

面向书画著录的文献循证与时空关联构建研究*

□高劲松 付家炜

摘要 书画著录是中国古代书画研究领域的基础性资料,对于传世书画作品的考据、赏析和鉴别具有重要参考价值。随着数字人文研究范式的兴起,在新技术条件下依托人文数据基础设施开展面向书画著录的文献循证实践,将为古代书画作品的时空分析与历史传承研究提供新的思路。文章在梳理相关研究现状的基础上,首先研究书画著录的时空语义描述与概念分层注释方法,提出面向书画著录的文献循证与时空关联构建过程,并以《石渠宝笈》书画著录为例开展案例实验。实验结果表明,文章提出的文献循证过程模型能够满足书画著录时空关联构建的需求,有助于为专业人士与社会公众提供多场景的书画著录知识服务。

关键词 数字人文 文献循证 时空关联 书画著录 知识组织

分类号 G203

DOI 10.16603/j.issn1002-1027.2022.05.004

1 引言

书画著录(Catalogue Literature of Calligraphy and Painting)是对中国古代书法、绘画作品进行目录性著述的专门文献,在书画研究中具有述流传、记传记、分派别、辨真伪等多种功能^[1]。在古代书画作品大量散佚、真伪相杂的今日,历代书画著录对于研究古代书画艺术特征和鉴别传世作品真伪具有重要参考价值。近年来,云计算、大数据、物联网等技术的发展催生了数字人文浪潮,为人文学科引入了新的思维模式和研究范式。数字人文的“大帐篷效应”促进了历史、文学、艺术等人文研究领域的交叉融合,书画著录研究作为涉及多个学科的交叉领域,在数字人文兴起的背景下有必要引入跨学科的研究视角,通过对人文研究方法与数字分析技术的有机整合,进一步开发书画著录潜在的隐性知识价值,从而更好地满足专业人士与社会公众对书画著录知识的利用需求。

文献循证是图情学科对数字人文研究的重要方法论贡献。循证研究即“基于证据的研究”,强调任何结论都需要从客观证据中得出,通过建立体系化

的原则、流程和方法以保障研究结论的客观性和科学性。文献循证是“基于文献的循证研究”,其实质是以文献材料中的客观事实和结论为证据,围绕研究问题形成一定的证据链,进而通过关系推导寻找可靠结论的过程^[2]。传统目录学、版本学、校勘学中的文献考据活动对于研究者的经验和能力具有较高要求^[3],而在数字技术的支持下,文献循证的证据来源不再局限于需要人工鉴别、翻阅的实体文献,还可扩展至文献资料中可被机器处理的各种事实知识,循证实践的应用场景大大扩展。在数字人文兴起的背景下,相关研究者依托人文数据基础设施,对面向文献事实知识的证据链构建和循证分析实践进行了探索^[4]。例如在南海历史文献研究中探索文献循证方法的数字化应用^[5-6];在分析古籍文献循证需求基础上提出基于文本可视化的古籍循证流程框架^[7];在历史人物研究中应用文献循证思想,通过量化分析和关联挖掘构建人物关系网络图谱^[8];在古籍资源建模研究中围绕文献循证的具体需求,构建面向异构资源融合的中文古籍数据模型^[9]。

时空分析是利用地理编码方法将原始信息在一

* 教育部人文社会科学研究规划基金项目“计算人文视角下文化遗产数字资源多模态知识组织与价值共创研究”(编号:22YJA870008)的研究成果之一。

通讯作者:高劲松,ORCID:0000-0003-0022-5923,邮箱:jsgao@ccnu.edu.cn。



定地理空间中进行时空关联和可视化表达的研究方法^[10]。时空分析将多种地理信息技术引入人文社会科学领域,运用定量、比较、计量等方法研究历史、社会、自然等多要素关联,现已成为数字人文的基本研究范式之一^[11]。随着时空分析在数字人文中应用的深入,如何实现时空关联构建的范式化、流程化,成为相关领域研究者需要关注的问题。时空关联构建是利用语义本体、自然语言处理、知识图谱等技术在同一时空基准上对不同尺度的地理数据进行相互关联,将不同的地理要素构成一个整体以实现更全面的关系表达的过程^[12-14],其在数字人文研究中的价值在于能够厘清各种时间、空间表述的定义、格式及关系,揭示各类要素及其属性在时空变化中所呈现的规律^[15-16]。目前,人文研究领域已在时空关联构建方面进行了一定探索,例如通过对文化记忆载体时空属性和关联的形式化编码以实现面向文化记忆领域的时空数据建模^[17];通过整合社会网络分析、文本结构分析、古地名提取等方法以满足古地图知识组织研究中的时空关联分析需求^[18];通过定义包含属性、事件、过程、状态的非物质文化遗产(以下简称非遗)时空关联描述模型,以开展非遗时空演化的量化分析研究^[19]。

伴随着数字人文研究范式的日益成熟,相关研究呈现层次更深、领域更专、粒度更细的发展趋势,不仅在供给侧对人文数据基础设施的建设水平提出了更高要求,也在需求侧呼唤新技术条件下人文研

究方法的迁移迭代。在时空分析成为数字人文基本研究范式之一的背景下,将文献循证方法应用于书画著录的时空关联构建研究,能够更好地满足相关领域对书画作品历史传承知识的需求。

2 书画著录的时空语义描述与概念分层注释

2.1 书画著录时空语义描述模型

书画著录真实地反映了历代书画作品的面貌,对揭示作品的历史传承具有重要价值。但是由于书画著录结构、内容的特殊性,将现有的古籍元数据标准应用于其时空关联构建时,存在时空属性揭示不足、领域描述需求难以适配等问题。因此在开展书画著录文献循证时,须针对其结构和内容特征构建专门的时空语义描述模型。本文以图1所示《江帆山市图》著录为例,对书画著录的结构和内容进行分析。在文献结构上,书画著录涉及后世记录的作品题名、创作年代、保管地点等外部要素,装帧、用纸、技法、尺寸等物理要素,以及题跋、印鉴等反映作品历史传承过程的时空要素。在文献内容上,书画著录可以理解为著录者基于自身认知和著录体例,对书画作品内容、形制所进行的半结构化描述,这一过程与现代图像学中对视觉艺术作品进行内、外部诠释的图像志研究具有相似性^[20-21]。换言之,书画著录中的著录条目亦可被理解为一种“文本态图像”,因此在书画著录时空语义描述模型构建中,有必要引入面向人文艺术图像的语义描述方法。

