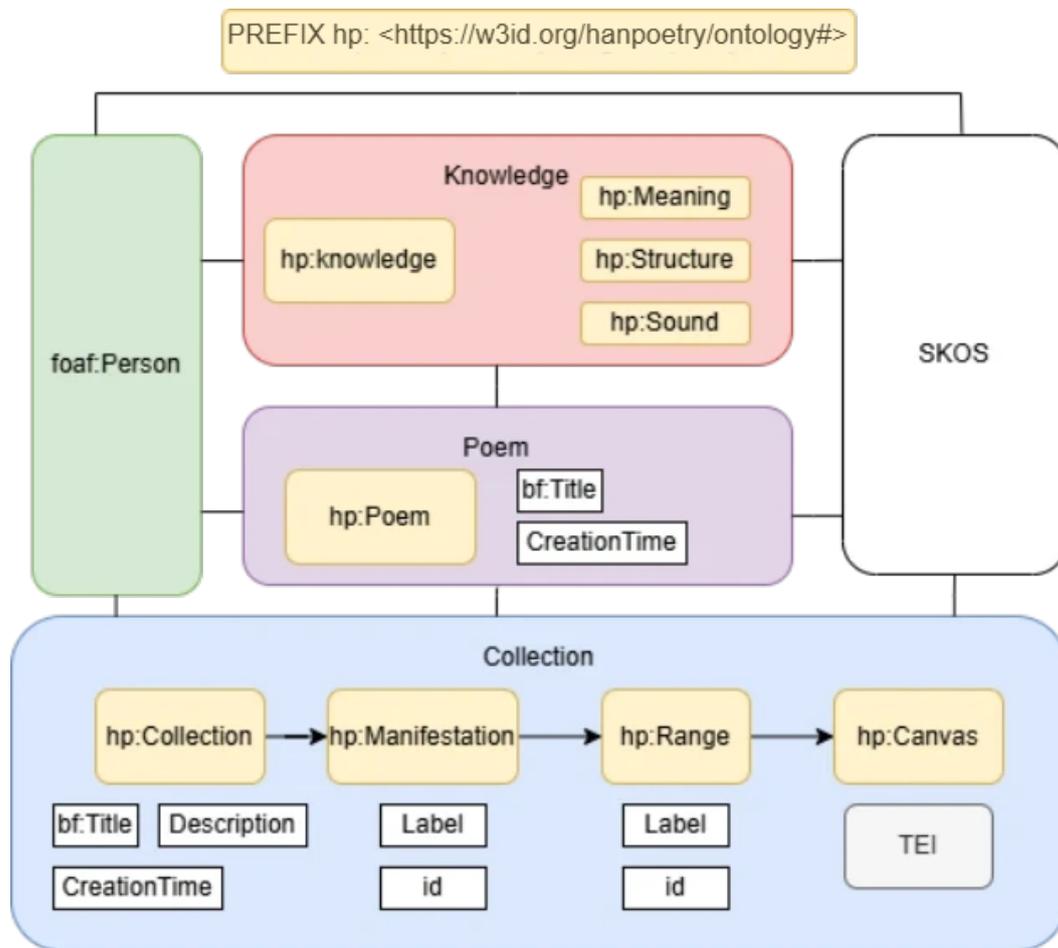
Figure 3. Core Entities and Functional Structure of the Ontology Layer for Classical Chinese poetry

3.3.2 Knowledge Class: Constructing Interpretive Knowledge Entities To facilitate the semantic representation of interpretive elements in Classical Chinese poetry, this model introduces a dedicated class, hp:Knowledge. Knowledge entities represent abstract interpretive units commonly used in literary analysis, such as imagery, emotional expression, rhetorical devices, and thematic orientations.

This class is aligned with the SKOS conceptual vocabulary introduced in Section 3.2. Each knowledge entity is associated with a corresponding SKOS concept through the property hp:hasSKOSConcept, which enables the alignment between the ontology layer and the lexical concept layer while preserving multilingual concept representation.

Rather than introducing separate ontology properties for different interpretive categories (e.g., imagery, emotion, or theme), the model adopts a unified linking mechanism in which poems are associated with knowledge entities through the property hp:hasKnowledge. In this design, interpretive categories such as imagery, emotion, and theme are represented as different instances of hp:Knowledge rather than as distinct relational properties.

This modeling strategy maintains a clear separation between lexical concepts and interpretive abstractions. SKOS concepts provide multilingual conceptual alignment, while hp:Knowledge entities represent specific interpretive units that can participate in different semantic configurations across poems.

Through the property hp:hasKnowledge, poems can be linked to multiple knowledge entities, forming a many-to-many relationship between poems and interpretive elements. This structure enables various analytical operations in Digital Humanities research, including semantic clustering of poems, co-occurrence analysis of interpretive elements,

and comparative studies of imagery, emotion, and thematic structures across different poetic traditions.

3.3.3 Collection Class: Organizing and Presenting Poetry Collections

In the modeling of collections, this paper adopts the framework design principles of the IIIF (International Image Interoperability Framework) Presentation API to address the issues of multimodal representation and structured organization of resources. The framework introduces four key entities: Collection, Manifestation, Range, and Canvas, which, through layered modeling and resource fragmentation, enable flexible organization and multimodal presentation of poetry collection resources.

(1)Collection: Represents a collection of poetry resources, such as single-volume or multi-volume poetry collections. Translated versions of poetry collections are also treated as independent collections, enabling the management of multilingual versions and related resources (such as annotations and revisions) while maintaining their links as independent entities.

(2)Manifestation: Corresponds to specific versions or physical representations of a poetry collection (e.g., translations, annotated editions, or printed versions). This entity models the various manifestations of a collection, ensuring flexibility in managing different versions and forms.

(3)Range: Describes the hierarchical structure within a collection, such as volumes, chapters, or sequences of poems, supporting structured navigation within the collection.

(4)Canvas: The smallest unit of display within a collection, linking the abstract knowledge of a poem (Poem class) to its physical manifestations (e.g., TEI texts, images, or multimedia resources). Each canvas aggregates multiple multimodal resources, such as poetry texts, IIIF image fragments, audio files, and semantic annotations, providing a comprehensive display.

This design allows each poem to appear in multiple collections, with different versions or manifestations of each poem represented by different canvases. The semantic integrity of the poem is maintained through its connection to the Poem class, while specific manifestations and performances (such as translations and annotations) are represented through the Manifestation and Canvas entities.

