
文物知识聚合与传播的初步研究

——以上海博物馆“宋徽宗与他的时代数字人文专题”为例

李峰¹

(上海博物馆 上海 200003)

【摘要】 博物馆作为一个知识生产和传播的机构，面对新的时代环境，需要创新的知识化范式。因此，博物馆要利用数字化的知识组织方法，实现基于本体模型和知识图谱的知识建模和知识表达，构建博物馆数字化新型的文物诠释和传播模式，可以提高公众对文物知识生产与获取的体验感，全面提升博物馆文化创新传播能力。除了理论研究外，上海博物馆还不断进行创新性探索，为我国博物馆利用新的数字化知识生产、组织工具，探索出一条具有特色的知识聚合、传播之路。

【关键词】 博物馆数字化 智慧博物馆 知识图谱 本体 文物知识聚合与传播 上海博物馆

【中图分类号】 G260 **【文献标识码】** A

一、引言

现代博物馆从诞生伊始就具有知识生产和知识传播的功能，三百多年来，这已经成为博物馆的一个核心价值。在博物馆知识信息聚合、生产和传播的过程中，博物馆既是旧知识的汇集之地，也是新知识诞生的沃土。信息时代以来，随着博物馆数字化建设的逐渐深入，“智慧博物馆”概念被提出，博物馆界也对知识属性予以充分的关注，提出“在教育与研究方面，通过系统挖掘和整理博物馆藏品的历史、艺术、科学和相关社会等方面的信息，建立新型的知识组织方式，推动研究与教育的互动，把博物馆及相关虚拟平台打造成为学生的第二课堂和公众终身教育的场所”^[1]。因此，从智慧博物馆建设的角度来说，知识的管理、生产和传播是博物馆智慧业务的核心要素，也是智慧管理、智慧保护、智慧服务的立基之所。我们如果要利用文物资源讲好中国故事、让文物真正“活起来”，亟需文物知识生产模式的改革和组织服务模式的创新，这里的关键就是文物知识的聚合与传播，而知识聚合与传播的基础是知识的组织。

当博物馆置身于新的时代语境，面对需要创新的知识化范式，就有必要引入新的手段和方法进行探索。如果说，从知识聚合(knowledge aggregation)到知识组织(knowledge organization)乃至知识传播(knowledge transmission)，这一过程形成了一个博物馆知识创新的关联形态，那么在方法论层面，有必要借助于已构成知识生产重要方式之一的数字人文，以研究展示为导向，以数字思维及跨学科的知识整合方法介入博物馆传统的专业研究活动，形成基于数字内容的知识生产与传播链。在技术层

¹**作者简介**：李峰（1964—），男，上海博物馆副馆长、副研究馆员，主要研究方向：博物馆学、博物馆文创、博物馆数字化。
基金项目：国家重点研发计划课题“文物知识聚合与传播关键技术研究”项目“基于知识图谱的文物知识组织和服务关键技术研发与示范”课题（2019YFC1521204）的阶段成果

面上，利用知识图谱（knowledge graph）与本体（ontology）等方法进行多元分层且系统化的知识体系构建；利用深度学习等智能技术进行博物馆文物数据的挖掘、分析和揭示；利用数据驱动的可视化技术进行展示，让数据以美观的图形方式而非数字方式呈现，清晰有效地传达与沟通信息。基于此，博物馆经过为时不短的数字化建设，长期采集加工所形成的大量包括图像、文本、视频和三维在内的数据资源才会有真正的用武之地。在数字时代，博物馆的核心业务终将与数字技术结合，创新出融合多学科、以技术协力人文而产生的全新业务形态，让传统的知识体系和数字化的知识系统在博物馆中形成竞争性的多层次知识结构，由此给社会提供一种新的选择，进一步满足人们对知识探索、学习和创新的需求。

本文将以上海博物馆（以下简称“上博”）“宋徽宗与他的时代数字人文专题”（以下简称“宋徽宗数字人文专题”）为例，探讨如何利用数字化的知识组织方法，实现基于本体模型和知识图谱的知识建模和知识表达，以全面完备的知识支撑文物叙事，构建博物馆数字化新型的文物诠释和传播模式，提高公众对文物知识生产与获取的体验感，全面提升博物馆文化创新传播能力。