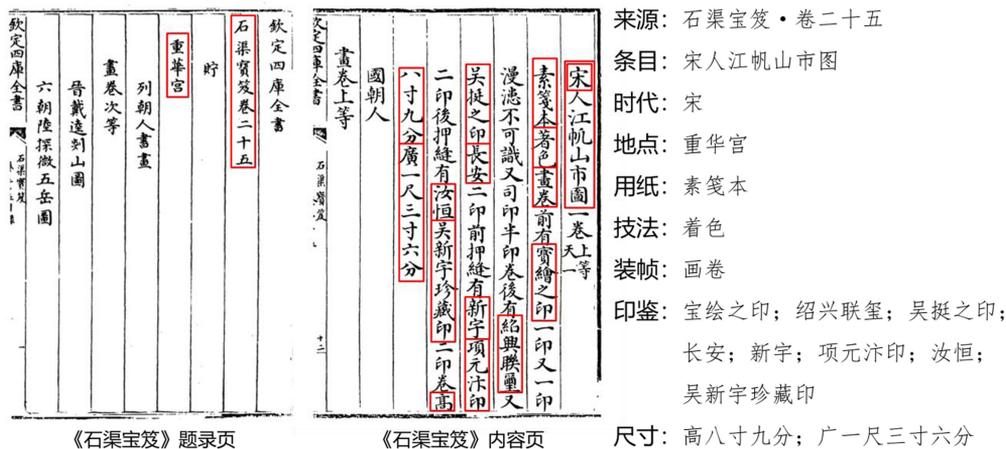


图1 书画著录的结构与内容示例



目前,将视觉艺术领域的图像学理论应用于人文图像语义描述的研究已有先例,例如王晓光(Wang X)等参考欧文·潘诺夫斯基(Panofsky E)的图像学理论构建了面向敦煌壁画数字图像的深度语义描述框架^[22-23];曾子明等针对数字人文领域的用户认知需求,在借鉴萨拉·沙特福德(Shatford S)的图像分层描述理论基础上提出了面向历史照片的语义描述模型^[24-25];朱学芳等基于对非遗图像用户的认知层次分析,应用 Panofsky-Shatford 模型构建了面向非遗领域的数字图像描述框架^[26]。

Panofsky-Shatford 模型是图像学领域的重要理论, Panofsky 模型主要通过前图像志 (Pre-iconography)、图像志 (Iconography)、图像学 (Iconology) 的三层模型实现图像内容描述, Panofsky-Shatford 模型在前者基础上进行扩展,将图像内容划分为通用概念、具体概念和抽象概念 3 个层次,每个层次又与人物、事物、时间、地点 4 个维度分别对应。本文在 Panofsky-Shatford 模型基础上结合书画著录的结构与内容特征,提出如表 1 所示的时空语义描述模型。

表 1 书画著录时空语义描述模型

维度	通用概念 General	具体概念 Specific	抽象概念 Abstract
人物/机构 Who	G1:一般性的人物身份或机构类型(如:创作者、题跋者、钤印者、收藏机构等)	S1:具体的人物或机构(如:董其昌、项元汴、故宫博物院等)	
事物 What	G2:概括性的事物总称(如:著录条目、鉴藏印名等)	S2:具体的事物名称(如:宋人江帆山市图、项元汴印等)	A:反映书画作品相关性质的各类名词概念 (如:题材、用纸、技法、装帧、尺寸等)
时间 When	G3:某一类时间表述的总称(如:朝代纪年、年号纪年、公元纪年、干支纪年等)	S3:具体的时间表述(如:明代、崇祯七年、公元 1636 年、丙子年等)	
空间 Where	G4:某一类空间表述的总称(如:历史地名、地理事物、建筑场所等)	S4:具体的空间表述(如:松江、洞庭湖、重华宫等)	

表 1 的模型包含 3 个概念层次和 4 个语义维度,其中 Who 维度、What 维度分别用于表示书画著录中的各类人物、事物,When 维度用于描述书画著录中各种时间表述,并在通用时间概念基础上扩展了朝代、年号、干支等中国古代特有纪年方式,Where 维度用于描述书画著录中各类空间表述。模型中,G 类字段表示通用或泛指性概念,S 类字段表示具体或专指性概念,A 类字段用于表示书画作品的题材、用纸、技法、装帧、尺寸等抽象概念信息。

2.2 书画著录概念分层注释模型

古代书画作品是承载文化记忆的媒介资料,其衍生的图文声像数据则是以数字形式重组和传播的文化记忆资源^[27]。媒介资料作为文化记忆的载体,具有整体性、延续性和系统性的普遍特征。在书画领域,媒介资料的整体性体现在书画作品自身蕴含的直接信息、书画著录记载的间接信息以及古今书画赏评鉴定活动所产生的他者信息之间存在对照、引述、印证等多维关联;其连续性体现在书画作品在

内容表达、作品创作、题跋钤印、修复装裱的过程中形成了“三度时空”的层次分野^[28];其系统性则体现在书画作品及其相关文献著述所承载的主、客观知识间存在广泛的事理逻辑关联,且这种内容载体关联在历史文化变迁中不断固化和丰富。

基于上述分析,本文沿用 Panofsky-Shatford 模型的概念分层描述思想,引入媒介关联性和时空连续性作为分层依据,面向书画著录的内容与载体关联,定义书画内容(Content)、书画创作(Produce)、书画流转(Transfer)三个概念层次,提出如表 2 所示的二维分层注释模型。模型遵循表 1 的语义划分维度,定义人物、事物、时间、空间 4 种概念类型,在此基础上对书画著录的概念层次进行注释,以实现同一类型下不同文本的分层处理。例如《江帆山市图》著录中,将“素笺本”注释为“Thing.Produce”,将“项元汴印”注释为“Thing.Transfer”,由此可在时空层次上对两个事物概念(Thing)进行区分。



表2 书画著录概念分层注释模型

概念层次	人物概念(Person)	事物概念(Thing)	时间概念(Time)	空间概念(Place)
内容	反映书画内容中各类人物姓名、角色、身份的著录文本(如:文人、青年、老者等)	反映书画内容主题或内容中事物名称的著录文本(如:宋人江帆山市图、帆船、货物等)	直接或间接反映书画内容中时间概念的著录文本(如:黄昏、暮春、三月等)	直接或间接反映书画内容中空间概念的著录文本(如:山峰、江水、寺院等)
载体	书画创作(Produce)	反映书画作品题材、形制等信息的著录文本(如:山水画、素笺本、画卷)	反映书画作品创作时间或时代的著录文本(如:宋、公元1100年)	反映书画作品创作地点或作者籍贯的著录文本(如:吴兴、汴梁)
	书画流转(Transfer)	反映书画传承过程中鉴藏者、题跋者姓名、身份的著录文本(如:吴新宇、项元汴)	反映书画传承过程中附加的题跋、钤印等信息的著录文本(如:吴挺之印、项元汴印)	反映书画传承过程中地点或地名节点的著录文本(如:嘉兴、中华宫、故宫博物院)