3.4 Resource Layer: Integration of RDF and TEI with Federated Querying

The resource layer plays a crucial role in the management of Classical Chinese poetry data by integrating RDF triples with TEI files, ensuring efficient storage, retrieval, and querying of poetry resources while consolidating textual data with related semantic metadata. By utilizing Virtuoso [48] and SPARQL queries, this layer enables decentralized data management and cross-database querying.

Due to current technological limitations, image resources are not yet integrated into the system. However, the architecture design has already reserved space for future visual data integration, thus expanding the multimodal capabilities of the platform.

3.4.1 Integration of RDF and TEI

In the resource layer, RDF triples are used to represent core entities within the system, such as poems, authors, and collections, along with their relationships and metadata. Each poem is represented as an RDF resource and assigned a unique URI, linking it to relevant metadata (such as author, genre, and creation date) and conceptual information (such as themes, emotions, and rhetorical devices).

TEI files contain the actual textual content of the poems, structured using elements such as <div>, <lg>, <l>, and <seg>. These TEI files are linked to the corresponding RDF resources through unique URIs, ensuring that the poem texts are integrated with the semantic metadata stored in the RDF database, as illustrated in Figure 4. The detailed mechanism for establishing these TEI–RDF connections is described in Section 4.4.

This integration allows semantic queries to be executed on both the textual data and related concepts and bibliographic metadata.

3.4.2 Decentralized Data Management and Semantic Access

The data architecture of this framework
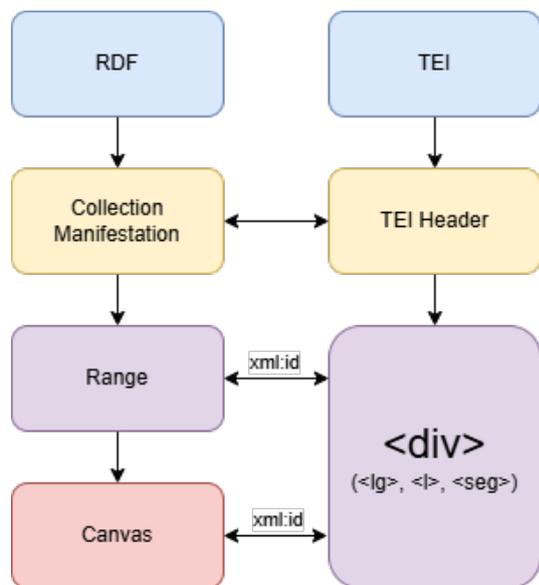
Figure 4. Mapping Between RDF Resources and TEI Files in the Collection

is designed to support decentralized publication while maintaining semantic interoperability across independently maintained datasets. In the current implementation, the RDF data generated by the project are distributed as independent RDF files through the project repository rather than through a continuously running public SPARQL endpoint. These datasets can be loaded into standard triple store systems such as Virtuoso, where multiple RDF graphs can be aggregated into a unified queryable environment.

Within the URI design introduced in Section 3.1, certain resource identifiers contain intermediate dataset segments that indicate the dataset to which a resource belongs. For example, a knowledge concept in the Heian corpus may be identified by a URI such as `https://w3id.org/hanpoetry/resource/heian/knowledge#ImageryMoon`. In this structure, the segment "heian" denotes the dataset, while "knowledge" indicates the resource type within the shared resource hierarchy. Similar patterns are used for other corpora, such as Tang poetry datasets.

Although these datasets are physically stored and published as separate RDF modules, all resources remain within the unified hanpoetry URI space and share the same ontology definitions and conceptual vocabularies. This design allows datasets to remain independently maintained by different institutions or projects while preserving semantic interoperability through shared identifiers and standardized data models.

When deployed in compatible triple store environments, these distributed RDF datasets can be integrated into a unified RDF graph, enabling semantic queries that span multiple poetry corpora. In this way, decentralized dataset publication and centralized semantic querying can coexist within the same Linked Data infrastructure.

In addition to resource datasets, the vocabulary system follows the same modular organization. Alongside the shared core vocabulary, individual datasets may maintain additional vocabulary modules that introduce corpus-specific concepts or annotations. Because these vocabularies follow the same SKOS structure and remain aligned with the shared ontology layer, they can be seamlessly combined when datasets are integrated. This modular design enables both resource data and conceptual vocabularies to participate in the same decentralized yet interoperable data architecture, as illustrated in Figure 5.
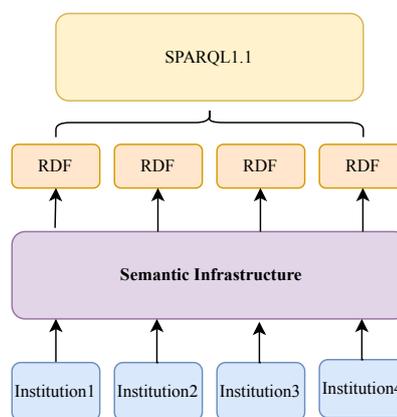
Figure 5. Federated Query Mechanism for Decentralized Poetry Data Integration

4 Semantic Modeling of Classical Chinese poetry Entities

This chapter presents an ontology-driven model of Classical Chinese poetry, based on the semantic infrastructure introduced in Section 3. Unlike the conventional data flow, which moves from collections to poems to concept extraction, this model adopts

a semantic-first approach. In this approach, the poem serves as the central unit of interpretation and integration. Each semantic layer (knowledge, structure, representation, and language) revolves around the poem as an anchor point, as shown in Figure 6.

In this chapter, the namespace declarations introduced in Section 3.1 are consistently adopted to simplify the representation of URIs in the RDF examples.

To improve readability and maintain consistency in the examples presented in this section, we briefly clarify how concept URIs are represented in the RDF snippets. In the underlying RDF dataset, concept identifiers often follow hierarchical URI structures that reflect the internal classification of the conceptual vocabulary. For example, the concept representing "Moon" is defined using a full URI such as:

```
Classical Chinese poetry Semantic Model Example

1  <https://w3id.org/hanpoetry/concept/↵
   ↪  tang#KnowledgeConcept/MeaningConcept/↵
   ↪  SceneConcept/NaturalScene/↵
   ↪  CelestialObject/Moon>
2     a skos:Concept.
```

This hierarchical URI structure preserves the semantic organization of the conceptual vocabulary and is maintained in the actual RDF dataset.

