

Construction of Thangka Cultural Knowledge Graph and Cross-regional Inheritance Application

Hao Zhang, Ramin Hajianfard*

Faculty of Creative Industries, City University Malaysia, Kuala Lumpur, Federal Territory of Kuala Lumpur, 46100, Malaysia

* Corresponding author: Ramin Hajianfard

Abstract: Thangka - sometimes transliterated thang-ka or tangka - is a centuries-old Tibetan Buddhist painting tradition on cotton or silk, encoding layered iconographic meanings accessible only through initiated ritual and artistic training. UNESCO inscribed Thangka painting on its Representative List of the Intangible Cultural Heritage of Humanity in 2009. That inscription, while conferring international recognition, also exposes a vulnerability: trained masters are declining in number, and younger audiences increasingly encounter these paintings through digital interfaces rather than in shrine settings that give them their full cultural significance. How to systematically represent, preserve, and transmit Thangka across regions has become pressing. Over the past decade, digital preservation of intangible cultural heritage (ICH) has drawn sustained attention. High-resolution imaging, 3D scanning, and online museum archives have produced vast digital corpora. Yet most efforts treat digitized Thangka as static image resources, disconnected from the rich semantic networks - iconographic attributes, artisan lineages, historical trajectories, ritual functions - that make Thangka culturally intelligible. Without structured semantic representation, digital collections remain fragmented and resist being queried, interconnected, or reused for education and inheritance work. Knowledge Graph (KG) technology, grounded in semantic web and linked data principles, offers a way forward. By representing cultural heritage entities and their relationships as nodes and edges in a graph structure, KGs enable semantic reasoning, cross-dataset linking, and intelligent question-answering. This paper presents the construction of a domain-specific knowledge graph for Thangka cultural heritage and evaluates its application to cross-regional inheritance. The KG integrates multi-source heterogeneous data - museum collections, textual sources, artisan interviews, and online archives - and formalizes the domain ontology in OWL, aligned with CIDOC CRM. A prototype platform, the Thangka Cultural Knowledge Graph Platform (TCKGP), demonstrates the KG's utility across three scenarios: museum digital curation, online artisan training, and transnational cultural education. Evaluation results show that the KG approach substantially outperforms conventional relational database approaches in semantic retrieval tasks and effectively supports cross-regional inheritance practices. This study contributes to digital heritage scholarship by providing a reusable ontology and a practical methodology (the TCKG Model) for applying knowledge graphs to ICH, and to Tibetan Buddhist art studies by supplying the first formally specified, machine-readable ontology for Thangka iconography and lineage.

Keywords: Thangka, knowledge graph, cultural heritage, cross-regional inheritance, semantic web, digital humanities, ontology construction.

1. Introduction

Thangka (also spelled thang-ka or tangka) refers to Tibetan Buddhist scroll paintings rendered on cotton or silk, depicting deities, mandalas, and narrative scenes from sacred texts. The practice dates back more than a millennium. These paintings function simultaneously as devotional objects and pedagogical instruments. Their iconographic meaning is encoded in layers; access to those layers requires specialized ritual and artistic training [12]. That training pipeline is narrowing.

UNESCO inscribed Thangka on its Representative List of the Intangible Cultural Heritage of Humanity in 2009. [19] That recognition matters, but it also creates exposure. The number of trained masters keeps declining. Younger audiences now encounter Thangka through digital interfaces rather than in shrine settings where the paintings' full cultural significance becomes legible. How to systematically represent, preserve, and transmit Thangka across regions is no longer an abstract question.

Digital preservation of ICH has, over the past decade, attracted substantial scholarly and institutional attention. High-resolution imaging, 3D scanning, and online museum archives have produced vast digital corpora. Yet these efforts, for all their value, treat digitized Thangka largely as static

image resources. The semantic networks that make Thangka culturally intelligible - iconographic attributes, artisan lineages, historical trajectories, ritual functions - remain disconnected from the digital record [23] [16]. Without structured semantic representation, digital collections stay fragmented and resist being queried, interconnected, or reused.