3 面向书画著录的文献循证与时空关联构建过程

事实知识是文献循证的基础,文献循证实践需要对大量事实知识进行结构化存储,并通过推理和一致性检验以构建“事实证据链”。在新技术条件下,面向书画著录的文献循证与时空关联构建可归纳为如图2所示的过程模型:首先,利用数字化文本标注工具从书画著录中抽取实体并通过本体建模实现其关系组织;其次,引入历史人物传记、历史纪年表、历史地名表等可以互相佐证的数据源(以下简称互证数据源)对标注实体进行语义对齐和一致性检验,并通过多轮实体匹配过程揭示标注实体的时空语义关联;再次,利用语义知识库进行书画著录时空关联证据链的持久化保存,实现文献循证数据存储和文献循证过程揭示。

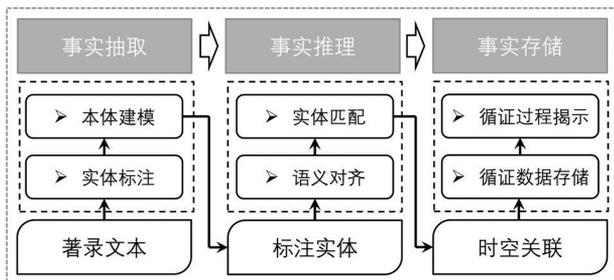


图2 面向书画著录的文献循证与时空关联构建过程

3.1 基于文本标注的书画著录事实抽取

书画著录事实抽取是对著录文本中的语义实体进行识别、标记并生成结构化数据集的过程。传统文献循证实践中主要通过“细读”实现文献事实抽取,即运用不同字形、字体、颜色、形状的标记对纸质文献中的人物、事物、时间、地点等实体进行标注^[29]。在新技术条件下,面向数字人文的文本标注

工具在实现数字化细读的同时可兼顾标注数据的管理组织,降低了文本标注的专业技术门槛,提升了相关方法在人文研究中的适用性^[30-31]。本文在引入数字化文本标注工具的基础上,提出面向书画著录的事实抽取流程,包含实体标注和本体建模两个关键环节,如图3所示。

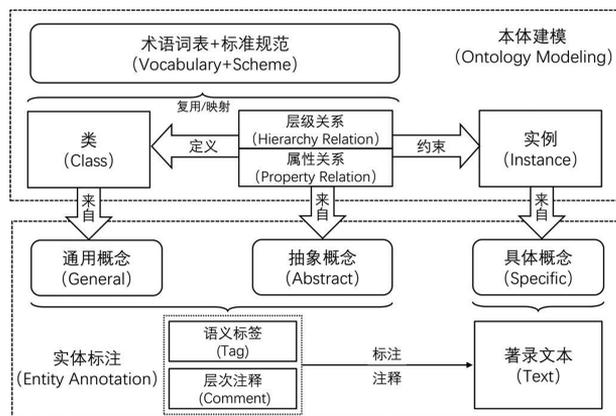


图3 基于文本标注的书画著录事实抽取流程

首先,遵循书画著录时空语义描述模型定义语义标签,并完成标签集构建。语义标签主要来源于模型中的通用概念和抽象概念,由字段类型和标签名称组合构成,例如“G1.创作者”。其次,将书画著录中的语义实体与时空语义描述模型中的概念定义进行匹配,并通过标注工具将著录文本转换为对应标签下的标注实体。再次,面向标注数据进行书画著录本体建模,在复用已有术语词表或标准规范的基础上,依据时空语义描述模型对本体类和本体属性进行定义,并通过属性约束实现标注实体的概念界定和关系揭示。其中本体类、层次关系、属性关系主要来源于时空语义描述模型中的通用概念和抽象

概念,本体实例则主要来源于书画著录的标注实体。在此基础上,为了反映著录文本的概念层次,在创建本体实例时中需复用 RDFS 词表的 `rdfs:comment` 属性,以存储其概念层次注释信息,例如“Thing.Content”“Person.Produce”等。

3.2 基于实体匹配的书画著录事实推理

书画著录事实推理是将标注实体与各种互证数据源进行对齐,再通过实体匹配揭示各型实体的时空属性和语义关系,进而建立时空关联的过程。在数字人文实践中,相关研究者主要通过面向特定领域的时空数据建模,以实现实体的时空属性定义和时空关系揭示。而在书画著录领域,受制于书画著录的文种体裁特性和历史文本的措辞用语特征,直接反映时空语义的实体关系往往相对稀疏,仅依赖时空数据建模难以有效揭示书画著录中人物、题跋、鉴藏印等非时空实体所隐含的潜在时空语义。因此,本文引入基于实体匹配的书画著录事实推理过程,如图4所示。

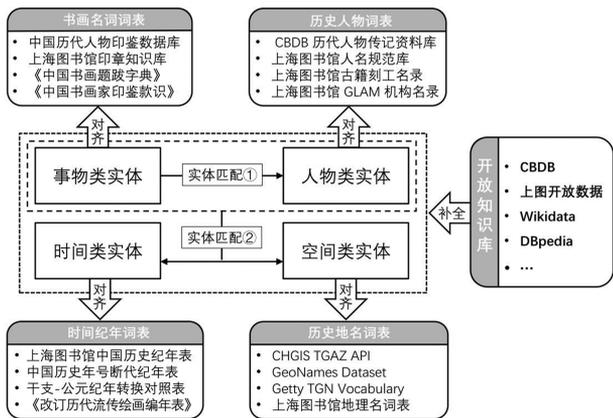


图4 基于实体匹配的书画著录事实推理流程

首先,依据书画著录标注实体所属维度与其他互证数据源进行语义对齐。分别针对人物、事物、时间、空间维度选取历史人物词表、书画名词表、历史纪年表、历史地名表等作为互证数据源;通过标注实体与互证资料的语义对齐实现概念消歧、缺省属性补全和关系修正。其次,在事物类实体与人物类实体间进行实体匹配,通过工具书查阅、数据库检索等途径将题跋、鉴藏印等事物实体与书画创作者、题跋者、鉴藏者等特定的人物实体建立关联,形成“事物→人物”的关联关系,作为进一步揭示其时空关联的中介。再次,在非时空类实体与时空类实体间进行实体匹配,其核心是以人物为线索,通过揭示历史人