However, repeatedly displaying such long URIs in the textual discussion would significantly reduce readability. Therefore, for the purpose of illustration in this article, we adopt a simplified notation when presenting examples. In the following RDF snippets, the same concept may be represented using an abbreviated identifier such as:

```
Classical Chinese poetry Semantic Model Example

1  concept:moon a skos:Concept.
```

This abbreviated form is used only as a shorthand representation in the examples shown in this paper. It should not be interpreted as the literal expansion of the concept namespace prefix. In the underlying RDF dataset, the complete hierarchical

URIs remain unchanged. Readers interested in the full structure can access the data through the open data repository provided at the end of the paper.

Together, these components constitute a layered, interoperable model that supports interpretive analysis, cross-version alignment, and multilingual access in digital literary studies.

4.1 Poem Class as a Semantic Anchor

In this ontology, each poem is modeled as an instance of the hp:Poem class, which extends the bf:Work class defined in BIBFRAME. This design reflects the dual status of poems as both bibliographic works and semantic anchors within the model. Rather than embedding poems as subordinate items within static collections, the ontology promotes them to first-class semantic entities capable of aggregating interpretive knowledge, bibliographic metadata, and multimodal representations.

This modeling approach is particularly necessary in the East Asian literary tradition, where Han poems have historically been compiled into numerous anthologies across dynasties, regions, and editorial intentions. A single poem may appear in Tang Shi San Bai Shou (Three Hundred Tang Poems), Quan Tang Shi (Complete Tang Poems), and Tang Shi Xuan (Selected Tang Poems), each differing in structure, annotation, and readership context. Modeling the poem as a standalone entity—rather than as a fixed component of a particular compilation—ensures bibliographic clarity and enables cross-collection comparison and alignment.

As a semantic anchor, the hp:Poem class connects core metadata—such as multilingual titles, authorship, and genre—with interpretive knowledge entities through the property hp:hasKnowledge. These knowledge entities represent interpretive elements commonly used in literary analysis, including imagery, emotional expression, rhetorical features, and thematic orientations.

It also supports links to multiple versions or manifestations—such as translations, annotated editions, or digital renderings—via the hp:isRepresentedIn property, which connects the
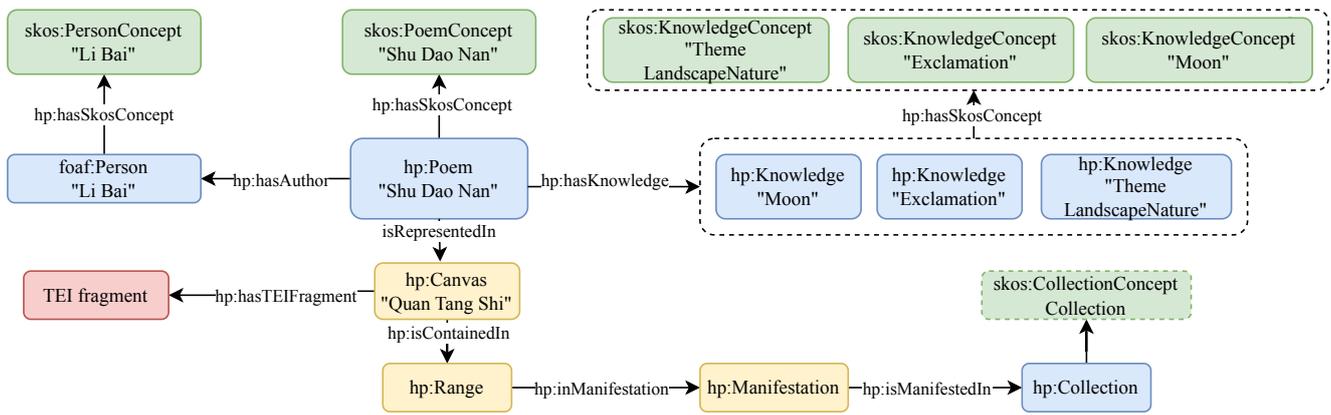
Figure 6. Multi-layer semantic model of Classical Chinese poetry centered on hp:Poem. The diagram shows how OWL entities (poem, person, knowledge, collection) are linked with SKOS concepts. The model also demonstrates multimodal integration of TEI fragments through hp:Canvas.

poem entity to one or more hp:Canvas instances representing specific textual or multimodal realizations.

The following RDF example illustrates this modeling structure using the poem Shu Dao Nan (The Road to Shu is Hard) by Li Bai. The poem is instantiated as an hp:Poem entity with multilingual titles, an author reference (person:LiBai), a symbolic knowledge link to knowledge:ImageryMoon, and associations with two hp:Canvas instances: one representing the original Chinese version in Quan Tang Shi, and the other representing an English translation in Selected Chinese Poems Translated into English Verse.

```
   Poem Entity with Titles, Author, Knowledge, and
   Representations

1  poem:shudaonan a hp:Poem;
2      bf:title" 蜀道难 "@zh ,
3          " The Road to Shu is Hard "@en;
4      hp:hasAuthor person:LiBai;
5      hp:hasKnowledge knowledge:ImageryMoon ,
6          knowledge:ThemeLandscapeNature;
7      hp:isRepresentedIn
↪          canvas:quantangshi-shudaonan ,
8          canvas:yingyitangshi-shudaonan.
```

4.2 Semantic Linking between Knowledge Entities and Poems

Building upon the lexical alignment mechanism established in the SKOS layer (Section 3.2), the ontology introduces the class hp:Knowledge to model interpretive variation at the poem level rather than at

the concept level. While SKOS ensures multilingual referential consistency, semantic interpretation is operationalized through relational configurations among knowledge entities within individual poems.

In the proposed model, symbolic, emotional, and thematic features of Classical Chinese poetry are represented as independently modeled knowledge entities using the hp:Knowledge class. Each hp:Knowledge instance corresponds to a single interpretive abstraction (e.g., ImageryMoon, Exclamation, LandscapeNature, Loneliness) and is linked to its lexical counterpart in SKOS via the property hp:hasSKOSConcept. This separation ensures that lexical alignment and interpretive modeling remain conceptually distinct while interoperable.

Although SKOS provides multilingual labeling and hierarchical organization, its lightweight structure does not aim to encode the interpretive ambiguity or cultural multiplicity inherent in literary texts. Identical lexical concepts—such as "moon" or "wind"—may participate in culturally distinct symbolic or thematic configurations across different linguistic traditions. Rather than embedding such meanings directly within the concept layer (e.g., via skos:related or skos:definition), the ontology models them relationally at the poem level.