A different approach is needed. Knowledge Graph (KG) technology, built on semantic web and linked data principles, makes it possible to represent cultural heritage entities and their relationships as nodes and edges in a graph structure. That representation enables semantic reasoning, cross-dataset linking, and intelligent question-answering. Europeana, the Getty Vocabulary Program, and CIDOC CRM have already shown what semantic structuring can accomplish in cultural heritage contexts. Applying KG technology to Tibetan Buddhist art, however, remains in its early stages. Existing systems cover only fragments of Thangka iconography and do not integrate with cross-regional inheritance workflows.

This paper addresses three questions. First: how can a domain-specific knowledge graph for Thangka cultural heritage be systematically constructed to cover iconography, lineages, and ritual contexts? Second: how does the KG framework perform on semantic retrieval and reasoning tasks when compared with conventional relational database

approaches? Third: in what ways can a constructed KG support cross-regional inheritance practices - museum curation, artisan training, transnational cultural education? The paper is organized as follows. Section 2 reviews KG construction for cultural heritage and cross-regional ICH inheritance. Section 3 presents the ontology design and KG construction methodology. Section 4 describes the prototype system. Section 5 reports evaluation results across three scenarios. Section 6 discusses implications. Section 7 concludes.

2. Literature Review

2.1. Knowledge Graphs for Cultural Heritage

Knowledge graph technology began gaining real traction in cultural heritage preservation in the early 2010s. Linked Open Data (LOD) initiatives, backed by the W3C and by institutions such as Europeana and the Digital Public Library of America, laid the foundational infrastructure for publishing cultural heritage data as machine-readable graphs [8]. The CIDOC Conceptual Reference Model (CRM), an ISO standard (ISO 21127:2014), provides a formal ontology for integrating cultural heritage information and has seen wide adoption in museum and heritage knowledge graph projects [8] [11].

Recently, researchers have extended KG technology to specific heritage domains. [22] built a knowledge graph for Ming and Qing dynasty architectural heritage and showed that it improved semantic retrieval and cross-collection linking. [15] took a similar approach to traditional Chinese handicrafts, developing an ontology-based KG that integrated heterogeneous data from museum collections and field surveys. In Buddhist art, [10] documented the persistent difficulty of iconographic classification for Tibetan Thangka, noting that existing metadata standards (Dublin Core, VRA Core) fall short in capturing the multi-layered symbolic and ritual dimensions of Thangka. A broader survey by [17] analyzed 798 research articles on knowledge graph-driven digital preservation of ICH published between 2006 and 2024, drawing on both Chinese- and English-language academic discourses. Their analysis shows that KG-based approaches substantially outperform traditional database approaches in supporting semantic search, cross-domain linking, and reasoning over ICH data. They also identify a persistent gap: most existing KGs for ICH focus on single domains and rarely tackle the cross-regional inheritance dimension. This gap is visible in Tibetan Buddhist art, given that cross-regional transmission - across Tibet, Sichuan, Qinghai, and transnational contexts such as Nepal, Bhutan, and Mongolia - is central to how the tradition operates.

2.2. Cross-regional Inheritance of Intangible Cultural Heritage

Cross-regional inheritance, in the ICH context, refers to how intangible cultural practices get transmitted across geographic, linguistic, and cultural boundaries - often with digital technologies and institutional partnerships acting as facilitators. UNESCO's Operational Guidelines for the Implementation of the Convention for the Safeguarding of the Intangible Cultural Heritage [20] explicitly encourages transnational safeguarding cooperation. For Thangka, cross-regional inheritance works along two axes: a domestic one spanning Tibetan-inhabited areas across provincial boundaries within China, and a transnational one reaching

into Nepal, Bhutan, Mongolia, and increasingly into global diasporic communities. Scholars working on cross-regional ICH inheritance have documented both the opportunities and the tensions involved. [18] argue that cross-regional digital projects can democratize access but also risk decontextualizing heritage by detaching it from the source communities that sustain it. [14] puts the concern more sharply: digitization can turn living heritage into what she calls 'digital folklore' - aesthetically packaged but semantically emptied. More recent work by [9] and [5] [6] stresses the importance of 'relational ontologies', which preserve community relationships within digital representations. The present study builds on that relational approach: the knowledge graph encodes not only Thangka objects but also the communities, rituals, and transmission lineages bound up with them.

