

FROM WORDS TO WORLDS: A MULTIMODAL ANNOTATION FRAMEWORK FOR POST-TANG CHINESE LYRIC

YAO SONG ^a

Abstract This paper presents a methodological framework for the digital curation of the post-Tang Chinese lyric — the *ci*, *qu*, and *sanqu* composed for musical realization between the late Tang and the end of the Qing (ninth century to 1911). The framework coordinates three mutually constraining annotation layers: a textual layer containing the canonical poem, variants, and the commentarial tradition; a multimodal layer linking the text to its notated music — chiefly the *jianzipu* (減字譜) tablature of the *qin*-song repertory, alongside the *suzipu* pitch notation of the *sung ci* — and to its calligraphic carriers (ink tracings, rubbings, engraved transcriptions); and a *Six-W* provenance layer capturing who wrote, addressed, and recited the poem, what it sets out to do, when and where it was composed, why it was occasioned, and how it was transmitted. Situated against contemporary annotation and provenance models (TEI, W3C PROV-O, FRBR, CIDOC-CRM, Dublin Core, and the W7 model), the paper articulates an AI-assisted annotation pipeline — NLP for the textual layer, computer vision and handwriting recognition for the calligraphic layer, optical music recognition for the score layer — under a human-in-the-loop protocol that preserves the philological prerogative of interpretive judgment. The argument is illustrated by one fully worked case drawn from that dataset, the *qin* song *Gu Yuan* (古怨) in Jiang Kui's *Baishidaoren Gequ*, whose surviving editions support a documented evidential gap that the schema is designed to carry rather than resolve. The framework is not a proposal awaiting implementation: the pipeline described here has been built and run across all three annotation layers, and the resulting dataset comprises approximately 1.3 million post-Tang lyric compositions, each carrying a *Six-W* provenance record, with the full three-layer treatment applied wherever a non-textual witness survives. The dataset is not yet publicly released; the schema, the pipeline, and the epistemic principles are set out here in full so that the data may be assessed independently of the release schedule.

Keywords: multimodal annotation, Chinese lyric poetry, *jianzipu* tablature, calligraphy, provenance ontology, digital philology

1. INTRODUCTION

The so-called *jintishi* of the High Tang and the lyric *ci* of the late Tang already presuppose a communicative situation in which the poem is performed, sung, inscribed, and circulated. The *ci* of the Song, the *qu* of the Yuan, and the *sanqu* and irregular *ci* of the Ming and Qing inherit and intensify this situatedness: each composition is bound up with a specific *cipai*

^a Academy of Chinese, History, Religion and Philosophy, Hong Kong Baptist University

(metric pattern), with a particular occasion and addressee, with a tradition of musical realization codified in the *ci pu*, with a calligraphic body — sometimes an ink trace by the poet himself, more often the work of an admired later hand — and with a commentarial and editorial history that can span almost a millennium. If a poem is, as Stephen Owen has argued, an event as much as a text, then a digital edition that returns only the bare characters surrenders most of what distinguishes the lyric from the regulated verse (Owen 2006). To put it differently: the post-Tang lyric has always been multimodal — the editing tradition alone is a multimodal chain — and the recent history of textual scholarship, in which philology has become gradually computerized, has simply inherited a category mistake that the present framework attempts to correct.

The established digital corpora for Chinese poetry are organized around the text and around the persons who wrote it: the *Quan Tang Shi* electronic editions supply the standard full-text corpus of Tang poetry, the *Siku Quanshu* derivatives a full-text archive of the imperial library, and the China Biographical Database (CBDB) a prosopographical record of the officials, literati, and kin documented in the sources (CBDB 2008; Chinese Text Project 2006; *Quan Tang Shi* 1706). Each of these resources is authoritative within the domain it was designed for, and each is text-first by design rather than by default. What none of them is designed to carry is the relation between a poem and its non-textual records — the notation under which it was sung, or the brushwork of the manuscript that transmits it. Where such images are held, they serve as illustrations rather than as analytically linked data. The next generation of digital curation for Chinese poetry has to do this bridging, and to do it methodologically before doing it technologically.

Two recent efforts are proximate antecedents. Wang Jun and colleagues at Peking University have developed the Wuyudian smart-annotation platform, which applies large language models to automatic punctuation, named-entity recognition, and relational tagging of pre-modern Chinese texts (Wuyudian 2023). The same group's *Evol* project, published in *Humanities and Social Sciences Communications*, integrates text reuse detection, diachronic *n*-gram analysis, and word co-occurrence for two millennia of Chinese literature (J. Wang et al. 2024). Wuyudian shows that a function-specific annotation environment is feasible at scale; neither project, however, is organized around the multimodal coupling of text, score, and calligraphy; both are designed for the text and its annotation.

The empirical foundation for the framework is the Tang-Song Literature Chronology Map, a continuously updated digital infrastructure that maps the composition, circulation, and reception of Chinese poetry from the Tang through the Song and, in its current extension, into the Five Dynasties, Yuan, Ming, and Qing. The Chronology Map links each poem to its year of composition (where recoverable), its geographical locus, the biographical data of its participants, and the editorial chain through which it has descended to the present. The Chronology Map was designed from the outset as a spatio-temporal database rather than as a text archive: every entry carries a date layer (reign year, *ganzhi* cycle, modern calendar equivalent), a place layer (historicized toponym, modern coordinate reference), and an agent layer (author, addressee, commentator, calligrapher). The framework presented in this paper extends the Chronology Map's architecture into the post-Tang period and adds the two modalities that the Map in its current form does not carry: the musical score and the calligraphic carrier. The result is a multimodal research infrastructure that combines verified spatio-temporal provenance with text, score, and calligraphy in a single interoperable schema.

Two constraints on the scope of the corpus should be stated at once, since they govern the reading of every later section. First, the full three-layer treatment — text, score, and calligraphic carrier bound to one another — is applied to those compositions for which at least one non-textual witness survives, whether a notated score, a calligraphic carrier, or both. Depth of treatment therefore varies with the surviving evidence, and this should be acknowledged rather than glossed: the fully layered subset is a corpus of well-documented works, not a representative sample of the tradition, and the two must not be confused. Second, the corpus itself — the schema's textual and provenance layers — comprises approximately 1.3 million post-Tang lyric compositions, a statement of catalogued holdings rather than a canon or a claim of complete coverage. Because a corpus of that size cannot be reviewed item by item before publication, review proceeds by sampling and increment: the pipeline annotates the corpus in full, and the human-in-the-loop protocol described in Section 4.2 governs which annotations are certified and in what order, the certified and the uncertified sets being released as separate streams (Section 6). Entries are graded accordingly: a composition attested in all three modalities receives the full layered treatment, while one attested in text and notation only is recorded as a partial entry with the missing modality marked unrecorded rather than left blank. This framework carries no claim to be the largest resource of its kind. The claim it does make is narrower and checkable: that it is, so far as we have been able to establish, the first to bind text, notation, and calligraphic carrier for the same poem within a single provenance schema.

A single example will clarify the stakes. Consider Jiang Kui's (姜夔, ca. 1155–1221) Yangzhou Man (扬州慢), a ci composed in 1176 (淳熙丙申) upon the poet's return to a war-ravaged Yangzhou. The cipai is not an inherited pattern but the poet's own invention — a ziduqu (自度曲, self-composed melody) assigned in the collection to the zhonglu gong (中吕宫) — and the poem comprises ninety-eight characters in two stanzas, fifty in the upper register and forty-eight in the lower, according to the normative description of the tune in Wan Shu's Cilü (《词律》, 1687) and the imperial Qinding Cipu (《钦定词谱》, 1715), both of which take Jiang Kui's own text as the standard form (Wan 1687; Y. Wang et al. 1715). The poem is also accompanied by a prose autocommentary (词序) that fixes the occasion and constitutes the primary evidence for the date of composition. The music survives as well: the Baishi Daoren Gequ (白石道人歌曲) transmits Jiang Kui's own notation for the melody in the abbreviated pitch script known as suzipu (俗字谱), the source on which the modern study of Song music largely rests. The same collection, however, contains the piece on which the present framework's score layer concentrates: Gu Yuan (古怨), a qin song and the only item in the collection notated in jianzipu (减字谱), the tablature of the qin. Gu Yuan is examined in Section 3.1 and supplies the worked case. Finally, the poem survives as a philological event: the earliest printed witnesses of the collection — the 1202 edition of Qian Xiwu and the edition Jiang Kui is reported to have cut himself — are both lost, and the text descends instead through the Lu Zhonghui edition of 1743, the Zhang Yishu edition of 1749, the Siku Quanshu copy, and the 1913 edition printed by Zhu Xiaozang from Jiang Bingyan's manuscript of 1737 (Jiang 1202). A digital edition that returns only the ninety-eight characters has surrendered the notation, the autocommentary, the edition history, and the commentarial chain. The Six-W schema asks, for this single ci, who (Jiang Kui as poet and composer; the unknown scribe; the Qing commentators), what (a ziduqu in the zhonglu gong), when (1176, with the edition dates kept distinct from the date of composition), where (Yangzhou as composition site; the editorial sites as transmission loci), why (the occasion recorded in the 词序), and how (text, notation, print). The framework's claim is that these six questions, asked together and answered with evidence

— and left explicitly unanswered where the evidence runs out — constitute the minimum unit of a multimodal digital edition.

Two terminological clarifications should be made at the outset. The first concerns the chronological scope, which is fixed here for the whole paper and used identically in the title and the abstract: by “post-Tang” this study means the lyric composed for musical realization from the late Tang (ninth century) through the end of the Qing (1911) — the *ci* tradition from Wen Tingyun onward, the *qu* of the Yuan, and the *sanqu* and irregular *ci* of the Ming and Qing. The framework is silent on exact periodization and admits a song-cycle from the late Tang as readily as a *ci* composed in 1810. The second concerns the sense of “multimodal”: it is the sense current in digital humanities after Manovich, namely a deliberate alignment of textual, visual, and musical evidence within a single analytical frame, rather than the narrower machine-learning sense in which “multimodal” refers to contrastive training between image and caption (Manovich 2020; Baltrušaitis et al. 2019). The framework treats every modality as evidence; the only question is how to align them — and, where the evidence is incomplete, how to say so.

2. RELATED WORK AND POSITIONING

The framework inherits from three converging strands of scholarship: an annotation turn in digital humanities, a computational turn in classical Chinese studies, and a multimodal turn in cultural analytics.