物的籍贯、生卒年限、生平轨迹等时空信息以构建“事物→人物→时空”的时空关联链条。此外,在标注实体与互证数据源的对齐、匹配过程中,需通过文献考证和逻辑推断将标注实体与外部开放数据集或知识库中的实体建立映射,通过引入外部实体以对书画著录时空关联证据链进行补全,同时提升证据链与外部数据网络的互操作性。

3.3 基于图数据库的书画著录事实存储

书画著录事实存储是通过构建语义知识库对文献循证的事实数据、互证数据和推理过程进行表示和存储并形成“事实证据链”的过程。在书画著录时空关联构建中,事实存储不仅要实现对原始数据、互证数据的有序组织,还需对书画著录的文献循证过程进行描述和揭示,从而适应知识利用过程中的多场景需求。图数据库是领域知识图谱构建的重要支撑技术,与传统关系型数据库相比,图数据库通过相互连接的节点和边实现知识表示,在存储文献循证数据、反映文献循证过程时具有显著优势。本文基于图数据库技术构建书画著录文献循证数据的事实存储框架,如图5所示。

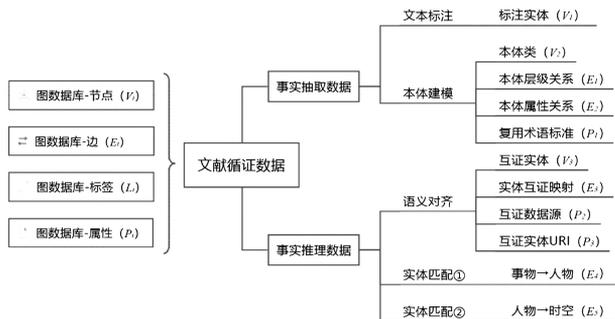


图5 基于图数据库的书画著录事实存储框架

首先,定义图数据库模型 $G = \langle V_i, E_i \rangle$,其中 V_i 为图数据库的节点(vertex)集合, E_i 为反映节间关系的边(edge)集合。 $V_i = \langle L_i, P_i \rangle$,其中 L_i 即节点标签(label),用于表示节点的类型; P_i 即属性(property),用于表示一类节点所具有的性质。 $E_i = \langle L_i, subject, object \rangle$,其中 L_i 表示边所描述的关系类型, $subject$ 和 $object$ 分别表示边所连接的头尾节点。其次,将文献循证过程中生成的标注实体(V_1)、本体类(V_2)、互证实体(V_3)分别通过图数据库的节点(V_i)进行表示,并利用节点标签(L_i)对实体类型进行区分。再次,将本体模型的层级关系(E_1)、属性关系(E_2),语义对齐生成的实体互证映射(E_3)以及“事物→人物→时空”的实体匹配记录(E_4, E_5)作为



图数据库的边(E_i)进行存储。最后,通过定义图数据库的节点属性(P_i),完成对概念层次注释(P_1)、复用术语标准(P_2)、互证数据源(P_3)以及互证实体 URI(P_4)等文本型数据的表示和存储。

4 案例实验:以《石渠宝笈》书画著录为例

《石渠宝笈·初编》(以下简称为石渠宝笈)是清代官修的内府书画集成,被视为中国古代书画著录的集大成者。该书共四十四卷,对清故宫收藏的书画作品进行了全面著录,详细记载了书画作品的形制、题跋、款识等细节,在书画研究领域颇具史料价值。石渠宝笈原书体量巨大、领域背景复杂,本文以

故宫博物院编著的工具书《故宫博物院藏石渠宝笈精粹》为依据^[32],借鉴其作品收录列表确定实验数据采集范围;并以始知阁古代文献数字化项目为数据源^[33],对实验所需著录条目进行查询、获取,形成总计 9865 字的原始语料。

4.1 《石渠宝笈》书画著录的事实抽取

(1)著录文本的语义标签集构建。本实验依据书画著录时空语义描述模型分析石渠宝笈原始语料的内容、文种结构,构建如表 3 所示的语义标签集,并以《洛神赋图卷》著录为例对各标签定义进行说明,表中“**「**”内为石渠宝笈的著录原文片段,“**【**”内为语义标签所对应的标注对象。

表 3 《石渠宝笈》书画著录语义标签集

标签字段	标签名称	标签编码	标注示例
G1 (人物维度)	创作者	G1.Creator	「晋 【 顾恺之 】 洛神图一卷」
	题跋者	G1.Inscriber	「 【 赵孟頫 】 行书洛神赋款识云」
	收藏者	G1.Collector	「又经 【 金章宗 】 秘藏」、「幸为 【 敬仲 】 博士所收」
G2 (事物维度)	著录条目	G2.RecordItem	「 【 洛神赋图卷 】 (宋摹本)」
	鉴藏印	G2.Seal	「下有 【 几暇临池 】 、 【 稽古右文之玺 】 二玺」
	题跋	G2.Inscription	「拖尾 【 李衍跋 】 云」、「又 【 虞集题 】 云」
G3 (时间维度)	朝代纪年	G3.Dynasty	「 【 晋 】 顾恺之洛神图一卷」
	年号纪年	G3.RegionTitle	「 【 大德三年 】 子昂」
	干支纪年	G3.StemBranche	「 【 大德丁未秋九月 】 蓟邱李衍」
G4 (空间维度)	历史地名	G4.PlaceName	「 【 蓟邱 】 李衍」、「 【 云间 】 沈度」、「 【 长洲 】 吴寛」
	建筑场所	G4.SiteName	「下有 【 松雪斋 】 图书印」
A (抽象概念)	用纸	A.PaperUse	「 【 素绢本 】 着色画」
	技法	A.Technique	「素绢本 【 着色画 】 」
	装帧	A.Binding	「晋顾恺之洛神图一 【 卷 】 」
	横长	A.Length	「画幅高八寸四分 【 广一丈七尺八寸七分 】 」
	纵宽	A.Width	「画幅 【 高八寸四分 】 广一丈七尺八寸七分」