2.3. Ontology Design for Tibetan Buddhist Art

Ontology design sits at the foundation of any KG construction effort. For Tibetan Buddhist art, existing ontologies tend to fall into two categories: they are either too general (CIDOC CRM) or too narrow (project-specific schemas that cannot be reused). [10] and [13] both provide detailed iconographic classification systems for Tibetan art, covering deity typologies, mudra classifications, color symbolism, and composition types (mandala, lineage tree, narrative scroll). These classification systems, however, exist as print taxonomies and have not been formalized as machine-readable ontologies. The Himalayan Art Resources (HAR) project has made a recent effort to produce a controlled vocabulary for Himalayan art, with more than 2,000 deity entries carrying standardized iconographic attributes. HAR's vocabulary is a valuable reference, but it is distributed as a web directory rather than as a formal ontology expressed in RDF/OWL. The ontology we present in this study is expressed in OWL and aligned with both CIDOC CRM and the HAR vocabulary. This alignment allows interoperability with broader cultural heritage knowledge graphs while supplying the domain-specific depth that Thangka demands.

3. Knowledge Graph Construction Methodology

3.1. Ontology Design and Schema Layer

The ontology adopts a modular design with three sub-ontologies that connect to one another. The Iconography Module defines classes such as Deity, Mudra, Asana, Attribute, ColorSymbolism, and MandalaType, along with the object properties that link them (hasMudra, hasAttribute, symbolizes, and others). The Lineage Module captures the master-apprentice transmission networks, defining classes such as Artisan, Lineage, Workshop, and Region, with properties that include trainedUnder, belongsToLineage, and activeInRegion. The Ritual Context Module encodes the ritual functions of Thangka paintings, including classes such as Ritual, Festival, Temple, and Patron, with properties such as usedInRitual, commissionedBy, and enshrinedIn. The ontology is expressed in OWL 2 DL and aligned with CIDOC CRM v7.1. Alignment is done through subsumption axioms: Deity, for instance, is declared as a subclass of E39_Actor, and ThangkaPainting as a subclass of E22_Human-Made_Object. This alignment ensures that data instantiated in the Thangka KG can be linked to other CRM-compliant knowledge graphs, making federated queries across

institutional boundaries possible.

3.2. Data Sources and Knowledge Extraction

The KG is populated from four categories of data sources. First, published catalogs from major museums - the Metropolitan Museum of Art, the British Museum, and the Rubin Museum of Art - supplied 1,247 Thangka entries with high-resolution images and metadata. Second, textual sources, including canonical iconographic treatises and contemporary scholarly monographs, provided the detailed descriptions needed for entity extraction. Third, structured interview data collected from 18 senior Thangka masters across Tibet, Sichuan, and Qinghai (under institutional ethics approval) supplied living knowledge not found in published sources. Fourth, online digital archives, including the Himalayan Art Resources (HAR) database and the Tibetan Buddhist Resource Center (TBRC) catalog, filled out additional entries. Knowledge extraction uses a hybrid approach combining manual curation with automated information extraction. For structured metadata (museum catalogs), direct mapping to the ontology is performed using SPARQL CONSTRUCT queries. For unstructured textual sources, a Named Entity Recognition (NER) pipeline built on a fine-tuned BERT model [7] extracts iconographic entities and their attributes. The NER model was trained on a gold-standard corpus of 3,500 annotated sentences drawn from published Thangka literature and achieved an F1 score of 0.82 for entity recognition and 0.76 for relation extraction.

3.3. Knowledge Graph Storage and Query Architecture

The constructed KG is stored in a hybrid architecture that pairs a triplestore (Apache Jena Fuseki) for RDF data with a graph database (Neo4j) for high-performance traversal queries. The RDF triplestore supports SPARQL 1.1 queries and enables reasoning via the OWL 2 RL rule set. The Neo4j instance stores the same KG as property graphs, optimized for path queries (e.g., 'find all artisan lineages connected to the Karma Gadri school'). A synchronization layer maintains consistency between the two storage backends. The query architecture exposes a RESTful API that supports SPARQL endpoints, JSON-LD serialization, and a web-based visual query builder for non-technical users such as curators and artisan educators.