The annotation turn is most closely associated with the Text Encoding Initiative (TEI) and with the work of Ide and Romary on abstract data models for linguistic annotation (Burnard 2014; Ide and Romary 2002). The position common to this work is that encoding decisions are hermeneutic decisions: the categories a markup scheme makes available constrain what can be said about a text, so that markup is a constitutive rather than a merely technical layer of interpretation (Ide and Sperberg-McQueen 1995). TEI provides granular tags for persons, places, dates, and editorial interventions; CIDOC-CRM offers a complementary vocabulary for cultural objects (ICOM/CIDOC 2021); Dublin Core supplies the lightweight fifteen-element set (DCMI 1999); W3C PROV-O formalizes provenance among entities, activities, and agents (W3C 2013); and FRBR distinguishes the work, expression, manifestation, and item layers (IFLA 1998). The framework treats these as a minimum usable intersection — communities (Chinese philology, *ci* studies, *qu* studies) are free to extend it.

The computational turn in Chinese classical studies is dense. Jockers’s text-analysis infrastructure has been adapted to Chinese (Duan et al. 2023; Jockers 2014). The SikuBERT family of pre-trained language models, trained on a billion characters of classical Chinese, has brought classical Chinese NLP into a usable regime (D. Wang et al. 2022). The Peking University Wuyudian platform operationalizes these capabilities for philologists — automatic punctuation, NER, relational tagging; the same group’s Evol platform integrates text reuse detection and diachronic n-gram analysis for a two-thousand-year corpus (J. Wang et al. 2024). Wang’s group has also published benchmark corpora such as ChisNERE (Tang et al. 2025). A more general point, argued by Schöch for the humanities at large, bears directly on the present framework: the value of such data is a function of its documentation rather than of its size, since large, smart, clean, and messy data are not degrees on a single scale, and a corpus that

does not record how its categories were assigned cannot support the inferences its size appears to license (Schöch 2013). That argument is the reason this paper claims a documentation scheme rather than a volume of material.

The multimodal turn begins with Manovich's Cultural Analytics (Manovich 2020). Elkins has argued for the methodological independence of visual analysis (Elkins 1999). Transkribus, eScriptorium, and related platforms have demonstrated that manuscript image data can be productively processed at scale (Coquenot et al. 2023; Transkribus 2015; eScriptorium 2019). For Chinese calligraphy, the question of whether a manuscript is an ink trace, an engraved transcription, or a rubbing is itself an analytical claim that should be carried forward into the digital edition. Aligning a manuscript hand with the musical score under which the poem was sung falls outside the remit of these platforms, which address the image and the text it carries; the present framework is, to the best of my knowledge, the first to make that alignment an explicit object of annotation.

Two provenance models bear more directly on the present framework than the general annotation standards do, and the Six-W schema is best understood against them. Ram and Liu's W7 model was developed as an ontological account of data provenance and comprises seven interconnected elements — what, when, where, how, who, which, and why — each of which tracks events affecting data during its lifetime (Ram and Liu 2009). Pamart, De Luca, and Véron have applied a W7-inspired scheme to the documentation of multimodal imaging surveys of cultural heritage, showing that a question-set of this kind can be made operational across heterogeneous capture data (Pamart, De Luca, and Véron 2022). The schema proposed here therefore does not claim the question set as its novelty. The six questions, and the “which” that Ram and Liu add to them, are prior art, and W7 in particular supplies a vocabulary for provenance as a property of events rather than of documents. What is claimed as new is representational. First, the object whose provenance is recorded is not a document but a poetic event with several distinct histories — composition, inscription, printing, cataloguing, and interpretation — which must be modelled separately if they are not to be conflated. Second, the schema is required to hold conflicting values side by side, each with its own warrant, because the philological record frequently does not converge on one. Third, the schema records the history of assertions about the poem as distinct from the history of the poem, a distinction for which the provenance literature has no occasion and which W7's event vocabulary does not by itself supply. The contribution, in short, is a set of representational decisions about what a poem is in a database, not the discovery of the questions one may ask of it.

Three questions that Chinese-language scholarship has long answered in prose remain, at the time of writing, largely unformalized in digital form. The first is that of a multimodally coherent ontology for *ci*: the existing *ci* philology (词学) is densely typological — *cipai*, 宫调, 平仄, 韵脚 — and these distinctions are understood with great precision, though they have not been expressed as relations in a digital schema (Ma, Wu, and Cao 1996; Lam 1998; Lam 2012). The second is the cross-modal alignment between the *ci pu* (词谱) and the texts themselves: the *ci pu* tells one how a *cipai* was realized musically, and it remains to link each character of the lyric to the corresponding pitch or syllable in the melody in machine-readable form. The third is the machine-readable continuity between a poet's draft and a contemporary calligraphy manual: the calligraphic evidence is normally treated as art-historical data, and its philological relation to the text has not been given a digital form. Taken together, these three questions define the space the framework sets out to occupy — the domain it addresses by

binding, rather than merely collocating, the textual, musical, and calligraphic records of the same poem.

A fourth question may be registered alongside these three. The commentarial tradition — the 笺注 (jianzhu), 评点 (pingdian), and 本事 (benshi) literatures that accompany virtually every canonical ci and qu from the Song onward — is not yet represented as a first-class data type in the digital infrastructure for Chinese poetry. CBDB records the commentators among its persons, and does so with a precision that makes it the natural reference point for the prosopography of commentary, though its design links them to the biographical record rather than to the individual poem they gloss; the Quan Tang Shi electronic editions present the poems as text, the commentarial apparatus being a feature of the print tradition they digitize rather than of the poems themselves. The framework treats the commentarial tradition as a first-class data type: every comment is an annotation event with its own Who (the commentator), When (the date of the commentary), and How (the medium through which the comment was transmitted — manuscript marginalia, printed appendix, oral tradition). This treatment follows the hermeneutic principle, common to the annotation literature from the TEI onward, that annotation is a constitutive rather than a merely descriptive component of a text’s meaning (Ide and Romary 2002; Table 1).

Table 1 Comparison of Digital Poetry Corpora and Annotation Platforms by Modality Coverage and Schema Features

Platform / Corpus	Text	Score	Calligraphy	Provenance	Spatio-temporal	AI-assisted	Open schema
Quan Tang Shi (electronic)	Yes	No	No	No	No	No	No
CBDB	Yes	No	No	Partial	Partial	No	No
Siku Quanshu derivatives	Yes	No	No	No	No	No	No
Wuyudian (PKU)	Yes	No	No	Partial	No	Yes	No
Evol (PKU)	Yes	No	No	Partial	No	Yes	No
Chronology Map (Tang–Song)	Yes	No	No	Yes	Yes	No	Partial
This framework (post-Tang lyric corpus; 1.3 million compositions)	Yes	Yes	Yes	Yes (Six-W schema)	Yes (Chronology Map, Tang–Qing)	Yes	Yes

Note: For the existing resources the entries report capabilities that have been implemented and documented by those projects; for the framework described here the entries report the state of the dataset described in Section 6, which has been built and is not yet publicly released. The resources also differ in purpose — full-text archive, prosopographical database, annotation platform, spatio-temporal database — and the comparison is therefore one of dimensions covered, not of overall completeness or quality. The Quan Tang Shi electronic editions and CBDB are authoritative reference resources in their respective domains — the Tang corpus and the biographical record of the Chinese elite across the imperial period — and the framework described here is designed to complement rather than to supersede them; a No records that a dimension lies outside a resource’s design scope, and implies no judgement of its quality. A modality is marked Yes only where the resource links that modality to the text at the level of the individual poem; for the framework described here the score and calligraphy columns are marked Yes because the pipeline has been run over those layers, while the depth of treatment varies with the surviving evidence, as Section 1 explains.

3. THE SIX-W PROVENANCE SCHEMA

The Six-W schema is the framework’s signature contribution. Its design is governed by three

principles. First, it should be at the same time a formal ontology — every field carries an unambiguous type — and a historian-friendly vocabulary — every field corresponds to a term in Chinese philological tradition. Second, it should be a minimum usable intersection with PROV-O, FRBR, CIDOC-CRM, and Dublin Core, so that any implementation that wishes to publish aligned data can do so. Third, it should be silently extensible: each W is defined as an interface rather than an enumeration, so that domain communities — regional philologists, Ming ci specialists, qu scholars — can add local vocabularies without invalidating the schema.

A fourth commitment should be made explicit. The schema treats the Six-W fields not as a checklist to be completed but as a question set against which the editor must justify each entry. To record the when of a ci as “1176” without saying on what evidence is to produce a confidently wrong fact; to record the what of a poem as “咏史” without also pointing to its literary tradition is to reduce philology to a tagging exercise. The fields record promises that the annotation will be paired with evidence; the framework explicitly treats an unevidenced entry as a withdrawn claim rather than a default value. This posture is continuous with the evidentiary discipline of Chinese collation scholarship. In his *Four Methods of Collation*, Chen Yuan (陈垣) reserves the most powerful of his procedures — emendation by reasoning (理校法) — for cases in which no earlier witness can be had, and warns that it is at once the most brilliant and the most dangerous of the four (Yuan Chen 1931). In his own practice he therefore restricts it sharply, holding that it may be applied only to the plainest of errors and that “without solid evidence one dare not invoke reasoning and rely on conjecture” (非有确证, 不敢藉口理校而凭臆见也). The framework’s treatment of an unevidenced entry as a withdrawn claim rather than a default value is the same rule applied to a database.

Each W is briefly characterized below; the consolidated field table is Table 2, and a machine-readable worked entry is given in Section 3.1.

Who. The conventional conception of authorship, inherited from CBDB and from prosopographical databases generally, treats the poet as the unique agent of the poem (CBDB 2008). The schema treats who more carefully. A poem is an event whose participants include the author and the addressee, and — in the commentarial tradition — the commenter, the rubricator, the music editor, the calligrapher, the collector, and the cataloguer. Each who is a uniform agent type with a name, a role, a temporal interval, and a geographical locus; the schema allows multiple authors for collected or revised poems, and multiple addressees for zengda (赠答) and changhe (唱和) exchanges. One further distinction is required by the field’s own logic. Anonymous, disputed, and inapplicable attributions are three different states and are recorded as three different values. A poem whose author is unknown carries the value unrecorded for agent.author, together with a warrant naming the witness that fails to name an author; a poem whose authorship is disputed carries two or more competing agent entries, each with its own warrant; and a field that a given work does not possess — an addressee, for a poem that has none — is marked not_applicable rather than left absent. Every field in the schema carries this three-valued status flag (asserted / unrecorded / not_applicable), so that a downstream query for “poems with no known author” does not also return poems whose authorship was simply never entered.