Each hp:Knowledge entity is connected to poems through the property hp:hasKnowledge. In this design, interpretive categories such as imagery, emotion, or thematic orientation are represented as different knowledge entities rather than as distinct

ontology properties. The poem thus functions as the semantic anchor: it provides the contextual environment in which multiple knowledge entities co-occur. Interpretive synthesis emerges from these co-occurrence patterns rather than from predefined ontological hierarchies.

```
  Constructing Knowledge Entities and Linking to
  Poems

1 concept:moon a skos:Concept;
2     skos:prefLabel " moon "@en;
3     skos:prefLabel " 月亮 "@zh;
4     skos:prefLabel " つき "@ja.
5
6 knowledge:ImageryMoon a hp:Knowledge;
7     hp:hasSKOSConcept concept:moon.
8
9 poem:shudaonan hp:hasKnowledge
10     knowledge:ImageryMoon;
11     knowledge:ThemeLandscapeNature.
```

For example, Shu Dao Nan by Li Bai is not modeled as encoding a fixed semantic assertion such as "moon equals loneliness" or "nature equals hardship." Instead, it is semantically linked to independently modeled knowledge entities such as Exclamation (representing rhetorical affect) and ThemeLandscapeNature (indicating its dominant thematic domain). These elements are not ontologically fused but simply co-occur within the same poem instance.

In this way, the same SKOS concept (e.g., hpc:Moon) may participate in different knowledge configurations across poems from distinct cultural and linguistic traditions. Semantic divergence is therefore preserved through instance-level relational structures without requiring predefined equivalence or culturally prescriptive definitions at the concept layer.

4.3 Collection-Level Modeling: Compilation Context and Resource Organization

In this ontology, poetry collections are modeled as hp:Collection entities, representing curated groupings of poems that reflect historical, editorial, or publication contexts. Such collections—like Quan Tang Shi, Tang Shi San Bai Shou, and Tang Shi Xuan —are critical to the transmission and reception of Classical Chinese poetry. A single poem may appear in multiple compilations, each with its own structure, annotations, and interpretive framing. Modeling collections as independent entities allows poems to maintain semantic integrity while participating in diverse institutional or cultural contexts.

Each hp:Collection contains one or more hp:Manifestation entities, representing concrete textual realizations such as printed editions or translated versions. These are further structured through hp:Range entities to capture internal divisions—such as volumes, chapters, or scrolls. At the lowest structural level, hp:Canvas entities represent the point of linkage between physical manifestations and the abstract semantic anchor: the hp:Poem instance.

This layered architecture—Collection → Manifestation → Range → Canvas—follows the logic of the IIIF Presentation API while adapting it to the structural conventions of classical East Asian poetry. The inclusion of intermediate layers supports fine-grained navigation, fragment-based referencing, and version-aware alignment across multiple manifestations of a collection.

While these structural layers organize how a poem is manifested and presented, they do not carry semantic interpretations themselves. Instead, they provide a contextual resource framework within which hp:Poem entities—defined as semantic anchors —can be situated and retrieved. This separation of structure and semantics allows for coherent integration of interpretive data (as discussed in 4.2) with publication metadata, and supports multi-version analysis across physical or digital repositories.

For example, Quan Tang Shi is modeled as follows:

```
Quan Tang Shi Collection to Canvas Structure

1  collection:quantangshi a hp:Collection;
2      rdfs:label " 全唐诗 "@zh;
3      rdfs:label " Complete Tang Poems "@en;
4      hp:hasAuthor person:Pengdingqiu;
5      hp:hasCreationTime
       ↪   "1705-01-01"xsd:date;
6      hp:hasCollectionConcept
       ↪   concept:quantangshi;
7      hp:hasManifestation
       ↪   manifestation:quantangshi.
8
9  manifestation:quantangshi a
   ↪   hp:Manifestation;
10     hp:isManifestedIn
       ↪   collection:quantangshi;
11     hp:hasRange range:quantangshi-vol20.
12
13 range:quantangshi-vol20 a hp:Range;
14     hp:rangePosition 20;
15     hp:inManifestation
       ↪   manifestation:quantangshi;
16     hp:hasCanvas
       ↪   canvas:quantangshi-shudaonan.
17
18 canvas:quantangshi-shudaonan a hp:Canvas;
19     hp:representsPoem poem:shudaonan;
20     hp:isContainedIn
       ↪   range:quantangshi-vol20.
```

The following example illustrates how Shu Dao Nan is linked to two different TEI resources—one in Chinese and one in English—via two separate hp:Canvas instances:

```
Canvas Entities with Language and TEI Fragments

1  canvas:quantangshi-shudaonan a hp:Canvas;
2      hp:representsPoem poem:shudaonan;
3      hp:hasLanguage "zh";
4      hp:hasTEIFragment
       ↪   <teires:quantangshi.xml#vol20-shudaonan>.
5
6  canvas:yingyitangshi-shudaonan a hp:Canvas;
7      hp:representsPoem poem:shudaonan;
8      hp:hasLanguage "en";
9      hp:hasTEIFragment
10     <teires:yingyitangshi.xml#shudaonan>.
```

Although the current model links only TEI fragments, the Canvas layer is designed to accommodate a wide range of multimodal resources—including images, audio recordings, and annotations—making it extensible for future integration of IIIF images and other media types. As such, Canvas acts as the bridge between abstract semantic representation and concrete resource expression, ensuring that interpretive analysis remains anchored in specific, retrievable forms.

4.4 Integrating TEI and Semantic Layers Through Canvas

The hp:Canvas class serves as the lowest-level anchoring unit in the resource layer, linking abstract hp:Poem entities to their concrete textual or multimodal manifestations. This structure is inspired by the IIIF Presentation API, where Canvas represents a discrete, displayable segment of a digital resource.

In the current implementation, each hp:Canvas instance connects a poem to a corresponding fragment in a TEI file. These fragments represent specific versions or translations of the poem and are distinguished by language and editorial context. The use of Canvas enables consistent referencing across different manifestations while maintaining a stable link to the central poem identity.

4.5 SKOS-Based Conceptual Mapping for Semantic Navigation

As introduced in Section 3.2, the SKOS concept layer provides a lightweight multilingual vocabulary system that supports label alignment, hierarchical categorization, and navigable concept schemes. While not used for interpretive modeling, it plays a key role in enabling semantic navigation and cross-resource aggregation.