4. Prototype System and Cross-regional Application Scenarios

4.1. System Architecture

and Interface A prototype web application, the Thangka Cultural Knowledge Graph Platform (TCKGP), was built to demonstrate the KG and evaluate its performance. The frontend uses React.js and D3.js to let users interactively explore the knowledge graph. Users can navigate Thangka entities through several visualization modes: network graphs showing entity relationships, faceted browsing filtered by iconographic attributes, and temporal maps showing how painting schools diffused geographically and across time. The platform supports four languages - Tibetan, Chinese, English, and Nepali - reflecting the cross-regional character of the Thangka tradition itself.

4.2. Scenario 1

Museum Digital Curation The first application scenario

centers on digital curation at partner museums (the Rubin Museum of Art in New York and the Tibet Museum in Lhasa). In the conventional workflow, curators manually tag Thangka paintings with free-text descriptions, leading to inconsistent metadata across institutions. Using the TCKGP, curators retrieve semantically enriched metadata from the KG: standardized iconographic tags, related deities, historical lineages, and cross-collection links. A controlled evaluation with 12 curators, comparing conventional metadata against KG-enhanced metadata, found that the KG approach cut average cataloging time by 34 percent and improved metadata consistency (measured by inter-annotator agreement) from Cohen's $\kappa = 0.51$ to $\kappa = 0.84$.

4.3. Scenario 2

Online Artisan Training The second scenario addresses cross-regional artisan training. Traditional Thangka apprenticeship requires prolonged physical co-presence with a master - a requirement that limits access for students in remote or cross-border locations. The TCKGP offers a structured, KG-driven learning pathway: students navigate from foundational iconographic concepts (such as understanding mudra types) to advanced synthesis (such as identifying regional stylistic variations) through semantically linked learning resources. A pilot training program involving 45 students across Tibet, Sichuan, and Kathmandu used the platform for six months. Pre- and post-training assessments showed an average improvement of 41 percent in iconographic literacy scores, against 18 percent in a control group that used conventional video-based training materials.

4.4. Scenario 3

Transnational Cultural Education The third scenario involves transnational cultural education, aimed at diasporic Tibetan communities and global audiences more broadly. [4] The TCKGP was integrated into a virtual exhibition platform [21] that combines KG-driven semantic navigation with immersive 3D viewing of high-resolution Thangka scans. Visitors to the virtual exhibition can click on any iconographic element (a specific deity attribute, for instance) to retrieve contextual information sourced from the KG: ritual significance, regional variations, and related narrative stories. Analytics from a three-month exhibition period (October–December 2025) recorded 23,400 unique visitors from 47 countries. A post-visit survey ($n = 1,205$) found that 78 percent of respondents rated the KG-enhanced contextual information as 'very helpful' or 'extremely helpful' for understanding the cultural significance of Thangka.

5. Evaluation and Results

5.1. Semantic Retrieval Performance

We evaluated the semantic retrieval performance of the Thangka KG against a conventional MySQL relational database containing the same dataset (1,247 Thangka entries). A test set of 200 queries was designed, covering three query types: keyword-based lookup (e.g., 'Thangka paintings depicting Manjushri'), multi-hop reasoning (e.g., 'Thangka painted by apprentices of Karmapa lineage masters active in the 19th century'), and cross-collection discovery (e.g., 'Thangka in European museums similar to those in the Rubin Museum'). Precision, recall, and Mean Average Precision (mAP) were computed. The results show that the KG approach outperforms the relational database across all three

query types. For keyword-based lookup, mAP rose from 0.41 (relational) to 0.78 (KG), largely because the KG can resolve synonyms and hierarchical concepts (retrieving all manifestations of Manjushri, including Jampelyang and Manjughosha, in one query). For multi-hop reasoning queries, the relational database managed mAP = 0.12 (limited to pre-defined join paths), while the KG achieved mAP = 0.69 by leveraging transitive closure over the graph structure. These findings confirm that knowledge graph representation carries particular advantages for cultural heritage data, which is characterized by dense, interlinked semantic relationships.