What. The what field identifies the poem’s genre, topic, and stylistic field. Genre inherits from a controlled vocabulary — shi, ci, qu, ge, fu — with the cipai or qupai recorded as a sub-vocabulary. Topic inherits a typology common to the ci tradition (咏物, 咏史, 题画, 送别,

艳情), but the schema is open. The stylistic field carries the jingdian (经典) references — the Shijing, the Chu ci, the major Tang models — that the poem has actualized, as well as the pai (派) or lineage to which it is conventionally assigned. The same three-valued flag applies here. For an anonymous or fragmentary work the genre may be asserted while the topic remains unrecorded; for a work whose tradition supplies no lineage, stylistic field is not applicable rather than empty. The distinction matters because it separates a slot that is merely empty from a slot that has been examined and found inapplicable, and only the second of these is a finding.

When. When is not a single date but a layered set: the date of composition (which for ci is sometimes recoverable only to a guanju interval of years), the date of inscription on a manuscript if distinct, the date of the first printed edition, the date of a cipu's first fitting of the cipai, and the date of the surviving calligraphic carrier. The layer inherits the time ontology for premodern Chinese texts proposed by Wang, Wang, and Wei, including the conventional reign-year calendar (帝王纪年), the ganzhi (干支) cycle, the month, the jieqi (节气) marker, and the shi-chen (时辰), together with their modern Gregorian conversions (Wang, Linxu, Wang, and Wei 2022; Wang, Linxu, Wang, and Wei 2024). The Chronology Map provides the temporal scaffolding for the When field at scale: its year-by-year mapping of Tang and Song poetic composition, now being extended into the Five Dynasties, Yuan, Ming, and Qing, allows the editor to inherit a verified chronology rather than reconstruct one from scratch for each entry. The Map's integration of reign-year, ganzhi cycle, jieqi marker, and modern Gregorian conversion supplies precisely the layered temporal record that the When field requires, and does so for a corpus whose chronological span exceeds a millennium.

The most important constraint on this field is not what it records but what it keeps apart. A date of composition and a date at which an editor adopted a proposed date of composition are different facts with different agents, and the schema stores them separately. The composition date belongs to the poem; the adoption of it belongs to the history of an assertion about the poem. The same holds for authorship: the poet who wrote the poem and the scholar who proposed an attribution are distinct agents acting at distinct times on distinct warrants. In practice this means that the schema does not permit a date to be entered directly on a text. It requires an assertion entity — a typed, reified statement with its own identifier, its own agent, its own date of proposal, its own warrant, and its own review history — which the scholarly record can then confirm, contest, or supersede without altering the object it is about. Three consequences follow. A value in the When field is always a pointer to an assertion rather than a literal. Contradictory assertions coexist and are never merged. And because each assertion carries its own review status, a dataset can be queried for the philological state of a claim independently of the state of the poem it describes.

Where. Where is layered: the place of composition, the place of transmission (the libraries, presses, and collections through which the text has passed), and the place of physical instantiation. Geographical references are recorded both as historicized place names (e.g., 扬州 in 1176) and as modern coordinate references linked through the China Historical GIS (CHGIS 2001). The Chronology Map's geographic layer, which links each historicized toponym to a modern coordinate through the China Historical GIS, provides the spatial infrastructure for the Where field at corpus scale. For the post-Tang extension, the framework inherits the Map's place-authority file and adds the transmission sites — libraries, printing houses, private collections — that the Map does not record but that are indispensable for tracing the itinerary of a ci from its composition site to its surviving manuscript or printed edition. The

result is a four-layer geography: composition place, inscription place, transmission itinerary, and current physical location, each carrying its own temporal stamp.

Why. Why captures the occasion and the motivating event — the Yangzhou Man of 1176 was occasioned by a huaichu (怀旧) feeling induced by the devastation left by the Jin incursion of 1161 (胡马窥江); a ci of 1670, by a shida (寿诞) song on the seventieth birthday of a friend. The schema distinguishes occasion (an external social fact: a birthday, a banquet, a military event) from cause (the inferred motivation); recording why is the most hermeneutically demanding of the Ws and the one that most rewards human-in-the-loop validation. In the Yangzhou Man of 1176 the occasion is recorded rather than inferred, since Jiang Kui's own prefatory note is the warrant; where no such note survives, the framework requires the inference to be labelled as one, so that a reconstruction is never mistaken for testimony.

How. How is the modality that distinguishes this framework from text-only corpora: it records not only the editorial transmission but also the performance pathway — whether the poem has been set to music, the cipai under which it was sung, and whether it has been inscribed on a calligraphic carrier (D. Wang et al. 2022). For surviving manuscripts, the how field records the carrier's material class (墨迹本, 刻本, 拓本), the calligrapher's identity (if distinct from the poet), and the layout (横卷, 立轴). The framework treats the how field as the place where textual and non-textual modalities most explicitly converge; it is the modal hinge that justifies the framework's commitment to a Three-Layer pipeline rather than to a textual corpus with appended images. The Chronology Map, in its current form, records the textual and prosopographical dimensions of this transmission but does not carry the score or calligraphy data; the framework's post-Tang extension has converted the Map from a text-and-biography database into a multimodal research infrastructure.

Table 2 Six-*W* Provenance Schema: Fields, Definitions, Traditional Equivalents, Reference Ontologies, and Example Value Types

Dimension	Field	Definition	Chinese Equivalent	Traditional	Ontology reference	cross-	Obligation	Example Value Type
WHO	agent.author	Creator of the poem	作者		PROV-O Agent		required	string / URI
WHO	agent.addressee	Recipient of the poem	受赠人 / 题赠对象		PROV-O Agent		optional	string / URI
WHO	agent.communicator	Reciter, performer, or vocalizer	歌者 / 演唱者		PROV-O Agent		optional	string / URI
WHO	agent.annotator	Commentator, 笺注, 评点	注者 / 评者		PROV-O Agent		optional	string / URI
WHO	agent.calligrapher	Inscription calligrapher	书作者 / 抄录者		PROV-O Agent		conditional	string / URI
WHAT	genre	Genre of the poem	体裁 — 诗/词/曲		TEI / FRBR (cross-reference)		required	controlled vocabulary
WHAT	cipai	Metric pattern	词牌 / 曲牌		local vocabulary		conditional	controlled vocabulary
WHAT	topic	Thematic category	题材 — 咏物/咏史/题画		local vocabulary		required	controlled vocabulary
WHAT	stylistic_field	Lineage or school	流派 — 婉约/豪放		local vocabulary		optional	URI
WHEN	date.composition	When composed	作年		PROV-O / W7 (cross-reference)		conditional	interval or date
WHEN	date.inscription	When inscribed on a carrier	题款日期		PROV		optional	interval or date
WHEN	date.publication	First printed edition	刊刻日期		FRBR		conditional	interval or date
WHEN	date.transmission	Catalog event dates	收藏 / 著录日期		PROV		optional	interval or date
WHERE	place.composition	Where composed	作地		TEI / CIDOC		conditional	historicized place URI
WHERE	place.transmission	Libraries, presses	传藏地		CIDOC		optional	URI
WHERE	place.instantiation	Where the carrier is held	收藏地 / 馆藏		CIDOC		conditional	URI
WHY	occasion	External social fact	场合 — 寿诞/悼亡/题画		CIDOC event		conditional	URI
WHY	motivation	Inferred cause	动机 / 本事		local vocabulary		optional	string
HOW	transmission	Editorial pathway	稿本 / 抄本 / 刻本 / 选本		FRBR		required	controlled vocabulary
HOW	performance	Realization mode	吟诵 / 歌咏 / 题壁		local vocabulary		optional	URI
HOW	music_realization	Connection to a <i>cipai</i> or <i>qupai</i>	词谱 / 宫调 / 减字谱		local vocabulary		conditional	URI
HOW	calligraphy.carrier	Material class	墨迹本 / 刻本 / 拓本		CIDOC		conditional	controlled vocabulary
HOW	calligraphy.layout	Composition arrangement	横卷 / 立轴 / 扇面		CIDOC		optional	controlled vocabulary

Note: The schema is intended as an extensible interface; each field may be replaced or refined by a community-specific vocabulary, in keeping with the principles of Section 5. PROV-O entries reference W3C recommendations. The Obligation column states whether a field must be filled and how many values it may take: required = exactly one; conditional = exactly one if the genre possesses the field, otherwise the value not_applicable; optional = zero or one. Cardinality is used in this paper in one sense only — of the relations recorded in the alignment block of Section 4, where it states how many units on one side correspond to a unit on the other (1:1, 1:n, n:0) — and the term is kept off the field table for that reason. Every

field carries the same three-valued status flag (asserted / unrecorded / not_applicable), and no value is stored as a literal — each is a pointer to an assertion entity. The Ontology Cross-reference column names the target of a mapping, not a dependency: the local Chinese term is primary and the ontology term is a cross-reference, as Section 5 requires. Two relations are mandated for every entry and constrain the whole record: asserts (an assertion entity to the work it is about) and is_warranted_by (an assertion entity to the witness or argument that supports it). No assertion that carries a value may be stored without at least one is_warranted_by link, and no field value may be stored without an assertion. A record whose claim is itself an absence — a field that is unrecorded, or one that does not apply — is the single exception: a negative finding has no witness to cite, and it carries a note stating how far the search for one extended instead.

3.1 A Worked Entry: Gu Yuan (古怨) in the Baishidaoren Gequ

The case used throughout this paper is the qin song Gu Yuan (古怨) in Jiang Kui's Baishidaoren Gequ. It is chosen for four reasons: it is the only qin piece in an otherwise predominantly vocal collection, and therefore the only item in it written in jianzipu tablature; it is the earliest extant notated qin song; its collection survives in a small, enumerable set of witnesses whose relations are documented; and the modern reading of its notation is genuinely contested, which makes it a direct test of whether the schema can carry an evidential gap rather than paper over it. The collection's earliest witnesses are lost: an edition cut by Qian Xiwu in 1202 and an edition Jiang Kui is reported to have cut himself. Four later witnesses stand behind all modern work — the Lu Zhonghui edition of 1743, the Zhang Yishu edition of 1749, the Siku Quanshu copy, and the 1913 edition printed by Zhu Xiaozang from Jiang Bingyan's manuscript of 1737 (Qiu 1956). Because the 1202 editions are lost, the date of composition and the dates of the surviving witnesses are separate facts, and the relation between any two witnesses is a hypothesis about transmission rather than an observed link. The framework records it as such.