To integrate this layer with the ontology's semantic core, the system adopts a dual modeling approach: instances belonging to the classes hp:Poem, hp:Person, and hp:Collection are represented both as OWL individuals and as SKOS concepts. These are linked via properties such as hp:hasTitleConcept, hp:hasPersonConcept, and hp:hasCollectionConcept. For instance:

```
   Mapping Poem and Person Concepts

1  poem:shudaonan hp:hasTitleConcept
   ↪    concept:shudaonan.
2  concept:shudaonan a skos:Concept;
3      skos:prefLabel " 蜀道难 "@zh;
4      skos:altLabel " The Road to Shu "@en;
5      skos:inScheme <.../PoemTitleConcept>.
6
7  person:LiBai hp:hasPersonConcept
   ↪    concept:LiBai.
8  concept:LiBai a skos:Concept;
9      skos:prefLabel " 李白 "@zh;
10     skos:prefLabel " Li Bai "@en;
11     skos:broader concept:TangPoet.
```

This separation of semantic representation (OWL) and conceptual labeling (SKOS) enables multilingual retrieval, faceted browsing, and cross-collection clustering—without interfering with the poem' s interpretive modeling. It bridges deep ontological structure with practical user-facing access, making the system both semantically expressive and navigable.

5 Use Case: Modeling and Federating Classical Chinese poetry

To demonstrate the practical applicability of the proposed semantic framework, this chapter presents a prototype application centered on Classical Chinese poetry. Rather than attempting to cover the entire corpus, the prototype strategically selects a small yet diverse set of samples. This selection is intended to showcase the framework' s core capabilities in structured semantic modeling, multilingual concept alignment, and distributed data federation.

The focus of this chapter is twofold: firstly, on the construction and structuring of a small cross-linguistic, cross-cultural corpus (Section 5.1); and secondly, on validating semantic integration and querying capabilities through SPARQL querying scenarios (Section 5.2). This case study aims to illustrate that the proposed semantic framework is not only technically viable but also capable of fostering cross-linguistic, cross-cultural, and multimodal literary knowledge networks.

As shown in Figure 7, the data processing workflow begins with TEI-encoded Classical Chinese poetry texts, which are subsequently transformed into RDF-based semantic representations. Through this process, textual data, entity information, and extracted knowledge elements are organized into structured datasets that support semantic integration and federated querying across distributed resources.
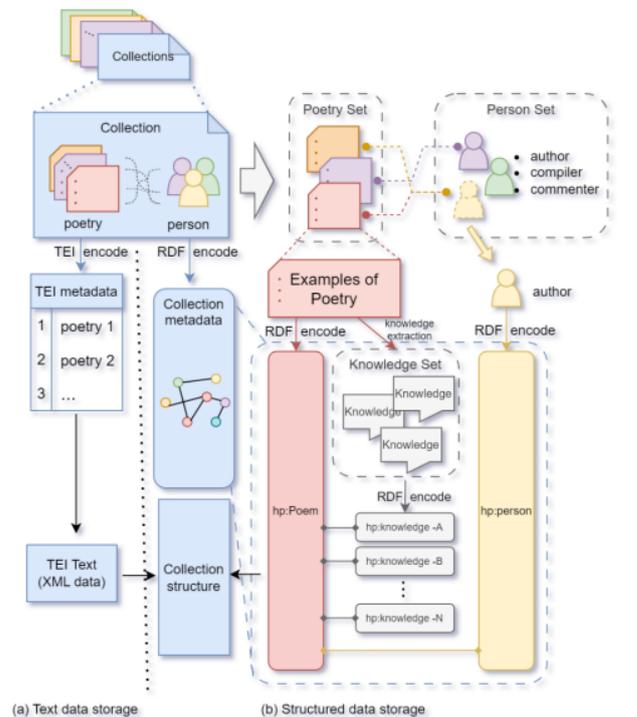
Figure 7. Data Processing Workflow for Classical Chinese poetry Semantic Infrastructure

5.1 Data Sources and Corpus Overview

To build a prototype corpus that demonstrates structured modeling, multilingual alignment, and distributed interoperability, we employed a minimalist sampling strategy. The corpus, comprising ten poems from both Chinese and Japanese Classical Chinese poetry, was carefully selected to represent linguistic diversity, cultural origin, shared imagery, and structural variety. This cross-linguistic and cross-cultural selection enables a focused evaluation of the system' s modeling and retrieval capabilities while ensuring coverage of essential semantic variations.

The corpus draws from three primary sources: Quan Tang Shi (Complete Tang Poems), Selected Chinese Poems Translated into English Verse [49], and Nihon Shiki (Japanese Records of Poetry) [50]. Quan Tang Shi provides original Chinese texts by poets such as Li Bai, Du Fu, and Bai Juyi, sourced from the Chinese Text Project. Selected Chinese Poems Translated into English Verse offers English translations of corresponding poems, validating cross-linguistic concept mapping. Nihon Shiki includes Japanese Classical Chinese poetry, showcasing the tradition's continuation and adaptation in a different cultural context.

The corpus structure follows the Collection–Manifestation–Range–Canvas model. Each poetry collection is represented as an hp:Collection entity, with different versions (e.g., original text, translations, annotated versions) modeled as hp:Manifestation instances. These are subdivided into Range and Canvas units for fine-grained access. Individual poems are modeled as hp:Poem entities, linking to specific Canvas units via the hp:isRepresentedIn property, preserving the poems' abstract semantic integrity while associating them with textual resources, translations, or multimodal representations.

All texts are encoded in TEI-XML at the stanza and line levels, allowing precise linking of textual components. In the knowledge layer, concepts such as "moon" imagery were modeled as parallel skos:Concepts in Chinese, English, and Japanese, linked to poem entities via properties hp:hasKnowledge. The Chinese and Japanese corpora were deployed in separate RDF triple stores and federated using SPARQL 1.1. Despite the limited sample size, the structural diversity and semantic depth of the corpus effectively validated key system capabilities, including cross-linguistic retrieval, multi-version aggregation, and concept alignment, laying a foundation for future expansion and multicultural integration.