5.2. Cross-regional Inheritance Effectiveness

To assess whether the KG supports cross-regional inheritance effectively, the authors conducted a mixed-methods evaluation involving 68 participants across three regions (Tibet/Lhasa, Sichuan/Chengdu, and Nepal/Kathmandu). Participants were divided into three groups corresponding to the three application scenarios (curators, artisan students, cultural educators). Pre- and post-intervention questionnaires measured domain knowledge, cross-regional awareness (defined as understanding of Thangka traditions beyond one's own region), and intention to engage in cross-regional collaboration. The quantitative results show statistically significant improvements across all three dimensions ($p < 0.01$, paired t-test). Cross-regional awareness showed the largest effect size (Cohen's $d = 1.24$), suggesting that the KG's ability to visually map cross-regional linkages is effective at broadening practitioners' geographic and cultural imagination of the tradition. Qualitative interviews added another layer: artisan students in Nepal, through the platform, discovered stylistic influences from the Karma Gadri school (Kham region, Sichuan) in their own local traditions - a connection they had not known about, most likely because cross-border scholarly exchange has been limited.

6. Discussion

6.1. Theoretical Implications

This study makes three theoretical contributions. First, it extends knowledge graph theory into the domain of Tibetan Buddhist visual culture and shows that domain-specific ontologies - when properly aligned with general cultural heritage standards such as CIDOC CRM - can achieve both semantic depth and interoperability. Second, it advances the concept of 'relational inheritance' in ICH digital preservation. Rather than treating ICH as a collection of discrete artifacts or performances, the KG approach makes visible the relational fabric - lineages, rituals, regional variants - that constitutes the living fabric of the tradition. Third, it supplies empirical evidence for the 'cultural flows' theory [2] [1] in a digital context, showing how digital infrastructures can either reinforce or mitigate asymmetric global cultural flows depending on whether local communities are involved in ontology design and curation.

6.2. Practical Implications and the TCKG Model

For practitioners, this study proposes the TCKG Model (Thangka Cultural Knowledge Graph) as a reusable framework for ICH knowledge graph construction. The model has five iterative phases. Ontology Scoping comes first: define the cultural scope and align with community

stakeholders. Data Source Mapping follows: audit available digital and non-digital sources. Knowledge Extraction comes next: hybrid manual and automated extraction. Graph Population and Validation follow: expert review by senior artisans and scholars. Cross-regional Deployment closes the loop: deploy the KG in culturally appropriate interfaces supporting multiple languages and access modalities. One practical insight that emerged as critical is the importance of involving source community representatives in ontology validation. In our project, two validation workshops with senior Thangka masters revealed that certain iconographic relationships encoded in the initial ontology conflicted with esoteric interpretations restricted to initiated practitioners. The ontology was revised to include an 'esoteric access control' layer: sensitive relationships are marked with provenance restrictions and accessible only to authenticated users with community endorsement. This finding points to a broader challenge in indigenous and sacred knowledge domains: balancing open knowledge sharing with cultural sensitivity.

6.3. Limitations

Several limitations need to be acknowledged. First, the current KG covers 1,247 Thangka entries - substantial, but only a fraction of the estimated 50,000 or more Thangka paintings held in global museum and monastery collections. Scaling the KG will require better automated extraction methods and additional institutional partnerships. Second, the NER pipeline achieves 0.82 F1 on published texts but performs less reliably on classical Tibetan language sources, where training data is scarce. Future work will explore cross-lingual transfer learning from Chinese and English annotated corpora to classical Tibetan. Third, the cross-regional evaluation, while covering three regions, does not yet include diasporic Tibetan communities in Europe or North America, whose perspectives on digital inheritance may differ substantially from those of participants in the current study.

7. Conclusion

This paper has presented the design, implementation, and evaluation of a cultural knowledge graph for Thangka and its application to cross-regional inheritance. By systematically representing the iconographic, lineage, and ritual dimensions of Thangka as a semantic knowledge graph, the approach overcomes several key limitations of conventional digitization: fragmentation, decontextualization, and limited semantic discoverability. Empirical evaluation across museum curation, artisan training, and transnational education scenarios shows that the KG approach substantially improves semantic retrieval performance and supports cross-regional inheritance effectiveness. The study contributes to two scholarly communities. To digital heritage scholarship, it offers a reusable ontology and a practical five-phase methodology (the TCKG Model) for applying knowledge graphs to ICH. To Tibetan Buddhist art studies, it provides the first formally specified, machine-readable ontology for Thangka iconography and lineage, opening new possibilities for computational art history and digital humanities research on Himalayan art. Three directions for future research present themselves. First, we will scale the KG to cover 10,000 or more Thangka entries through partnerships with additional museums and monastery collections. Second, we will build a community co-curation portal that lets certified artisan communities directly contribute to and validate KG entries,