The distinction between the history of the poem and the history of an assertion about the poem can be shown on the Yangzhou Man. The composition date of 1176 is asserted on the warrant of Jiang Kui's own prefatory note: the agent is the poet, the date of the assertion is the date of the poem, and the warrant is the 词序. The acceptance of that date a century and a half later into a chronological edition is a second assertion — same value, different agent, different date, different warrant — and the schema stores it as a separate entity rather than as a confirmation of the first (Xia 1958). Were an editor to propose a different year, the two values would coexist, each with its own support, and neither would be privileged by the database.

The second distinction is between a field that is unknown and a field that does not apply. Gu Yuan is a qin song: it has no cipai, because the cipai is a category of the sung ci and not of the qin repertory, so cipai is not applicable. Its occasion is not recorded in any surviving witness, so occasion is unrecorded. A query for works with no recorded occasion returns Gu Yuan; a query for works without a cipai does not. Under a two-valued convention — present or absent — the two cases would be indistinguishable, and the difference between what the tradition did not say and what the schema does not model would be lost.

The evidential gap is not hypothetical. Five modern transcriptions of the same tablature disagree, as the collation that compared them sets out (Yingshi Chen 1986). Yang Yinliu's reading, first published in 1957 from the Zhang Yishu edition and revised in 1979, reads the tablature on the sanfen sunyi (三分损益) temperament in 1957 and on a just-intonation (纯律) one in 1979; Zha Fuxi's, published in 1962 from an eclectic collation of the editions, differs from it; Laurence Picken's of 1957, from the Lu Zhonghui edition, differs from both; and Rulan Chao Pian's of 1967 and Gong Yi's of 1979, both from the Zhu edition, preserve the Song practice of indicating stopped notes between the hui markers instead of converting them into hui-fraction values, which the earlier transcriptions had done, and Pian's alone supplies no metre. The disagreements cluster on four points: the tuning of the 侧商调 mode, the pitch centre and its internal modulation, the reading of inter-hui stopped notes, and metre and rhythm, for which the source carries no notation at all (Yingshi Chen 1986). This is precisely the situation in which a text-first edition would select one transcription and discard the rest. The schema does the opposite: it records all five as assertions, and it leaves the metre unspecified.

```
{
  "@context": {"sixw": "https://example.org/sixw#", "prov":
    "http://www.w3.org/ns/prov#"},

  "work": {
    "id": "sixw:work/baishi-gu-yuan",
    "title": "古怨 Gu Yuan",
    "genre": "qin song (琴歌)",
    "cipai": {"value": null, "status": "not_applicable",
      "note": "cipai is a category of the sung ci, not of the qin
      repertory"},
    "stylistic_field": {"value": null, "status": "not_applicable"}
  },

  "assertions": [
    {"id": "sixw:a/0007", "field": "who", "role": "author",
      "value": "姜夔 Jiang Kui (ca. 1155-1221)", "status": "asserted",
      "agent": {"name": "夏承焘 Xia Chengtao", "date": "1958"},
      "is_warranted_by": "sixw:w/jiang-baishi-ci-biannian-1958"},

    {"id": "sixw:a/0019", "field": "when", "role": "date.composition",
      "value": "1195-1204 (interval)", "status": "asserted",
      "agent": {"name": "editorial team", "date": "2026"},
      "is_warranted_by": "sixw:w/jiang-baishi-ci-biannian-1958",
      "note": "the source preface states the tuning, not a date; the interval
      is an editorial proposal, warranted by the chronological collation"},

    {"id": "sixw:a/0020", "field": "when", "role": "date.publication",
      "value": "1743", "status": "asserted",
      "agent": {"name": "editorial team", "date": "2026"},
      "is_warranted_by": "sixw:w/lu-zhonghui-1743"},

    {"id": "sixw:a/0031", "field": "why", "role": "occasion",
      "value": null, "status": "unrecorded",
      "note": "no surviving witness states an occasion; no inference is
      stored"},

    {"id": "sixw:a/0044", "field": "how", "role": "calligraphy.carrier",
```

```
    "value": null, "status": "not_applicable",
    "note": "the work survives in printed editions only; no calligraphic
    carrier is attested, so the field does not apply"}
],

"expressions": [
  {"id": "sixw:e/lu-1743", "manifestation": "陆钟辉刻本 (1743)",
    "notation": "减字谱 jianzipu; 侧商调 tuning stated in the source"},
  {"id": "sixw:e/zhang-1749", "manifestation": "张奕枢刻本 (1749)",
    "notation": "减字谱 jianzipu"},
  {"id": "sixw:e/siku", "manifestation": "四库全书本",
    "notation": "减字谱 jianzipu"},
  {"id": "sixw:e/zhu-1913", "manifestation": "朱孝臧刻本 (1913, from 江炳炎
    ms of 1737)",
    "notation": "减字谱 jianzipu"}
],

"alignments": [
  {"id": "sixw:al/0001", "layer": "score",
    "unit_a": "lyric syllable (字)", "unit_b": "jianzipu sign (pluck + string
    + hui)",
    "cardinality": "1:n", "status": "asserted",
    "note": "one character may be set to several plucks; targets stored as an
    ordered set",
    "is_warranted_by": "sixw:w/lu-zhonghui-1743"},

  {"id": "sixw:al/0002", "layer": "score",
    "unit_a": "jianzipu sign", "unit_b": "pitch (scale degree)",
    "cardinality": "1:1", "status": "conditional",
    "precondition": "侧商调 tuning as specified in the source preface;",
    "ambiguity": "the tuning itself is read differently by Picken (1957) and
    Yang (1979); reconstructed as a pure-temperament tuning (Chen, Yingshi
    1984)",
    "is_warranted_by": "sixw:w/preface-of-source"},

  {"id": "sixw:al/0003", "layer": "score",
    "unit_a": "pluck sequence", "unit_b": "metrical position",
    "cardinality": "n:0", "status": "unrecorded", "target": null,
    "note": "jianzipu records fingering, not duration; the witnesses mark
    section divisions but no metre, and the modern transcriptions diverge (4/4,
```

```
4/2, mixed, or no metre at all); no value is inferred",
  "downstream": "consumers receive null, never a default reconstruction",
  "is_warranted_by": "sixw:w/lu-zhonghui-1743"}
],

"structural_absences": [
  {"id": "sixw:sa/0001", "layer": "score",
   "unit_a": "jianzipu sign (harmonic section, 泛音)", "unit_b": "lyric
syllable (字)",
   "cardinality": "n:0", "status": "asserted", "target": null,
   "note": "the harmonic section carries plucks and no lyric; the empty side
is a property of the source, and the record is kept in this register rather
than in the alignment list, where an empty target would read as missing
data",
   "is_warranted_by": "sixw:w/lu-zhonghui-1743"}
]
}
```

The listing above — given here in JSON, and expressible in the same shape as TEI with a PROV-O profile — carries the argument in three places. First, every value in the who and when fields is a pointer to an assertion that names both the agent who proposed it and the witness that warrants it; the dataset stores a scholar’s proposal about a date as a claim made in a particular year on the basis of a particular collation, not as a property of the poem. Second, the status flag distinguishes a field that is unrecorded (the occasion) from one that is not_applicable (the cipai, a category the qin repertory does not have); a query for works lacking a cipai will not return Gu Yuan, and a query for works whose occasion is unknown will. Third, the alignment block makes each relation’s units, cardinality, and failure mode explicit, and a separate register holds the mappings whose empty side is a property of the source rather than a gap in the evidence; alignment records carry a status vocabulary of their own, distinct from the field-level flag — asserted, conditional (holding only under the stated precondition), unrecorded, and lacuna (the sign is illegible or the witness damaged). Syllable to sign is one-to-many, because a single character may be set to several plucks. Sign to pitch is one-to-one only under a stated tuning — which is exactly the point at which the five transcriptions diverge, and the reason the tuning is a precondition rather than a footnote. Pluck to metrical position is n:0, many plucks and no recorded metre, and is stored as unrecorded rather than filled with a reconstruction (Yingshi Chen 1986). This last case is where a multimodal edition earns its keep. The framework’s answer to an unresolved evidential problem is to model it, not to solve it: an absence of evidence is a first-class value, and nothing in the pipeline is permitted to substitute a default for it.

4. THE TEXT-SCORE-CALLIGRAPHY PIPELINE

The pipeline is organized in three layers, each tethered to the Six-W schema by means of a shared key.

Layer 1: Textual Layer. The textual layer stores the canonical poem, selected variants, the commentarial tradition, and the editorial chain through which the text arrived at its current form (Tang et al. 2025). Each entry is keyed to a who agent (author, annotator, editor) and to a when provenance field; the encoding inherits TEI conventions for variants, punctuation, and editorial interventions. The textual layer is not merely a storage container for the canonical reading; it is a versioning system. Each variant is recorded with its source — a manuscript siglum, a printed edition, a commentator’s proposal — and with the philologist’s judgment on its authority. The layer borrows from the TEI Critical Apparatus module the distinction between a lemma (the reading being annotated) and a witness (the source that attests it), but extends this distinction to accommodate the specifically Chinese practice of 校勘 (jiaokan, collation), in which the editor’s judgment is itself a datum that must be preserved rather than silently resolved. The layer also carries the autocommentary — a genre peculiar to Chinese lyric poetry, in which the poet himself (as in Jiang Kui’s Yangzhou Man) provides a prose gloss that is neither paratext nor paraphrase but a first-order interpretive act. The framework records the autocommentary as a distinct text type, linked to the poem by a has-autocommentary relation rather than merged into a single text node.