5.2 SPARQL Querying Scenarios and Integration Validation

To systematically validate the decentralized, multilingual, and concept-driven semantic infrastructure proposed in this study, a set of targeted SPARQL querying scenarios was designed and executed. All RDF datasets were deployed and managed using the Virtuoso triple store, where different poetry corpora were stored in independent named graphs and interconnected through SPARQL 1.1 federated querying mechanisms.

The query examples presented in this section illustrate the logical structure of the retrieval tasks. For readability, the query logic is presented in a simplified pseudocode form using the namespace prefixes defined in Section 3.1. The complete executable SPARQL queries used in the experiments are provided in the open data repository associated with this study.

It should also be noted that TEI resources were not directly stored within the Virtuoso triple store. Instead, TEI documents were maintained in an external HTTP-accessible repository and linked to the RDF data through persistent URIs. This design preserves the integrity of the TEI encoding while allowing the semantic infrastructure to reference and retrieve TEI fragments through linked data mechanisms.

Each querying scenario focuses on validating a specific functional capability of the system, including cross-repository concept retrieval and multimodal resource integration. Together, these scenarios demonstrate the system's semantic interoperability across multilingual datasets, distributed repositories, and heterogeneous resource formats.

5.2.1 Concept-Based Cross-Repository Semantic Retrieval

To verify the system's capability for semantic retrieval across distributed repositories and linguistic environments, a concept-centered query was designed around the imagery concept "moon". The query retrieves poems associated with the knowledge entity representing moon imagery from both the Tang poetry database (Tang_db) and the

Heian poetry database (Heian_db). It further traces the thematic knowledge entities linked to each poem.

For readability, the query logic is illustrated below in simplified pseudocode form using the namespace prefixes defined in Section 3.1. The complete executable SPARQL queries used in the experiment are provided in the open data repository accompanying this study.

SPARQL Query Logic: Cross-Repository Retrieval of Moon Imagery (Simplified)

```
1 SELECT dataset, poem, title, themeLabel
2 FROM Tang_db, Heian_db
3
4 WHERE
5     poem a hp:Poem
6     poem bf:title title
7     poem hp:hasKnowledge
       ↪   knowledge:ImageryMoon
8
9 OPTIONAL
10    poem hp:hasKnowledge theme
11    theme rdfs:label themeLabel
12
13 ORDER BY dataset, title
```

To illustrate the results of this concept-centered retrieval, Table 1 presents a summary of poems related to the moon imagery retrieved from both databases. The table lists the source dataset, the URI of each poem, the English title, and the thematic label associated with each work. This structured presentation highlights the system's ability to integrate data from distinct linguistic and cultural repositories, enabling cross-repository semantic retrieval.

The results presented in Table 1 demonstrate two key capabilities of the proposed framework. First, by using shared conceptual anchors within the multilingual SKOS vocabulary, the system enables federated querying across independently managed corpora with different linguistic backgrounds. Second, through the alignment between SKOS concepts, OWL knowledge entities, and poem instances, the system can retrieve not only poems associated with a given imagery concept but also their related thematic knowledge entities. This capability illustrates the

effectiveness of integrating conceptual vocabularies with ontology-based knowledge modeling for semantic literary analysis.

By facilitating concept-centered semantic retrieval, the framework provides a technical foundation for cross-cultural poetic comparison, thematic imagery analysis, and distributed literary knowledge research.

5.2.2 Linking TEI Resources via Canvas Entities To further validate the system's multimodal integration capabilities, this section presents a retrieval scenario focusing on the TEI-encoded text fragment of Li Bai's poem "Shu Dao Nan" (The Road to Shu is Hard) and its associated metadata. This example demonstrates how the Canvas entity functions as an intermediary layer connecting semantic poem entities with concrete documentary resources.

The retrieval task aims to obtain the poem title, the associated collection, the Canvas identifier representing the textual fragment, and the corresponding TEI resource. Through this process, the system establishes a complete semantic linkage from poem entities to specific textual manifestations encoded in TEI.

For clarity, the logical structure of the retrieval task is illustrated below using simplified pseudocode. The complete executable SPARQL query is provided in the open data repository accompanying this study.

Table 1. Concept-Based Cross-Repository Semantic Retrieval Results for Moon Imagery.

Dataset	Poem (URI)	Title (en)	Theme Label (en)
Heian_db	poem:yueyejianmeihua	Seeing Plum Blossoms on a Moonlit Night	Sentiment / Lyric
Heian_db	poem:qiuye	Autumn Night	Sentiment / Lyric
Tang_db	poem:guanshanyue	THE MOON OVER THE PASS	Border Campaign
Tang_db	poem:yonghuaiguji3	AN ANCIENT STORY	Nostalgia / Antiquity
Tang_db	poem:yebolvwang	Night Mooring and Traveler's View	Homesickness
Tang_db	poem:yuexiaduzhuo	WE THREE	Sentiment / Lyric
Tang_db	poem:qiuxing7	ODES TO AUTUMN No. 7	Nostalgia / Antiquity
Tang_db	poem:shudaonan	The Road to Shu is Hard	Landscape / Nature
Tang_db	poem:songchuyizhiwuchang	TO WUCHANG	Farewell
Tang_db	poem:xianyong	Idle Chant	Sentiment / Lyric

SPARQL Query Logic: Retrieving TEI Resources via Canvas (Simplified)

```
1 SELECT titleEn, collectionEn, canvas,
  ↪  teiFragment
2 FROM Tang_db
3
4 WHERE
5     poem = poem:shudaonan
6     poem bf:title titleEn
7
8     canvas hp:representsPoem poem
9
10 OPTIONAL
11     canvas hp:hasTEIFragment teiFragment
12
13 OPTIONAL
14     canvas hp:isContainedIn range
15     range hp:inManifestation manifestation
16     manifestation hp:isManifestedIn
      ↪  collection
17     collection rdfs:label collectionEn
18
19 ORDER BY collectionEn
```

The retrieval results for TEI resources linked via Canvas entities are presented in Table 2. The table lists the poem title, the corresponding poetry collection, the Canvas URI representing the document fragment, and the TEI URI that points to the encoded textual content.

As shown in Table 2, the poem entity hp:Poem (e.g., "Shu Dao Nan") is linked to one or more hp:Canvas instances through the property hp:representsPoem. Each Canvas entity corresponds to a specific textual manifestation of the poem within a particular documentary context. Through the property hp:hasTEIFragment, the Canvas entity further links to TEI fragments that encode the actual textual content.