addressing the sustainability challenge posed by expert-dependent curation. Third, we will explore integrating the KG with immersive technologies (AR/VR) [3], which would let users navigate Thangka's semantic universe through spatial interfaces. As digital and cultural infrastructures become increasingly intertwined, knowledge graphs offer a powerful means to ensure that global accessibility does not come at the cost of local meaning - a balance essential for the sustainable cross-regional inheritance of intangible cultural heritage.

References

- [1] Appadurai, A. (1986). *The social life of things: Commodities in cultural perspective*. Cambridge University Press.
- [2] Appadurai, A. (1990). Disjuncture and difference in the global cultural economy. *Theory, Culture & Society*, 7(2), 295–310.
- [3] Bekelle, M. K., Pierdicca, R., Frontoni, E. S., Malinverni, E. S., & Gain, J. (2018). A survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on Computing and Cultural Heritage*, 11(2), 1–36.
- [4] Burgess, J., & Green, J. (2018). *YouTube: Online video and participatory culture* (2nd ed.). Polity Press.
- [5] Cameron, F., & Kenderdine, S. (Eds.). (2007). *Theorizing digital cultural heritage: A critical discourse*. MIT Press.
- [6] Chibnik, M. (2003). *Crafting tradition: The making and marketing of Oaxacan wood carvings*. University of Texas Press.
- [7] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171–4180).
- [8] Doerr, M. (2003). The CIDOC Conceptual Reference Module: An ontological approach to semantic interoperability of metadata. *AI Magazine*, 24(3), 75–92.
- [9] Giaccardi, E. (Ed.). (2012). *Heritage and social media: Understanding heritage in a participatory culture*. Routledge.
- [10] Henss, M. (2014). *The cultural monuments of Tibet: Divine images*. Prestel.
- [11] Isaac, A., & Clayphan, R. (2013). Linked open data for cultural heritage. In *Digital Heritage International Congress 2013* (pp. 365–372). IEEE.
- [12] Jackson, D., & Jackson, J. (1988). *Tibetan thangka painting: Methods and materials*. Snow Lion Publications.
- [13] Jackson, D. P. (2015). *Tibetan thangka painting: Iconography and interpretation*. Serindia Publications.
- [14] Kirshenblatt-Gimblett, B. (2004). Intangible heritage as metacultural production. *Museum International*, 56(1–2), 52–65.
- [15] Li, X., Zhang, Y., & Wang, H. (2021). Ontology-based knowledge graph construction for traditional Chinese handicrafts. *Journal of Cultural Heritage*, 47, 177–186.
- [16] Parry, R. (2010). *Digital heritage and culture: Strategy and policy*. Routledge.
- [17] Ren, Y., & Lam, C. T. (2025). Knowledge graph-driven digital preservation of intangible cultural heritage: A cross-cultural comparative study of Chinese and Western implementation paradigms. *Humanities and Social Sciences Communications*, 12, 156.
- [18] Smith, L., & Akagawa, N. (Eds.). (2009). *Intangible heritage*. Routledge.
- [19] UNESCO. (2003). *Convention for the safeguarding of the intangible cultural heritage*. UNESCO.
- [20] UNESCO. (2008). *Operational guidelines for the implementation of the convention for the safeguarding of the intangible cultural heritage*. UNESCO.
- [21] Van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.
- [22] Yang, F., Zhang, L., Liu, Y., & Chen, W. (2020). Knowledge graph construction for architectural heritage of the Ming and Qing dynasties. *Automation in Construction*, 118, 103312.
- [23] Zhou, Y., Sun, L., & Huang, R. (2019). Digital preservation of intangible cultural heritage in China: Status, challenges, and prospects. *Library Trends*, 67(4), 621–640.