Layer 2: Score Layer. The score layer stores the notated musical evidence for the poem. Three families of notation must be kept apart, because they record different things. The first is the cipu (词谱) — the metrical catalogue, normative rather than musical, which for each cipai records the line divisions, the tonal pattern (平仄), and the rhyme positions, as in Wan Shu’s Cilü (《词律》, 1687) or the imperial Qinding Cipu (《钦定词谱》, 1715) (Wan 1687; Y. Wang et al. 1715; Long 1978). A cipu tells the editor what shape a text must have; it is not evidence for how any particular text was sung, and the framework never treats it as such. The second family records pitch: the gongche system (工尺谱), in its Song-era abbreviated form known as suzipu (俗字谱), which is the notation in which Jiang Kui’s seventeen songs in the Baishidaoren Gequ are set out, and the lülüpu (律吕谱), which appears in court and ritual contexts and is transmitted in the same collection. Notation of this family correlates text with pitch, and the correlation is sometimes written directly against the character, as Zhang Yan observes of contemporary practice in the upper scroll of his Ciyuan (《词源》) (Ciyuan 1297; Xia 1963). For the qu and sanqu repertoires the corresponding metrical and pitch sources are the qupu (曲谱), of which the Jiugong Dacheng Nanbei Ci Gongpu (《九宫大成南北词宫谱》, 1746) is the standard compendium (Jiu Gong Da Cheng 1746). The third family is the jianzipu (减字谱), the tablature of the qin. This is the system on which the present framework concentrates, and the one most often misdescribed. Jianzipu is not a variant of gongche and not a simplification of it: it is a tablature in the strict sense, a notation of instrumental action rather than of pitch. A jianzipu sign compounds a right-hand plucking instruction with a string number and a left-hand position given as a hui (徽) or as a point between hui; the resulting sound is determined only once the tuning of the seven strings is known. The system descends from the earlier descriptive notation (文字谱) of which the Tang manuscript Qie Shi Diao You Lan (《碣石调·幽兰》) is the sole survivor, and it is the notation of the earliest extant anthology of qin music, Zhu Quan’s Shenqi Mipu (《神奇秘谱》, 1425) (Zhu Quan 1425). Because pitch in jianzipu is a function of tuning and position rather than a recorded value, the score layer’s representation of a jianzipu source must carry the tuning as a first-class datum — an imperative that has no counterpart in the suzipu and gongche materials, where the pitch is written out. Each notation is stored both as a facsimile image, linked to the calligraphy layer when the score

survives in the poet's hand, and as a structured encoding in MEI, which preserves pitch where pitch is determined, duration where duration is notated, and the alignment between text and notation where an alignment is warranted. The choice of MEI over MusicXML is motivated by MEI's support for editorial metadata and variant readings — features that align with the framework's commitment to provisionality and polysemy preservation. Where a source carries no metre, MEI's critical apparatus allows the absence to be recorded as an absence rather than filled.

Layer 2 aligns to Layer 1 through a relation whose unit on each side is declared explicitly, since the units are not commensurable across notations. Four alignment types are defined, and only the first two belong to the score layer. Text-to-note alignment pairs a character of the lyric with a notated pitch in a pitch notation such as *suzipu* or *gongche*; the unit on the text side is the syllable, the unit on the notation side is the pitch, and the relation is one-to-one or, where a character is set to several notes, one-to-many. Text-to-action alignment pairs a character with a *jianzipu* sign; here the notation side is neither a syllable nor a pitch but a compound of plucking technique, string, and position, so the relation is character-to-sign, and the derivation of pitch from the sign requires the tuning as a precondition. Iconic alignment pairs an image (意象) activated by the lyric — the 扬州 of Jiang Kui's *Yangzhou Man*, for instance — with a region of a paired *tihua* (题画) painting or of a surviving image in a *huaci* (画册); this is a relation between the text layer and the calligraphy layer. Spatial alignment pairs a line of the lyric with a line of a manuscript, a rubbing, or a printed page; this too is a calligraphy-layer relation (Owen 1992). Treating them together makes it difficult to see what is being aligned at each layer; they are kept apart here so that each layer's alignments can be validated on their own terms. Each relation additionally declares its own cardinality and its own behaviour under partial data. Where a mapping is one-to-many the targets are stored as an ordered set rather than collapsed to a single value. Where the notation supplies no value for a position — as *jianzipu* supplies no metre — the alignment is stored with the status *unrecorded* and a null target, never with an inferred default. Where a sign is illegible or the witness is damaged, the alignment is stored with the status *lacuna*, and the extent of the lacuna is recorded so that a later examiner can see how much of the line is affected. Where the empty side of a mapping is itself a property of the source — as in the harmonic section (泛音) of *Gu Yuan*, where the tablature runs without lyric — the record is kept in the register of structural absences carried by the worked entry of Section 3.1 rather than in the alignment list, so that a deliberate blank is not read as a gap in the evidence.

Layer 3: Calligraphy Layer. The calligraphy layer stores the visual carriers of the poem. For each carrier, the layer records (i) the character-level data — *bi-hua* (笔画) analysis, the calligrapher's hand, and the script style (*xingshu*, *kaishu*, *caoshu*); (ii) the layout-level data — column ordering, title and *qianzi* (题款) placement, seals, and *zhuangbiao* (装裱); and (iii) the material-level data — substrate (绢本, 纸本, 拓本), ink color, dimensions, and any colophonic data. The three sub-layers are deliberately nested: character-level analysis is uninformative without the layout context (a single character out of context can be confused for a Tang piecework or a Song fake), and the material layer is uninformative without its inscription-pedigree (a *huangma zhi* (黄麻纸) of the right date can mislead if one does not know which workshop produced it). The framework therefore insists that any calligraphy-layer entry be released as a bundle, not as a roster of component features; this is the analog, at the visual layer, of the textual layer's editorial-chain requirement. The calligraphy layer is, in one

sense, the most straightforwardly visual of the three layers; in another sense, it is the most semantically dense. A single calligraphic carrier may carry information about the poet's handwriting (if it is an autograph), about the calligraphic tradition (if it is a later transcription by a recognized master), about the material culture of textual transmission (if it is an engraved stele or a woodblock print), and about the reception history (if marginalia by later collectors are present). The layer records all four dimensions: the bi-hua (stroke-level) analysis is stored as vector data extracted by computer vision; the calligrapher's hand is identified by a trained classifier and verified by a human expert; the material class is recorded as a controlled vocabulary (ink trace, engraved transcription, rubbing, woodblock print, modern facsimile); and the collector seals and colophons are transcribed and linked to the CBDB person records of their owners. The layer thus transforms the calligraphic carrier from an art-historical illustration into a structured philological datum.

4.1 AI-Assisted Annotation across Three Channels

The textual, score, and calligraphy layers are each subject to AI-assisted annotation, but the techniques differ in maturity and risk. For the textual layer, SikuBERT-class pre-trained models and contemporary Chinese-language large language models — applied through the Wuyudian-style smart-annotation interface — have demonstrated effective performance on automatic punctuation, classical Chinese named-entity recognition, and relation extraction tasks (Wuyudian 2023; J. Wang et al. 2024). For the calligraphy layer, attention-based handwritten text recognition (HTR) pipelines identify the calligrapher's hand and segment the strokes; the most actively developed infrastructure is eScriptorium and Transkribus-derived pipelines (Zhao et al. 2025; eScriptorium 2019). For the score layer the picture divides by task, and the tasks differ sharply. Character recognition on historical Chinese notations has advanced quickly: on the Baishidaoren Gequ corpus, recent work reports a character error rate of 7.1 per cent for suzipu across seventy-seven highly imbalanced classes and 0.9 per cent for lülüpu, against an average human error rate of 15.9 per cent on the same material, with the dataset extended to all 109 pieces of the collection across five historical editions and covering jianzipu as well as the two pitch notations (Repolusk and Veas 2025). Comparable work exists for gongche (He et al. 2025). What character recognition does not by itself settle are the three tasks downstream of it: deriving pitch from jianzipu, which requires the tuning to be settled; recovering metre, which jianzipu does not record and which must be supplied by editorial decision (Yingshi Chen 1986); and aligning the readings of one edition against another, where the transcriptions diverge, as they do for Gu Yuan (Yang et al. 2022). The framework has been run across all three channels over the corpus described in Section 6, and it treats each channel as a candidate-proposing layer rather than an authoritative one: every machine-proposed annotation enters the dataset with a model identifier, a confidence score, and an evidence pointer, and is subject to the review protocol of Section 4.2. The three channels are not equally demanding — textual annotation inherits a decade of pre-trained language-model development, calligraphic annotation draws on the broader handwriting-recognition ecosystem, and score annotation asks most of its models, since pitch and metre must be derived from the tablature and the tuning together rather than read off the page.

4.2 Human-in-the-Loop at Three Levels

The framework insists that AI-assisted annotation is provisional, and it separates two things that are easily run together: the review status of an annotation, which is a state it occupies, and the review operations that move it between states, which are events. There are three statuses.

Candidate: the annotation was proposed by a model and carries a model identifier, a confidence score, and an evidence pointer. Reviewed: a philologist has examined the candidate and recorded a verdict, and the original candidate remains in the provenance metadata alongside it. Endorsed: for high-stakes fields — variant readings, dates, attributions — the annotation has additionally been offered to a community forum, an extension of the *xueshu* (学术) channel that has long existed in Chinese philology. The *why* field is reserved for endorsed annotation, since it almost invariably crosses into interpretation. Against these statuses stand two operations. Conflict escalation occurs when two models, or a model and a legacy database such as CBDB or the Chronology Map, produce contradictory annotations for the same field; the conflict is escalated to the philologist rather than resolved algorithmically, and the framework’s function is to surface the disagreement, not to settle it. Rejection occurs when a philologist declines to carry a candidate forward; it is logged with a reason code, and the annotation is retained rather than deleted. The reason codes are deliberately not equivalent to falsity, and this distinction is load-bearing for the training policy. Three codes are recognised: erroneous, meaning the candidate is factually wrong; unverifiable, meaning the evidence currently available is insufficient to publish it; and superseded, meaning a later candidate displaces it. Of these, only erroneous material may be used as a negative training example. Unverifiable annotations are withheld from training altogether, because their being wrong has not been established, and using them would teach the model to reject readings that may simply be under-evidenced. Superseded annotations are retained as positive historical data with their dates intact, because a superseded reading remains the reading of a datable witness and is often the very thing a scholar needs to see. The statuses are not sequential gates: an annotation may be proposed, escalated, reviewed, and rejected within a single editorial session, and the framework logs the complete decision chain, from model output through philologist verdict, as a provenance trail that is itself auditable.