This layered structure enables the system to trace semantic entities to concrete textual representations while preserving the distinction between abstract works and their documentary manifestations. Moreover, the results demonstrate that a single poem may be associated with multiple collections and textual realizations, illustrating the system's capability to integrate distributed documentary resources within a unified semantic framework.

By establishing these semantic connections, the system supports fine-grained literary analysis and cross-resource navigation, allowing researchers to move seamlessly from conceptual knowledge structures to specific textual evidence. The same mechanism can also be extended to integrate additional media resources, such as images or audio recordings, thereby expanding the multimodal scope of the poetry knowledge infrastructure.

6 Conclusion

This study proposes and validates a decentralized, multilingual, concept-driven semantic infrastructure for organizing and federating Classical Chinese poetry resources within East Asian cultural contexts. Through the construction of multilingual SKOS vocabularies, the reuse of OWL ontologies based on BIBFRAME and FOAF, and

Table 2. Retrieval Results for TEI Resources Linked via Canvas Entities.

Title (en)	Collection (en)	Canvas URI	TEI URI
The Road to Shu is Hard	Complete Tang Poems	canvas:quantangshishudaonan	teires:quantangshi.xml#vol20shudaonan
The Road to Shu is Hard	Selected Chinese Poems Translated into English Verse	canvas:yingyitangshishudaonan	teires:yingyitangshi.xml#shudaonan

the implementation of an RDF resource layer supporting SPARQL federated queries, the system enables semantic interoperability across heterogeneous poetry datasets from different institutions and linguistic backgrounds.

The prototype use case demonstrated the framework' s capabilities in enabling cross-cultural retrieval based on shared imagery (e.g., the "moon" motif), semantic aggregation queries driven by thematic concepts, and precise linking of TEI-encoded textual resources to semantic entities via the Canvas mechanism. The results show that the architecture supports multilingual alignment, multimodal resource integration, and decentralized knowledge federation, providing strong feasibility and flexibility.

This research addresses the technical challenges of fragmented, inconsistent, and non-interoperable Classical Chinese poetry resources, while also emphasizing the philosophical and sociocultural complexities of literary heritage modeling. In particular, the study explores the dynamic balance between local interpretative diversity and global semantic integration in contexts where textual forms are unified, but deep semantic structures vary.

Despite its contributions, several limitations remain. The experimental corpus is relatively small, and many complex poetic phenomena, such as metaphor networks and intertextual structures, have not been fully modeled. Future work will focus on expanding the corpus, integrating natural language processing and large language models for annotation automation, deepening ontology modeling, and strengthening multimodal integration. Additionally, connecting with global cultural heritage platforms like Wikidata and adhering to open data principles will enhance global semantic interoperability.

While this study centers on Classical Chinese poetry, the semantic framework has the potential to extend to other poetic traditions. The design maintains decoupling between conceptual and resource layers, making it adaptable to various linguistic and cultural contexts, such as Japanese waka, English sonnets, and Latin elegies. This flexibility establishes a strong foundation for the global interconnection and collaboration of literary cultural heritage.

In summary, this study offers a technical prototype and a sustainable pathway for exploring cross-linguistic, cross-institutional, and cross-modal semantic integration of texts, knowledge, and cultural heritage within the field of Digital Humanities.

Acknowledgements

Data Availability

All datasets, ontology models, TEI-encoded texts, and SPARQL query examples developed in this study have been openly released on GitHub, accompanied by documentation detailing the procedures for reproducing the experimental results. These resources are made available to support subsequent research and future extensions.

References

[1] Leonardo Candela, Donatella Castelli, Nicola Ferro, Yannis Ioannidis, Georgia Koutrika, Carlo Meghini, Pasquale Pagano, Seamus Ross, Dagobert Soergel, Maristella Agosti, Milena Dobreva, Vassiliki Katifori,

and Heiko Schuldt. *The DELOS Digital Library Reference Model: Foundations for Digital Libraries.* DELOS Network of Excellence on Digital Libraries, Pisa, Italy, 2007.

[2] International Image Interoperability Framework Consortium. IIIF Presentation API 3.0 Specification. https://iiif.io/api/presentation/3.0/. Retrieved April 27, 2025.

[3] Harvard University and Academia Sinica and Peking University. China Biographical Database (CBDB). https://projects.iq.harvard.edu/cbdb. Retrieved April 27, 2025.

[4] National Institute of Japanese Literature. Japanese biographical database (jbdb). https://jbdb.jp/. Retrieved March 5, 2026.

[5] W3C. RDF 1.1 Concepts and Abstract Syntax. https://www.w3.org/TR/rdf11-concepts/, 2014.

[6] W3C OWL Working Group. OWL 2 Web Ontology Language Document Overview (Second Edition). https://www.w3.org/TR/owl2-overview/, 2012.

[7] A. Miles and S. Bechhofer. SKOS Simple Knowledge Organization System Reference. https://www.w3.org/TR/skos-reference/, 2009.

[8] S. Harris, A. Seaborne, and E. Prud' hommeaux. SPARQL 1.1 Query Language. https://www.w3.org/TR/sparql11-query/, 2013.

[9] M. L. Zeng. Knowledge organization systems (KOS). *Knowledge Organization*, 35(2–3):160–182, 2008.

[10] Getty Research Institute. Art & Architecture Thesaurus (AAT). https://www.getty.edu/research/tools/vocabularies/aat/, 2021. Retrieved April 7, 2025.

[11] Wikimedia Foundation. Wikidata: A Free Collaborative Knowledge Base. https://www.wikidata.org/. Retrieved April 7, 2025.

[12] A. Jordanous, K. F. Lawrence, M. Hedges, and C. Tupman. Exploring manuscripts: Sharing ancient wisdoms across the semantic web. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, pages 1–12, 2012.

[13] P. Proutskova, D. Wolff, G. Fazekas, K. Frieler, F. Höger, O. Velichkina, and S. Dixon. The Jazz Ontology: A semantic model and large-scale RDF repositories for jazz. *Journal of Web Semantics*, 74:100735, 2022.

[14] M. Doerr, C. Bekiari, P. LeBoeuf, and Bibliothèque nationale de France. FRBRoo, a conceptual model for performing arts. In *2008 Annual Conference of CIDOC*, pages 15–18, Athens, September 2008.

[15] Martin Doerr. The CIDOC Conceptual Reference Model (CRM): A Standard for Semantic Interoperability in Cultural Heritage. Technical report, International Council of Museums (ICOM) CIDOC CRM Special Interest Group, 2009. http://www.cidoc-crm.org/.