(2)著录文本的实体标注。本实验选取“码库斯(MARKUS)古籍标注平台^[34]”进行著录文本的实体标注,其界面如图 6 所示。首先依据语义标签集定义,在 MARKUS 中对标签的颜色、格式、编码进行设置。其次利用 MARKUS 的实体关系标注功能,对著录文本中人物、事物、时空等实体关系进行初步组织,为本体建模提供依据。再次将实体标注数据以结构化格式(CSV)导出,共包含 492 个标注实体,其中人物类实体(G1 字段)99 个、事物类实体(G2 字段)226 个、时间类实体(G3 字段)50 个、空间

类实体(G4 字段)19 个、抽象类实体(A 字段)98 个。

(3)著录文本的本体建模。本实验以表 3 的语义标签集为基础,在复用已有术语词表和标准规范的基础上,构建石渠宝笈著录本体概念模型,如图 7 所示。由图 7 可知,该本体复用了 FOAF、CIDOC-CRM、Dublin Core Terms、OWL-Time 等术语标准,使用“zhulu”前缀表示本体命名空间,通过定义子类关系以揭示各个本体类之间的概念层级,通过定义对象属性以表示本体类之间的语义关系。对象属性中,“hasRecord”用于表示著录条目类与其他标

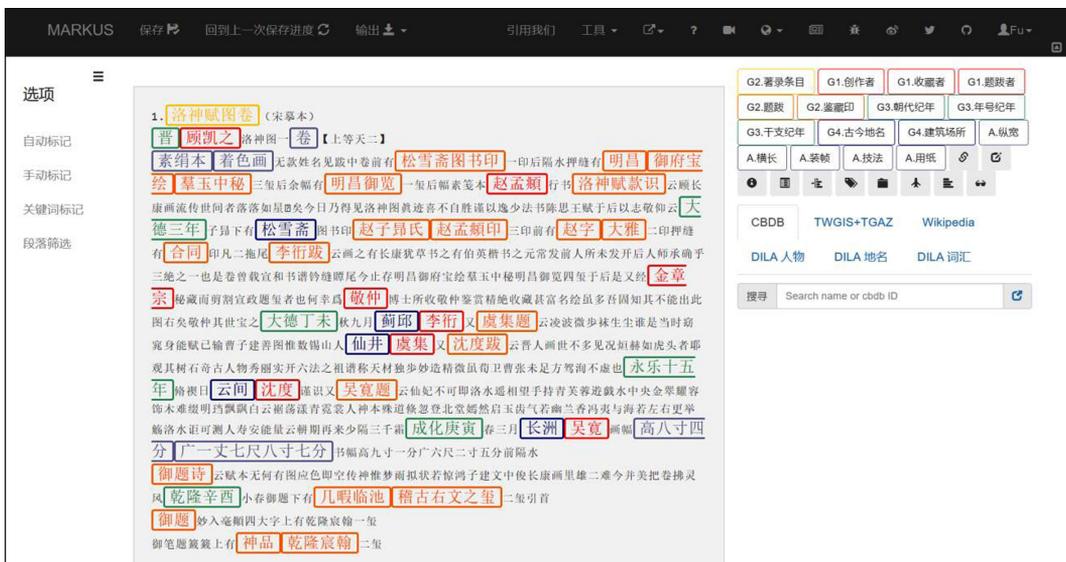


图6 MARKUS 文本标注界面

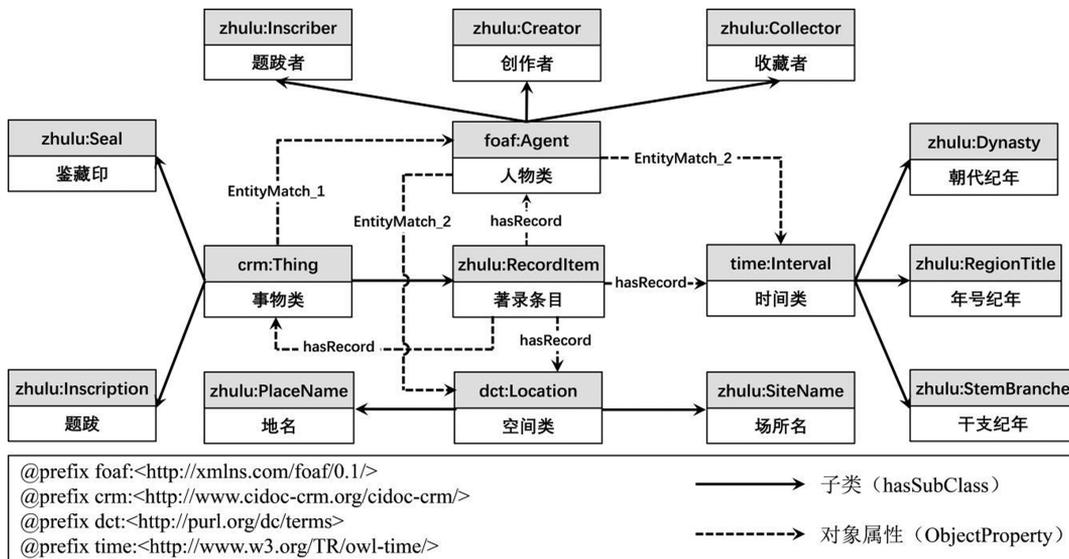


图7 《石渠宝笈》著录本体概念模型

注实体之间的对应关系,“EntityMatch_1”用于表示事物类实体和人物类实体之间的推理关系,“EntityMatch_2”用于表示人物类实体与时空实体间的推理关系。对于标注数据中反映著录文本概念层次的注释信息,以及反映用纸、技法等抽象概念的标注实体,则作为各个本体类的数据属性以文本形式存储。

4.2 石渠宝笈标注实体的事实推理

(1)标注实体的语义对齐。本实验分别从人物、事物、时间、空间维度选取标注实体的互证资料来源,具体包括:中国历代人物传记资料库(CBDB)^[35]、中国历史地理信息系统(CHGIS)^[36]、浙江图书馆历代印鉴数据库^[37]以及上海图书馆发布的历史人名规

范库、中国历史纪年表、地理名词表^[38]。基于上述互证数据源对492个标注实体进行语义对齐,引入外部互证实体128个,建立对齐关系134条。语义对齐过程如下:人物类实体分别通过历史人名规范库和CBDB资料库进行检索,获取对应的互证实体和URI标识,并实现同名消歧。事物类实体主要与历代印鉴数据库进行对齐,通过印章释文检索鉴藏印主人身份和ID标识。时间类实体按朝代、年号、干支等纪年方式分别计算其对应的公元纪年,并在中国历史纪年表中获取对应的互证实体和URI标识。空间类实体中,地名实体通过CHGIS TGAZ API检索历史地名^[39],再通过地理名词表获取对应