A schematic representation of the alignment of the three layers along the Six-W dimensions, with the AI-assisted three-channel pipeline and the human-in-the-loop loop, is provided in Figure 1.

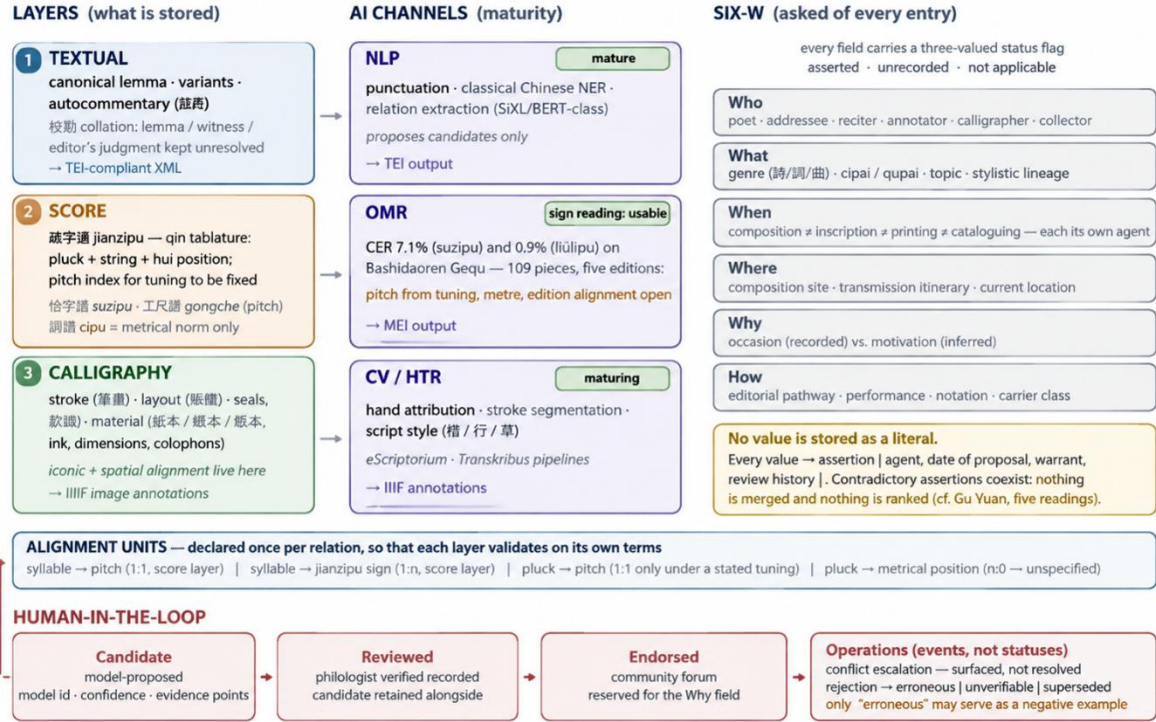


Figure 1 The three-layer multimodal pipeline, the AI channels that propose annotations and their present maturity, the Six-W dimensions with their three-valued status flag, the alignment units declared for each relation, and the human-in-the-loop statuses and operations. Statuses describe where an annotation stands; operations describe what is done to it. Dashed line: reviewed annotations return to the layers with their review status attached.

4.3 The Implementation and Its Limits

The pipeline has been implemented and run across all three layers over the corpus described in Section 6; its data types, techniques, and output formats are summarized in Table 3. What the implementation does not remove is the unevenness of the source material. The three layers are not equally informative: the textual layer rests on a continuous written record, the calligraphic layer on a body of surviving carriers of very unequal quality, and the score layer on notations that in the qin repertory record instrumental action rather than pitch, so that pitch must be derived from the tablature and the tuning together and metre is frequently not recorded at all. The framework's answer is not to withhold such an annotation but to grade it. The three-valued status flag of Section 3 allows a field to be published as asserted, unrecorded, or not applicable, and the review protocol of Section 4.2 allows an assertion to be published with its certification state attached; a reader can therefore distinguish what the evidence supports from what the pipeline could not settle, without either overstating the one or suppressing the other.

Table 3 Three Layer Multimodal Annotation Pipeline: Data Types, AI Techniques, Implementation Status, and Output Formats

Layer	Data Type	AI Technique	Implement ation status	Human-in-the-Loop status	Output Format
Textual	Canonical text,	SikuBERT NLP; NER; relation	Implement	Levels 1–3	TEI-compliant

	variants, commentaries	extraction	ed		XML
Score	Jianzipu tablature (qin); suzipu and lülüpu (pitch notations)	OMR character recognition; pitch derivation from tablature and tuning; edition alignment	Implement ed	Levels 1–2	MEI-compliant notation, with critical apparatus
Calligraphy	Ink tracings, rubbings, engravings	Computer vision; HTR; stroke analysis	Implement ed	Levels 1–3	IIIF-compliant image annotations

Table 3 sets out what each layer requires of its models. The asymmetry in difficulty it registers is a structural feature of the material rather than a defect to be hidden; the epistemic principles articulated in the next section are designed precisely to ensure that it does not compromise the integrity of the dataset.

5. EPISTEMIC PRINCIPLES

The framework’s epistemic posture is shaped by four principles. Each is a commitment the framework makes about how the implementation behaves and how any further implementation must behave; none can be relaxed without compromising the framework’s coherence as a digital-humanities contribution.

Provisionality. Every annotation is, in the framework’s terms, a hypothesis with provenance — a claim that has either survived or not yet survived the philologist’s review. No annotation, AI-suggested or human-confirmed, is treated as definitive, which is the position Sahle takes when he treats a scholarly edition as a critical representation that is never closed (Sahle 2016). The model for this posture is PROV-O, the W3C Recommendation that provides a vocabulary for describing and exchanging provenance information among entities, activities, and agents (W3C 2013). PROV-O itself requires nothing of the data published on the web: it defines terms, it does not legislate. The requirement that every datum carry provenance is therefore adopted here as a project-internal editorial policy, expressed in PROV-O’s terms, and not as an obligation the Recommendation itself imposes. The provisionality principle also has a temporal edge: every annotation is time-stamped, and the framework records when, by whom, and against which version of which source the annotation was confirmed. An annotation that holds for ten years is more durable than one that holds for ten minutes, and the framework records durability as a function of community review, not of individual confidence. The practical consequence for release is a confidence map: every public release carries a per-field annotation indicating whether the entry has been endorsed, is under review, or is still a machine candidate awaiting first-pass validation. A downstream user who queries the When field of a ci and receives “1176 (candidate, model-suggested, pending review)” knows that the date is not yet philologically certified. This transparency is, in the framework’s view, a minimum requirement for scholarly credibility: a dataset that conceals its own uncertainty is worse than no dataset at all.

Polysemy preservation. A single text may have multiple historically legitimate readings; the framework never forces one reading to absorb the others. Each variant, each reading tradition, each xue’an (学案) annotation is preserved as a separate expression in the FRBR sense (IFLA 1998). The choice is deliberate rather than conventional. FRBR distinguishes the work from its expressions, and a variant reading, a commentator’s emendation, and a modern editor’s

collation are all distinct realisations of the same work, produced by identifiable agents on identifiable warrants; treating them as separate expressions preserves their mutual independence, whereas treating them as alternative values of a single property would make them look like competing answers to one question. A statement about the text is a different kind of entity again: it is not modelled as an expression at all but as the assertion entity described in Section 3, so that the history of the text and the history of claims about the text remain separable. The tradition supplies the material for this. The editions of the Baishidaoren Gequ yield slightly different readings of Gu Yuan, and the five modern transcriptions of its tablature yield different pitches; Zhang Yan’s Ciyuan (《词源》), a foundational early discussion of ci music and prosody, already laments that the old music was little understood in its own day (Ciyuan 1297). The framework holds these apart without ranking them. The polysemy principle distinguishes it from mainstream NLP practice, in which the goal is typically to select the single best reading. In philology, by contrast, a variant reading is not an error to be corrected but a historical datum to be preserved. The framework implements this by allowing multiple assertions for a single Six-W field, each carrying its own provenance: the When field of a ci might simultaneously carry a composition date proposed in a Song preface and the same date re-asserted by a twentieth-century editor on a different warrant, or, where the sources genuinely disagree, two incompatible dates with neither superseding the other. The dataset thus becomes a record of disagreement rather than a forced consensus.

Tradition-aware mapping. The framework declines to overwrite Chinese philological vocabulary with CIDOC-CRM or PROV-O entries. The terms cipu, jianzipu, suzipu, cipai, qupai, gongdiao, zengyin, dunju, biaodiao and others are the legitimate analytical vocabulary of the field; the framework offers bidirectional mappings, not substitutions. The jianzipu case makes the point concretely. There is no Western musical term that corresponds to a compound sign specifying a plucking technique, a string, and a position between two hui; the nearest English description — “tablature” — names the family and not the thing. To map jianzipu onto a generic tablature class and then read the class as a definition would discard precisely the information the framework exists to preserve. The tradition-aware principle is not merely a courtesy to Chinese philology; it is an epistemic necessity. The term cipai (词牌) is not a “genre” in the FRBR sense, nor is gongdiao (宫调) a “key” in the Western musical sense. Each Chinese term carries a network of associations — historical, musicological, prosodic — that a one-to-one mapping to a Western ontology would silently truncate. The framework preserves the Chinese term as the primary label and records the ontology cross-reference as a secondary annotation, reversing the usual direction of mapping in which the local term is reduced to an instance of the universal one.

Annotator as interpreter and layered openness. AI does not replace the philologist; it supplies a pre-fabricated datum — the kind of datum that a philologist would otherwise have to assemble by hand (Yao et al. 2025; Zhang et al. 2024). Each of the three layers, together with the schema and the interface specification, has been implemented, and each can be released independently, in keeping with the design principles of the IIF ecosystem (IIF 2012). The annotator-as-interpreter principle has a corollary for the dataset’s licensing model. The framework adopts a layered openness policy: the Six-W schema, the ontology definitions, and the API specifications are published under an open license (CC-BY); the annotation data itself is released incrementally, with each batch carrying a confidence map and a review-status stamp. The framework rejects the binary choice between “open data” and “certified data”: it releases data that is open but labeled with its own uncertainty, allowing downstream users to make

informed decisions about which entries to trust for which purposes.