[16] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.

[17] B. Haslhofer and A. Isaac. data.europeana.eu: The europeana linked open data pilot. In *Proc. of the Int. Conf. on Dublin Core and Metadata Applications.* Dublin Core Metadata Initiative, 2011.

[18] L. Isaksen, R. Simon, E. T. Barker, and P. de Soto Cañamares. Pelagios and the emerging graph of ancient world data. In *Proc. of the 2014 ACM Conf. on Web Science*, pages 197–201, June 2014.

[19] E. Hyvönen. "Sampo" model and semantic portals for digital humanities on the semantic web. In *Digital Humanities in the Nordic Countries*, pages 373–378. CEUR, October 2020.

[20] R. Duan. The semantic networking of the thesaurus based on the simple knowledge organization system: A case study of the chinese archival thesaurus. *Journal of Library Science in China*, 37(03):54–65, 2011.

[21] X. Wang, X. Hou, H. Cheng, and S. Xia. Construction of the Dunhuang mural thesaurus and linked data publishing. *Journal of Library Science in China*, 46(04):69–84, 2020.

[22] S.-J. Chen and H.-H. Chen. Lexical-semantic mapping between chinese and english controlled vocabularies in the domain of chinese art. *Journal of Library and Information Science Research*, 13(2):161–208, 2015.

[23] C. Xia and L. Zhang. The application of linked data in Shanghai Library' s service of genealogy digital humanities. *Library Journal*, 35(10):26–34, 2016.

[24] X. Wang, L. Xu, and G. Li. Semantic description framework research on Dunhuang fresco digital images. *Journal of Library Science in China*, 40(01):50–59, 2014.

[25] T. Chen and J. Zhang. Framework design and exploration of the ontology service center. *Library*

Journal, 41(03):118–125, 2022.

[26] Lev Snydman, Robert Sanderson, and Tom Cramer. The international image interoperability framework (iiif): A community technology approach for web-based images. *D-Lib Magazine*, 21(7/8), 2015.

[27] National Institute of Japanese Literature. Union Catalogue Database of Japanese Texts. `https://kokusho.nijl.ac.jp/`. Retrieved April 7, 2025.

[28] Center for Open Data in the Humanities (CODH). IIIF Curation Viewer. `https://codh.rois.ac.jp/software/iiif-curation-viewer/`. Retrieved April 7, 2025.

[29] TEI Consortium. TEI: Guidelines for Electronic Text Encoding and Interchange (P5 Version 4.9.0, revision f73186978). `https://tei-c.org/release/doc/tei-p5-doc/en/html/`. Retrieved April 7, 2025.

[30] National Institute of Japanese Literature. Digital Tale of Genji. `https://genji.dl.itc.u-tokyo.ac.jp/`. Retrieved April 7, 2025.

[31] National Museum of Japanese History. Engi Shiki Database. `https://khirin-t.rekihaku.ac.jp/engishiki`. Retrieved April 7, 2025.

[32] International Institute for Digital Humanities. TEI Viewer Platform. `https://tei.dhii.jp/teiviewer4eaj`. Retrieved April 7, 2025.

[33] Chinese Text Project. Chinese Text Project. `https://ctext.org/ens`. Retrieved April 7, 2025.

[34] Y.-F. Chen. 寒泉古典文獻全文檢索資料庫 [Hanquan Classical Texts Full-text Search Database]. `https://skqs.lib.ntnu.edu.tw/dragon/`. Retrieved April 7, 2025.

[35] chinese poetry. Chinese poetry [Computer software]. `https://github.com/chinese-poetry/chinese-poetry`. Retrieved April 7, 2025.

[36] Y. Liu, B. Wu, and T. Bai. Construction and analysis of classical Chinese poetry knowledge graph. *Computer Research and Development*, 57(06):1252–1268, 2020.

[37] L. Zhou, L. Hong, and Z. Gao. Construction of knowledge graph of Chinese Tang poetry and design of intelligent knowledge services. *Library and Information Service*, 63(02):24–33, 2019.

[38] Y. Cong and M. Takaku. The study on structured data for Chinese poetry: Focus on Tang poems. *Journal of Japan Society of Information and Knowledge*, 32(1):15–38, 2022. In Japanese.

[39] Library of Congress. BIBFRAME 2.0: Model and design. `https://www.loc.gov/bibframe/docs/`, 2019. Retrieved April 27, 2025.

[40] Y. Li and S. Chen. Chinese metaphor recognition using a multi-stage prompting large language model. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)*, pages 1456–1468. Association for Computational Linguistics, 2024.

[41] T. Chen, X. Yang, Y. Xia, and R. Su. Research and design of the interconnecting framework of ancient book knowledge base. *Journal of Academic Libraries*, 41(04):58–64, 2023.

[42] T. Berners-Lee. Linked data. `https://www.w3.org/DesignIssues/LinkedData.html`, 2006. Retrieved April 27, 2025.

[43] D. Vila-Suero, A. Gómez-Pérez, E. Montiel-Ponsoda, J. Gracia, and G. Aguado-de Cea. Publishing linked data on the web: The multilingual dimension. In *Towards the Multilingual Semantic Web: Principles, Methods and Applications*, pages 101–117. Springer, 2014.

[44] Eun-Ju Kim and Seul Lee. 韩国月亮文化初探 [a preliminary study on korean moon culture]. *The Journal of Chinese Cultural Studies*, (13):259–273, 2008. In Chinese.

[45] Man-Li Hsu. A study about the poetry of the moon : concerning with folklore and belief. *Geibun-Kenkyu: Journal of Arts and Letters*, 65:288–309, 1994. In Japanese.

[46] Hiroshi Sasagawa. 平安文学と月 [the moon in heian literature]. 大阪大谷大学 *STEAM Lab* 紀要, 1:15–17, 2021. In Japanese.

[47] Information and documentation —romanization of chinese, 2015.

[48] OpenLink Software. Virtuoso Universal Server. `https://virtuoso.openlinksw.com/`. Retrieved April 7, 2025.

[49] W. J. B. Fletcher. *Selected Chinese Poems Translated into English Verse*. China Pictorial Publishing House, 2019.

[50] K. Takashima, editor. 日本詩紀*: 本文と总索引 [Nihon shiki: Text and general index]*. 汲古書院, 2003.