的规范地名实体和 URI 标识。场所实体在考证其所在地名称后,参照地名实体进行语义对齐。

(2)标注实体的时空关联构建。本实验遵循图 4 所示的实体匹配流程,在从外部数据源中引入人物实体 19 个、时间实体 62 个、空间实体 33 个进行补充的基础上,建立“事物→人物”关联 143 组,“人物→时间”关联 110 组,“人物→空间”关联 100 组。表 4 以《洛神赋图卷》著录中部分实体为例,阐述基于实体匹配的时空关联构建过程,其中“【】”表示标注实体,“[]”表示经语义对齐或知识库匹配引入的

外部实体,“=>”表示语义对齐,“→”表示推理关系。在“事物→人物”匹配中,依据鉴藏印实体的印章主人 ID 在印鉴数据库中检索相关人物资料,并与人物实体进行匹配,据此在事物实体和人物实体间建立推理关联。在“人物→时空”匹配中,首先进行基于上下文的实体匹配,将人物实体与著录上下文中的时空实体建立关联;其次进行基于人物数据库的实体匹配,通过互证数据源获取人物实体的生卒、籍贯等时空信息,以生卒年取均值作为关联时间实体,以籍贯地名作为关联空间实体。

表 4 基于实体匹配的石渠宝笈时空关联构建

事物实体	人物实体	时空实体	时空关联构建示例
【明昌】 (G2.Seal)	【金章宗】	[公元 1188 年] [阿城]	【金章宗】=>[完颜璟];【明昌】→[完颜璟]→<[公元 1188 年],[阿城]>
【几暇临池】 (G2.Seal)	[爱新觉罗·弘历]	【乾隆辛酉】 [北京]	【乾隆辛酉】=>[公元 1741 年];【几暇临池】→[爱新觉罗·弘历]→<[公元 1741 年],[北京]>
【虞集题】 (G2.Inscription)	【虞集】	[公元 1310 年] 【仙井】	【仙井】=>[抚州];【虞集题】→【虞集】→<[公元 1310 年],[抚州]>
【吴宽题】 (G2.Inscription)	【吴宽】	【成化庚寅】 【长洲】	【成化庚寅】=>[公元 1470 年];【长洲】=>[苏州];【吴宽题】→【吴宽】→<[公元 1470 年],[苏州]>

4.3 石渠宝笈时空关联的事实存储

(1)石渠宝笈文献循证数据的存储。本实验基于 Neo4j 图数据库搭建石渠宝笈文献循证数据的存储环境,定义包含 3 类节点和 7 类节点关系的图数据库模型,如表 5、表 6 所示。遵循该模型将石渠宝

笈书画著录的文本标注、本体建模、语义对齐、实体匹配数据分别导入图数据库管理系统,构建包含 636 个节点、1410 条节点关系、1893 项节点属性的石渠宝笈时空关联图数据库。

表 5 图数据库的节点定义

节点类型	节点标签	节点属性	属性取值说明
标注实体	AnnoEntity	Name	标注实体的原始文本,如“赵孟頫”
		Tag	标注实体的语义标签,如“G1.Inscriptor”
		Comment	标注实体的概念层次注释,如“Person.Transfer”
本体类	OntoClass	Name	著录本体的类名,如“人物类”
		URI	本体类复用术语标准的 URI,如“http://xmlns.com/foaf/0.1/agent/”
外部互证实体	AlignEntity	Name	外部互证实体的标签名,如“完颜璟”
		Source	外部互证实体的来源数据库,如“CBDB”
		URI	外部互证实体的 URI,如“https://cbdb.fas.harvard.edu/cbdbapi/person.php?id=339716”

表 6 图数据库的节点关系定义

关系标签	Subject 节点	Object 节点	关系功能说明
SubClassOf	OntoClass	OntoClass	表示本体类之间的层级关系
InstanceOf	AnnoEntity	OntoClass	表示标注实体与本体类的 is-A 关系
hasRecord	AnnoEntity	AnnoEntity	表示著录条目实体(G2.RecordItem)与其他标注实体(G1、G2、G3、G4 字段)的隶属关系
FeatureOf	AnnoEntity	AnnoEntity	表示抽象概念实体(A 字段)与著录条目实体(G2.RecordItem)的物理特征描述关系
sameAs	AnnoEntity	AlignEntity	表示标注实体与互证实体的语义对齐关系
EntityMatch_1	AnnoEntity	AnnoEntity	表示事物实体(G2 字段)与人物实体(G1 字段)的推理关系
EntityMatch_2	AnnoEntity	AnnoEntity	表示人物实体(G1 字段)与时空实体(G3、G4 字段)的推理关系

(2)石渠宝笈时空关联的查询与呈现。本实验在构建石渠宝笈时空关联图数据库的基础上,通过 Cypher 查询工具进行时空关联的可视化呈现。图 8 以《洛神赋图卷》为例,通过 Cypher 语言构建查询式,输出该作品的时空传承概况,查询式为:“match (m:AnnoEntity{Name:"洛神赋图卷"})-[[:hasRecord]]->(n),(n)-[:sameAs]]->(o),(o)-[:EntityMatch_2]]->(x) with m,n,o,x match (p:AnnoEntity{Name:"洛神赋图卷"})-[[:hasRecord]]->(q),(q)-[:EntityMatch_1]]->(r),(r)-[:EntityMatch_2]]->(y) return m,n,o,x,q,r,y”。由图 8 可知,《洛神赋图卷》时空关联证据链是以标注实体“洛神赋图卷”为核心的多层环状结构,自内向外分别是反映著录文本的标注实体层,通过“事物→人物”匹配(EntityMatch_1)形成的人物关联层以及通过“人物→时空”匹配(EntityMatch_2)形成的时空关联层。在图数据库框架下,证据链中的时空节点分别与公元纪年和现代规范地名进行对齐,反映了特定书画作品的历史传承轨迹;同时,人物、时间、空间节点均通过节点属性建立了与 CBDB 或历史人名规范库、历史纪年表、地理名词表的 URI 映射,可通过访问外部开放知识库中对应的互证实体获取更多相关资料。

图 9 是对图数据库中存储的文献循证记录进行遍历查询所生成的石渠宝笈时空关联证据链图谱,包含 457 个节点和 438 条节点关系,查询式为:“match (m)-[:EntityMatch_1]]->(n),(n)-[:EntityMatch_2]]->(o) with m,n,o match (p)-[:sameAs]]->(q),(q)-[:EntityMatch_2]]->(x) return m,n,o,p,q,x”。图 9 中,时空关联证据链图谱的各个节点形成了规模各异的多个团簇