6. IMPLICATIONS AND A RESEARCH AGENDA

For digital humanities. The framework extends the annotation turn in DH into a domain — Chinese poetry — that has been more often indexed than modeled. The resources described in Section 2 were designed for other tasks; binding a lyric to its *cipai*, to its surviving notation, and to a corroborated occasion within a single schema is the gap this dataset was built to occupy, and so far as we have been able to establish no such binding had been attempted. The framework also offers a test for the often-expressed claim that DH is a generalisable methodology: by stating the schema and pipeline explicitly, it invites implementation in other non-Western traditions (Japanese *waka*, Persian *ghazal*, the Arabic ode, Sanskrit *subhasita*) where the questions of who, what, when, where, why, and how are equally load-bearing. The framework also contributes to the ongoing debate within digital humanities about the relationship between distant reading and close reading. A multimodal dataset does not eliminate close reading; it enriches it. The philologist who consults the framework’s entry for Yangzhou Man sees not only the ninety-eight characters but the notation, the autocommentary, the edition history, and the commentarial chain — a thicker close reading than any printed edition can offer. At the same time, the structured nature of the dataset enables distant reading operations that are impossible with conventional editions: one can query all *ci* composed within a decade of a given military event, or all *ci* inscribed by a given calligrapher, or all *ci* whose *cipai* belongs to a given *gongdiao*. The framework thus mediates between the two reading modes rather than choosing between them.

For multimodal AI. The framework supplies richly-provenanced, low-noise, medium-volume training data of the kind that contemporary multimodal AI badly needs (Manovich 2020; Schuhmann et al. 2022). Most multimodal AI training corpora are built from web-scale text–image pairs with negligible provenance (Schuhmann et al. 2022). A corpus in which each annotation carries a controlled PROV-O entry is rare; the dataset described here is one. The dataset is also deliberately scoped for the model rather than scoped for the philologist: every annotation ships with the metadata that a multimodal model would need to learn a justified reading rather than a memorized caption. The dataset’s value for AI training lies not only in its volume but in its structure. Each entry carries a provenance trail that links the text, score, and calligraphy to a verified spatio-temporal context — the kind of metadata that web-scraped image-text pairs lack entirely. Whether such structure translates into capability is an open question, and the framework states it as three research questions rather than as a claim. First, does provenance-bearing training data of this kind measurably improve a model’s handling of culturally specific relations — for instance, its ability to distinguish a text-to-note alignment from a text-to-action alignment — relative to unstructured web corpora of comparable size? Second, does the presence of explicit, dated, and conflicting assertions during training change a model’s behaviour with respect to uncertainty, teaching it to report an unspecified value as unspecified rather than to fill it? Third, can a model trained on one notated tradition transfer to a cognate one, such as the Japanese or Korean repertoires discussed below, without retraining on the target notation? None of these questions is answered here. The framework’s claim is narrower: that such questions can be asked only of corpora that record provenance in the first place, and that this is why the data was built before the model.

For classical Chinese scholarship. The framework sustains philological practice at scale. By assigning provenance to every Six-W entry, the framework enforces the fundamentum of Chinese textual scholarship — the rule that every claim must be tied to an authority — as a default. The framework also recovers, in technical form, the very thing that made the cipai tradition intellectually productive: the alignment of lyrics to melodies and to material carriers, in a way that recovers not only how a cipai was sung but also how a ci was seen, copied, displayed, and republished. For specific cases (Jiang Kui’s Yangzhou Man in the Baishi Daoren Gequ, or the transmission of that same collection through the Jiang Bingyan manuscript of 1737, in which one hand carries both the words and the tablature), the framework offers a way to move between textual, musical, and visual evidence without losing the philological thread (Jiang 1202; Bush 1971). The framework also addresses a structural problem in the institutional organization of Chinese literary studies: the separation of ci philology (词学) from musicology (音乐学) and from calligraphy studies (书法研究). In the academic division of labor, a ci scholar edits the text, a musicologist reconstructs the score, and an art historian studies the calligraphy; the three rarely consult one another’s evidence. The framework’s multimodal schema forces these strands back into a single analytical frame, not by dissolving the disciplinary boundaries but by providing a shared data infrastructure in which each discipline’s contribution is visible to the others.

For comparative East Asian literary studies. The framework’s Six-W schema is, in principle, portable to cognate traditions beyond the Chinese ci and qu. The Japanese waka, the Korean sijo, and the Vietnamese lục bát all share with the Chinese lyric a coupling of text to musical realization and to calligraphic inscription; the provenance fields — who, what, when, where, why, how — transfer with minimal adaptation. A pilot comparison between a Song ci and a Heian waka, both annotated under the Six-W schema, would expose structural homologies and differences that text-only comparison cannot reach: the waka’s kana calligraphy and the ci’s cipai score are not parallel data in a text-first framework, but they become comparable once the framework treats each as a How-field entry under a shared ontology. The Chronology Map’s extension to the Korean and Japanese traditions is a medium-term objective of the research program.

The Chronology Map, in its current Tang–Song form, already provides verified spatio-temporal data for a substantial corpus of poems and their participants. The post-Tang extension — covering the ci, qu, and sanqu traditions from the late Tang through the Qing — carries the Six-W provenance fields, the three annotation layers, and the pipeline described in Section 4, over a corpus of approximately 1.3 million compositions, with the full three-layer treatment applied, as Section 1 explained, wherever at least one non-textual witness survives. The dataset is designed for interoperability: its ontology is expressed in PROV-O and CIDOC-CRM, its geographic data in GeoJSON linked to CHGIS, and its textual data in TEI-compliant XML. Release proceeds in two streams, and the distinction between them is what keeps a corpus of this size from overwhelming the review protocol of Section 4.2. The candidate stream contains machine-proposed annotations that have not been reviewed; it is published as it is generated, each item carrying its confidence score and its evidence pointer. The reviewed stream contains only annotations that a philologist has signed off; it is published batch by batch as review is completed. The two streams share identifiers, so that an item can be traced from the first to the second. What is never published is a mixture of the two in which the distinction is invisible. Construction began with the Song ci corpus as a proof of concept and was then extended to the Yuan qu and the Ming–Qing sanqu and irregular ci repertoires. The Chronology Map functions

simultaneously as the framework's data spine and as its quality-control mechanism: the spatio-temporal record supplies the verified coordinates against which every new multimodal annotation is checked.

The research program. This framework is the anchor of a five-year research program situated within a national-level research program on the digital-humanities transmission of Chinese traditional symbols and within an international cooperative agenda on digital humanities. The program has moved from schema definition to data: the dataset now built anchors approximately 1.3 million post-Tang lyric compositions across the Song, Yuan, Ming, and Qing, each carrying a Six-W provenance record, and the next stages extend the comparative work to the Japanese waka, the Korean sijo, and the Vietnamese *lục bát* traditions, and to a sustained reflection on the epistemic implications of multimodal scholarship for the digital humanities more generally. The schema is set out here in full because it is the most enduring and contested component of the program; nothing precludes later revisions, but the schema as it stands is intended to be the first stable layer of a scholarly public.

7. CONCLUSION

The post-Tang lyric poem has long resisted the textual edition precisely because it is a composite event — a text that has been sung, transmitted through musical notation, inscribed on a calligraphic carrier, and woven into a commentarial tradition. A digital edition that returns only the characters is, in the most literal sense, a decontextualization. The present paper has proposed and implemented a methodological framework that couples the textual layer to the score layer, the calligraphy layer, and the Six-W provenance schema in a single analytical contract.

Three claims have been argued: that the schema of a multimodal resource is methodologically prior to its implementation, because the interface between text and score / calligraphy is, in the end, an ontological claim about the poem; that the Six-W schema inherits from and extends the existing annotation ontologies, by binding five established annotation standards to the philological vocabulary of Chinese *ci* studies rather than translating that vocabulary into them; and that the auditability of a large multimodal dataset depends on its schema rather than on its scale, since a corpus that does not record how its categories were assigned cannot support the inferences its size appears to license. The schema is published here in the spirit of Sahle's account of the scholarly digital edition as a critical representation that is never closed: a document that is deliberately provisional but deliberately public, inviting correction of the schema rather than its abandonment (Sahle 2016). The Chronology Map provides the empirical ground; the Six-W schema provides the conceptual architecture; the three-layer pipeline provides the operational specification; the epistemic principles provide the quality contract. What remains is the labor of review — patient, incremental, and irreducibly philological.

By stating the schema, the pipeline, and the epistemic principles in full, the framework makes the dataset they carry both inspectable and accountable. The work ahead — comparative volumes, joint editorial methods, and the extension of the schema to cognate traditions — remains substantial; the harder half, which this paper attempts, is to say, in some detail, what it is we have built and on what warrant each part of it rests.