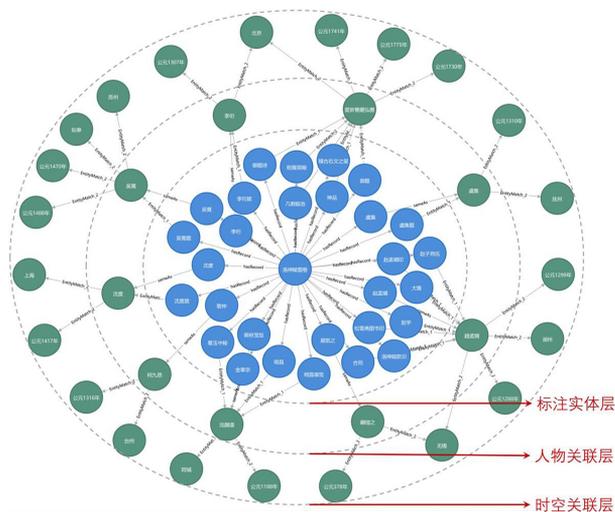


图 8 《洛神赋图卷》时空关联证据链图谱

(Cluster):其中规模最小的团簇至少包含 4 个节点,并通过节点关系形成“事物→人物→时空”的基本证据链,例如“寄傲→项元汴→〈公元 1557 年,嘉兴〉”;而虚线范围中以“苏州”为中心节点,以“公元 1481 年”“杭州”“爱新觉罗弘历”等为中介节点的节点群落构成了图谱中规模最大的团簇(包含 212 个节点)。由图 9 可知,通过对书画著录时空关联进行可视化呈现,能够对书画著录中以著录条目为基本单位的知识结构进行重组与再现,进而直观揭示书画著录中时间、空间及人物实体的多维度关联,为揭示古代书画作品传承轨迹提供了新的分析视角。

4.4 实验结果分析与讨论

上述实验以石渠宝笈书画著录为案例,对基于文献循证方法构建书画著录时空关联的有效性进行检验。在数据获取方面,实验以石渠宝笈书画研究领域的权威工具书为依据,合理确定书画著录调研范围,在书画形制、创作年代、题跋铃印、流传地域等

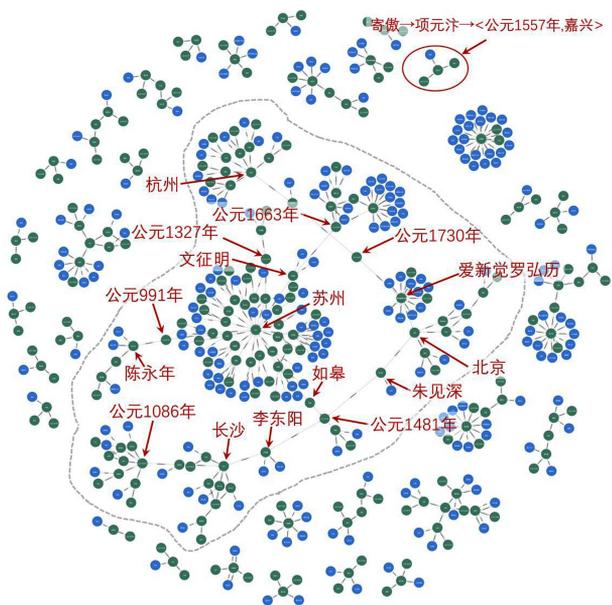


图9 《石渠宝笈》时空关联证据链图谱

方面涵盖了我国传世书画作品的基本分布特征,具有一定的典型性和代表性。实验在完成石渠宝笈书画著录的事实抽取、推理和存储基础上,利用图数据库可视化工具输出了石渠宝笈时空证据链图谱,论证了以“事物→人物→时空”为基本单元的事实推理链条能够满足书画著录时空关联证据链的构建需求。

基于对实验过程与结果的分析,现阶段通过文献循证方法构建书画著录时空关联仍面临以下不利条件:首先,文献循证过程中文本标注、语义对齐的自动化程度有待进一步提升。其次,需在图数据库框架下引入更加高效的数据查询与分析技术。再次,需以更加直观、易用的方式对书画著录时空关联进行可视化、交互式呈现。因此,在未来研究中,一方面需要引入人机结合的文本标注与资料比对机制,以提升事实抽取与推理阶段的效率;另一方面需将整体网分析、图计算、GIS分析等方法引入时空关联查询和呈现过程,以丰富时空关联揭示维度,提升隐性知识挖掘深度。

5 结语

本文重点研究了新技术条件下文献循证方法在书画著录时空关联构建中的应用模式。围绕研究目标,首先研究书画著录的时空语义描述和概念分层注释方法,提出面向书画著录的文献循证与时空关联构建过程,在此基础上以石渠宝笈书画著录为例

开展案例实验。实验结果表明,本文提出的事实抽取、事实推理与事实存储相结合的文献循证过程模型能够满足书画著录时空关联证据链的构建需求,并支持在时空语义揭示的基础上匹配多种书画著录知识利用场景。在未来研究中,还需进一步提升文献循证过程的自动化程度,在文本标注、资料比对、图谱查询、时空可视化等环节不断引入技术驱动、人机结合的数据处理机制,为书画著录时空关联数据集的大规模自动构建提供理论和实践参照。

参考文献

- 1 《中国大百科全书》第三版网络版.书画著录[EB/OL].[2022-03-25]. <https://www.zgbk.com/ecph/words? SiteID = 1& ID = 41931>.
- 2 夏翠娟.面向人文研究的“数据基础设施”建设——试论图书馆学对数字人文的方法论贡献[J].中国图书馆学报,2020,46(3):24-37.
- 3 夏翠娟,贺晨芝,刘倩倩,等.数字人文环境下历史文献资源共建共享模式新探[J].图书与情报,2021(1):53-61.
- 4 刘炜,林海青,夏翠娟.数字人文研究的图书馆学方法:书目控制与文献循证[J].大学图书馆学报,2018,36(5):116-123.
- 5 李文化,陈虹,李冬蕊.数字人文视域下的南海更路簿综合研究[J].大学图书馆学报,2020,38(2):91-98.
- 6 李文化,袁冰.数字人文方法下的《郑和航海图》暹罗湾地名考证[J].图书馆杂志,2021,40(8):88-93.
- 7 欧阳剑,任树怀.数字人文研究中的古籍文本阅读可视化[J].图书馆杂志,2021,40(4):82-89,99.
- 8 潘俊.面向数字人文的人物分布式语义表示研究——基于 CBDB 数据库和古籍文献[J].图书馆杂志,2020,39(8):94-102.
- 9 夏翠娟,林海青,刘炜.面向循证实践的中文古籍数据模型研究与设计[J].中国图书馆学报,2017,43(6):16-34.
- 10 秦昆,康朝贵.计算社会科学的时空分析理论与方法[J].贵州师范大学学报(社会科学版),2016(6):46-48.
- 11 夏翠娟.中国历史地理数据在图书馆数字人文项目中的开放应用研究[J].中国图书馆学报,2017,43(2):40-53.
- 12 胡菊菊.基于 Linked Data 的地理空间知识语义关联模型研究[D].徐州:中国矿业大学,2020.
- 13 姚堪.地理资源知识关联工具研究与实现[D].徐州:中国矿业大学,2018.
- 14 孙一贺,于浏洋,郭志刚,等.时空知识图谱的构建与应用[J].信息工程大学学报,2020,21(4):464-469.
- 15 张雪英,张春菊,吴明光,等.顾及时空特征的地理知识图谱构建方法[J].中国科学:信息科学,2020,50(7):1019-1032.
- 16 夏翠娟,姜秀明,潘威,等.数智时代的知识组织方法在历史地理信息化中的应用初探——兼论图情领域与人文研究的跨学科融合范式[J].图书情报知识,2021,38(3):37-49.
- 17 夏翠娟,陈刚.支撑城市记忆项目的时空数据基础设施建设[J].数字人文研究,2021,1(1):96-104.
- 18 潘威,夏翠娟,张光伟,等.历史地理信息化与图情研究融合的必要性与可行性——以“数字历史黄河”为中心的考察[J].图书情报知识,2021,38(3):50-60,37.
- 19 谈国新,张立龙.非物质文化遗产文化空间的时空数据模型构建



- [J].图书情报工作,2018,62(15):102-111.
- 20 沈语冰.如何趋近绘画:论图像解读的主要理论与方法[J].中国美术,2020(5):86-105.
- 21 魏雅青.从图像学视角分析《韩熙载夜宴图》[J].美与时代(下),2020(10):58-60.
- 22 Wang X, Song N, Zhang L, et al. Understanding subjects contained in dunhuang mural images for deep semantic annotation [J]. Journal of Documentation, 2018,74(2):333-353.
- 23 Panofsky E. Iconography and iconology: an introduction to the study of renaissance art[M]//Panofsky E. Meaning in the visual arts: papers in and on art history. New York: Doubleday, 1955:26-54.
- 24 曾子明,周知.面向数字人文的图像语义描述模型研究[J].情报理论与实践,2018,41(1):116-121.
- 25 Shatford S. Analyzing the subject of a picture: a theoretical approach[J].Cataloging & Classification Quarterly,1986,6(3):39-62.
- 26 王若宸,朱学芳.面向非物质文化遗产的图像语义描述研究[J].信息资源管理学报,2021,11(2):128-137.
- 27 夏翠娟.文化记忆资源的知识融通:从异构资源元数据应用纲要到一体化本体设计[J].图书情报知识,2021(1):53-65.
- 28 夏翠娟.构建数智时代社会记忆的多重证据参照体系:理论与实践探索[J/OL].中国图书馆学报:1-21[2022-05-01].
http://kns.cnki.net/kcms/detail/11.2746.G2.20220413.1505.003.html.
- 29 Janicke S, Franzini G, Cheema M F, et al. Visual text analysis in digital humanities[J]. Computer Graphics Forum, 2017,36(6):226-250.
- 30 王军,张力元.国际数字人文进展研究[J].数字人文,2020(1):1-23.
- 31 项洁,胡其瑞.历史文本的词汇标记及其应用[J].数字人文研究,2021,1(1):48-64.
- 32 李澍.故宫博物院.故宫博物院藏石渠宝笈精粹[M].北京:故宫出版社,2019.
- 33 Github. 殆知阁古代文献[EB/OL].[2022-03-25].<https://github.com/garychowcmu/daizhige20>.
- 34 Ho H, De Weerd H. Markus-text analysis and reading platform[EB/OL].[2022-03-25].<https://dh.chinese-empires.eu/markus/beta/>.
- 35 哈佛大学.中国历代人物传记资料库(CBDB)[EB/OL].[2022-03-26].<https://projects.iq.harvard.edu/chinesecbdb>.
- 36 哈佛大学. China historical gis[EB/OL].[2022-03-26].
<https://sites.fas.harvard.edu/~chgis/>.
- 37 浙江图书馆.中国历代人物印鉴数据库[EB/OL].[2022-03-26].
<http://diglweb.zjlib.cn:8081/zjtsg/zgjcj/index1.htm>.
- 38 上海图书馆.上海图书馆开放数据平台[EB/OL].[2022-03-26].
<http://data.library.sh.cn/index>.
- 39 哈佛大学. Temporal gazetteer api[EB/OL].[2022-03-26].
<http://maps.cga.harvard.edu/tgaz/>.

作者单位:华中师范大学信息管理学院,湖北武汉,430079

收稿日期:2022年3月31日

修回日期:2022年6月23日

(责任编辑:李晓东)

Study on Evidence-based Literature Research and Spatio-temporal Correlation Construction for Calligraphy and Painting Catalogue

Gao Jinsong Fu Jiawei

Abstract: Catalogue literature of calligraphy and painting plays not only an important role in the research of ancient calligraphy and painting but also has a significant reference value for the research on the historical inheritance of ancient calligraphy and painting works. With the rise of Digital Humanities, carrying out the evidence-based literature research upon the humanities data infrastructure will provide new viewpoints for the research about spatio-temporal analysis and inheritance path of ancient calligraphy and painting works. Based on the literature review, this paper firstly studies the method of spatio-temporal semantic description and concept hierarchy annotation for calligraphy and painting catalogue, and then puts forward the process model of evidence-based literature research and spatio-temporal correlation construction for calligraphy and painting catalogue. On this basis, it takes the catalogue of *Shiqubaoji* as an example to carry out an experiment. The result shows that, the evidence-based literature research process model proposed in this paper could meet the requirements of spatio-temporal correlation construction of calligraphy and painting catalogue, as well as provide multi-scenario knowledge service for professionals and the public.

Keywords: Digital Humanities; Evidence-based Literature Research; Spatio-temporal Correlation; Catalogue Literature of Calligraphy and Painting; Knowledge Organization