ORCID

Yao Song  <https://orcid.org/0000-0001-6075-6211>

REFERENCES

- Baltrušaitis, Tadas, Chaitanya Ahuja, and Louis-Philippe Morency. 2019. “Multimodal Machine Learning: A Survey and Taxonomy.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (2): 423–43.
- Burnard, Lou. 2014. *What Is the Text Encoding Initiative?* OpenEdition.
- Bush, Susan. 1971. *The Chinese Literati on Painting*. Cambridge, MA: Harvard University Press.
- Chen, Yingshi [陈应时]. 1984. “Lun Jiang Baishi de ‘Ce shang diao tiaoxianfa’ [论姜白石的《侧商调调弦法》].” *Yinyuexue Congkan* [音乐学丛刊], no. 3.
- Chen, Yingshi [陈应时]. 1986. “Qinqu ‘Ce shang diao Gu Yuan’ pu kaobian [琴曲《侧商调·古怨》谱考辨].” *Yinyue Yanjiu* [音乐研究], no. 3.
- Chen, Yuan [陈垣]. 1931. *Yuan dianzhang jiaobu shili* [元典章校补释例], juan 6, “Xiaofa sili [校法四例].” Reprinted as *Jiaokanxue shili* [校勘学释例]. Beijing: Zhonghua Shuju, 1959.
- China Biographical Database (CBDB). 2008. <https://cbdb.fas.harvard.edu/>.
- China Historical GIS (CHGIS). 2001. <https://chgis.fas.harvard.edu/>.
- Chinese Text Project. 2006. <https://ctext.org/>.
- Ciyuan* [词源]. 1297. By Zhang Yan [张炎]. Composed ca. 1297–1317; consulted in the Xia Chengtao annotated edition, 1963.
- Coquenot, Denis, Clément Chatelain, and Thierry Paquet. 2023. “End-to-End Handwritten Paragraph Text Recognition Using a Vertical Attention Network.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45 (1): 508–24. <https://doi.org/10.1109/TPAMI.2022.3144899>.
- DCMI (Dublin Core Metadata Initiative). 1999. “Dublin Core Metadata Element Set, Version 1.1.” <https://www.dublincore.org/specifications/dublin-core/dces/>.
- Duan, Siyu, Jun Wang, Hao Yang, and Qi Su. 2023. “Disentangling the Cultural Evolution of Ancient China: A Digital Humanities Perspective.” *Humanities and Social Sciences Communications* 10: 310. <https://doi.org/10.1057/s41599-023-01811-x>.
- Elkins, James. 1999. *The Domain of Images*. Ithaca, NY: Cornell University Press.
- eScriptorium. 2019. <https://escriptorium.openhistorymaps.org/>.
- Gong, Yi [龚一], and Guohua Xu [许国华]. 1979. “Jiang Baishi Gu Yuan zhi pouxi [姜白石《古怨》之剖析].” *Renmin Yinyue* [人民音乐], nos. 11–12.
- He, Zhizhou, Yuqian Zhang, Liumei Zhang, and Yuanjiao Hu. 2025. “Precise Recognition of Gong-Che Score Characters Based on Deep Learning: Joint Optimization of YOLOv8m and SimAM/MSCAM.” *Electronics* 14 (14): 2802. <https://doi.org/10.3390/electronics14142802>.
- ICOM/CIDOC. 2021. *CIDOC Conceptual Reference Model*. Version 7.1.
- Ide, Nancy, and C. M. Sperberg-McQueen. 1995. “The TEI: History, Goals, and Future.” *Computers and the Humanities* 29 (1): 5–15. <https://doi.org/10.1007/BF01830313>.

- Ide, Nancy, and Laurent Romary. 2002. "Standards for Language Resources." In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*. Las Palmas, Canary Islands: European Language Resources Association.
- IFLA. 1998. *Functional Requirements for Bibliographic Records (FRBR)*.
- International Image Interoperability Framework (IIIF). 2012. <https://iiif.io/>.
- Jiang, Kui. 1202. *Baishi Daoren Gequ* [白石道人歌曲]. The editions of 1202 are lost; consulted in the Lu Zhonghui edition (1743), the Zhang Yishu edition (1749), the *Siku Quanshu* copy, and the Zhu Xiaozang edition of 1913 printed from Jiang Bingyan's manuscript of 1737.
- Jiu Gong Da Cheng Nan Bei Ci Gong Pu* [九宫大成南北词宫谱]. 1746.
- Jockers, Matthew L. 2014. *Text Analysis with R for Students of Literature*. Cham: Springer.
- Lam, Joseph S. C. 1998. *State Sacrifices and Music in Ming China: Orthodoxy, Creativity, and Expressiveness*. Albany: State University of New York Press.
- Lam, Joseph S. C. 2012. *Historical Studies on Song Dynasty Music: Theories and Narratives* [宋代音乐史论文集: 理论与描述]. Shanghai: Shanghai Conservatory of Music Press.
- Long, Yusheng [龙榆生]. 1978. *Tang Song ci geli* [唐宋词格律]. Shanghai: Shanghai Guji Chubanshe.
- Ma, Xingrong, Xionghe Wu, and Jiping Cao, eds. 1996. *Zhongguo Cixue Dacidian* [中国词学大辞典]. Hangzhou: Zhejiang Education Publishing House.
- Manovich, Lev. 2020. *Cultural Analytics*. Cambridge, MA: MIT Press.
- Owen, Stephen. 1992. *Readings in Chinese Literary Thought*. Cambridge, MA: Harvard University Press.
- Owen, Stephen. 2006. *The Late Tang: Chinese Poetry of the Mid-Ninth Century (827–860)*. Cambridge, MA: Harvard University Asia Center.
- Pamart, Anthony, Livio De Luca, and Philippe Véron. 2022. "A Metadata Enriched System for the Documentation of Multi-modal Digital Imaging Surveys." *Studies in Digital Heritage* 6 (1): 1–24. <https://doi.org/10.14434/sdh.v6i1.33767>.
- Pian, Rulan Chao. 1967. *Song Dynasty Musical Sources and Their Interpretation*. Cambridge, MA: Harvard University Press.
- Picken, Laurence. 1957. "China." In *The New Oxford History of Music*, vol. 1, *Ancient and Oriental Music*, edited by Egon Wellesz. London: Oxford University Press.
- Qiu, Qionsun [邱琼荪]. 1956. *Baishidaoren Gequ tongkao* [白石道人歌曲通考]. Beijing: Yinyue Chubanshe.
- Quan Tang Shi* [全唐诗]. 1706.
- Ram, Sudha, and Jun Liu. 2009. "A New Perspective on Semantics of Data Provenance." *CEUR Workshop Proceedings* 526: 35–40.
- Repolusk, Tristan, and Eduardo Veas. 2025. "KuiSCIMA v2.0: Improved Baselines, Calibration, and Cross-Notation Generalization for Historical Chinese Music Notations in Jiang Kui's *Baishidaoren Gequ*." In *Document Analysis and Recognition – ICDAR 2025*, 116–32. Cham: Springer. https://doi.org/10.1007/978-3-032-04630-7_7.
- Sahle, Patrick. 2016. "What Is a Scholarly Digital Edition?" In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 19–40. Cambridge: Open Book Publishers.

- Schöch, Christof. 2013. “Big? Smart? Clean? Messy? Data in the Humanities.” *Journal of Digital Humanities* 2 (3).
- Schuhmann, Christoph, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, et al. 2022. “LAION-5B: An Open Large-Scale Dataset for Training Next Generation Image-Text Models.” *Advances in Neural Information Processing Systems* 35: 25278–94.
- Tang, Xuemei, Zekun Deng, Jun Wang, and Qi Su. 2025. “ChisNERE: A Premodern Chinese Corpus with Named Entity and Relation Annotation.” *Digital Scholarship in the Humanities* 40 (2): 617–38. <https://doi.org/10.1093/llc/fqaf001>.
- Transkribus. 2015. <https://transkribus.eu/>.
- W3C. 2013. *PROV-O: The PROV Ontology*. W3C Recommendation. <https://www.w3.org/TR/prov-o/>.
- Wan, Shu [万树]. 1687. *Cilü* [词律]. 20 juan.
- Wang, Dongbo, Chang Liu, Zihe Zhu, Jiangfeng Liu, Haotian Hu, Si Shen, and Bin Li. 2022. “SikuBERT 与 SikuRoBERTa: 面向数字人文的《四库全书》预训练模型构建及应用研究 [SikuBERT and SikuRoBERTa: Construction and Application of Pre-trained Models for Siku Quanshu for Digital Humanities].” *图书馆论坛* [Library Forum] 42 (6): 31–43.
- Wang, Jun, Siyu Duan, Binghao Fu, Liangcai Gao, and Qi Su. 2024. “Evol Project: A Comprehensive Online Platform for Quantitative Analysis of Ancient Literature.” *Humanities and Social Sciences Communications* 11: 291. <https://doi.org/10.1057/s41599-024-02763-6>.
- Wang, Linxu, Jun Wang, and Tong Wei. 2024. “Using Ontology to Model Time Description in Historical Chinese Texts.” *Digital Scholarship in the Humanities* 39 (1): 423–29. <https://doi.org/10.1093/llc/fqad092>.
- Wang, Linxu, Tong Wei, and Jun Wang. 2022. “Modelling Time Ontology in Ancient Chinese Texts.” In *Digital Humanities Conference 2022 (DH2022)*, Tokyo.
- Wang, Yiqing [王奕清], et al., comps. 1715. *Qinding cipu* [钦定词谱]. 40 juan.
- Wuyudian. 2023. <https://wyd.pkudh.org/>.
- Xia, Chengtao [夏承焘]. 1958. *Jiang Baishi ci biannian jianjiao* [姜白石词编年笺校]. Beijing: Zhonghua Shuju.
- Xia, Chengtao [夏承焘], ed. 1963. *Ciyuan zhu* [词源注]. By Zhang Yan [张炎]. Beijing: Renmin Wenxue Chubanshe.
- Yang, Bowen, Makoto Sato, Shun Kuremoto, Mamiko Koshiba, Shingo Mabu, Hiroo Hieda, and Takashi Kuremoto. 2022. “Recognition of Guqin Notation Using Deep Learning.” In *Proceedings of Innovative Application Research and Education (ICIARE2022)*, 61–64.
- Yang, Yinliu [杨荫浏], and Falu Yin [阴法鲁]. 1957. *Song Jiang Baishi chuangzuo gequ yanjiu* [宋姜白石创作歌曲研究]. Beijing: Yinyue Chubanshe.
- Yao, Xinyu, Mengdi Wang, Bo Chen, and Xiaobing Zhao. 2025. “WenyanGPT: A Large Language Model for Classical Chinese Tasks.” In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2025)*, 8339–47. <https://doi.org/10.24963/ijcai.2025/927>.
- Zha, Fuxi [查阜西], ed. 1962. *Guqin quji* [古琴曲集]. Vol. 1. Beijing: Yinyue Chubanshe.
- Zhang, Yuqing, Baoyi He, Yihan Chen, Hangqi Li, Han Yue, Shengyu Zhang, Huaiyong Dou, et al. 2024. “PhiloGPT: A Philology-Oriented Large Language Model for Ancient Chinese Manuscripts with Dunhuang as Case Study.” In *Proceedings of the 2024 Conference on Empirical Methods in*

- Natural Language Processing (EMNLP 2024)*, 2784–2801. Miami, FL: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.163>.
- Zhao, Yixin, Yuyi Zhang, and Lianwen Jin. 2025. “MCCD: A Multi-Attribute Chinese Calligraphy Character Dataset Annotated with Script Styles, Dynasties, and Calligraphers.” In *Document Analysis and Recognition – ICDAR 2025*, 538–55. Cham: Springer.
- Zhu, Quan [朱权]. 1425. *Shenqi mipu* [神奇秘谱].