二、博物馆文物知识聚合与传播的现状与问题

1. 国内博物馆数字化建设的现状和瓶颈

国内博物馆的数字化建设，如果从1984年上博成立电脑组算起，逾时近四十年。经过以基础建设为主的信息化时期和以分类业务系统建设为主的数字化建设时代后，2015年开始进入了“一种以物、人、数据动态双向多元信息传递模式为核心”^[2]的智慧博物馆建设高潮。虽然多年的数字化建设取得了非常大的成就，但也存在着明显的缺陷，观察各地全面铺开的智慧博物馆建设，我们不难发现一个明显的特征，就是各馆的建设内容同质性相当高：不外乎是通过二维或三维的文物数字化采集，建设以藏品为主的业务行政管理系统、以传播为主的网站、观众导览及预约等服务系统，间或有一些数字展示的软硬件。但缺乏真正能对博物馆转型或核心业务发展具有推动作用的产品，尤其是能对博物馆的文物研究发挥关键作用的平台，比如协作性文物研究平台；同时也很少产生博物馆知识传播的精品类产品，比较有影响力的反而是那种“网红”产品。出现这种状况的原因当然有很多，其中最主要的是：从主观上讲，我们缺少从知识生产、知识传播角度进行智慧化建设的思考、实践；从客观上说，虽然文物类型众多、文物数据的数量每年都在大幅增长，但文物数据多为多源异构，标准化程度差，因此给文物数据的组织和关联带来相当大的阻碍。很多博物馆在手握大量二维、三维数据的情况下，却茫然不知所措，不知下一步该怎么走，而且这一问题普遍存在，可以说形成了博物馆智慧化建设的一个巨大瓶颈。

2. 国外博物馆知识数据组织的启示

如果我们能体认到在博物馆的数字化传播方面，绝不能仅仅走流行路线、靠卖萌搞怪赚取流量这一招鲜，那么，博物馆智慧化建设就应该进行某种深层次的改变，真正以内容为中心展开知识性传播，体现博物馆的自身价值和初衷。当然，要实现这一点，需要博物馆在现有传播基础上进行一次实质性的“融合创新”。所谓“让文物活起来”，并不简单地等同于让文物动起来。以数字化的一个最基本的应用——展品阐释来说，不是几张图片、一段文字就能完成，也不是简单地让一张画从平面变立体并有动态展示即可，其背后应该有一个知识体系的支撑。而在这方面，国外博物馆似乎比国内博物馆先行一步。2004年，芬兰各博物馆与公司联合推出了基于 Sampo 语义模型的 Museum Finland 门户网站。2006年，国际文献工作委员会（International Committee for Documentation, CIDOC）开发的基于本体的数据模型——CIDOC 概念参考模型（CIDOC Conceptual Reference Model, CIDOC CRM）成为国际标准化组织（ISO）的推荐标准^[3]。CIDOC CRM 提供了一个通用并且可扩展的语义框架，使任何文化遗产信息都能用该框架描述，并且定义了文化遗产领域的各类实体（概念）、属性（关系），形成通用的术语词表，从而推动实现了不同文化遗产信息源之间的信息交换、集成和互操作，现该模型已逐渐成为文化遗产领域的一个通用的主流模型^[4]。近年来，文化遗产机构之间的协同合作被视为一种趋势，比如信息资源管理领域的 GLAM（即美术馆/Gallery、图书馆/Library、档案馆/Archives 以及博物馆/Museum）的馆际协作及资源整合方式已经逐渐成熟，GLAM 也因此成为这一领域的指称。在这一合作的过程中，图情界的一些成熟的知识组织方法，如元数据、主题分类、本体等也逐渐为博物馆界所应用。例如在数字化建设方面堪称楷模的美国克利夫兰博物馆（Cleveland Museum of Art）在2013年开放的 GalleryOne 项目^[5]，一般人可能会更关注其由12米宽超大显

示屏所组成的“藏品墙”（Collection Wall）以及被称为“镜头”（Lens）的互动屏，被它的丰富内容和充分的个性化体验所吸引，但很少有人知道在这些屏幕背后的定制型数据融合管理系统（PictionCMS）的强大支撑作用，以及通过元数据进行的数据控制和组织。正是这些数据库和内容管理系统，才使得这一着眼于激发观众的创造力及其对馆藏作品全面理解的项目目标能够全面实现。而运用主题分类最为引人注目的是英国维多利亚和阿尔伯特博物馆（Victoria & Albert Museum）所做的“中国图像志索引典”（Chinese Iconography Thesaurus, CIT）^[6]。图像志原是一种研究欧洲艺术史的方法，被用来纪录和检索艺术史中的图像。由于没有针对中国艺术进行图像志主题分类的方案，过去欧洲及北美博物馆的中国文物通常只能按照西方的分类体系进行编目著录。CIT 首次改变了这一状况，设计出了符合中国文物艺术品概念的词汇方案。这一系统“期望为博物馆、图书馆与文献库的专业人士创建一套标注准则，提供标准词汇（受控词表），从而提升馆藏文物编目的质量和准确性，促进数字图像跨收藏的信息获取及相互操作”^[7]。CIT 系统以自然界、人类、社会与文化、宗教、神话与传说、历史与地理及文学作品为基本分类，每一类以树形图模式呈现多种层次及其语义联系。借助索引典，研究人员可以选择适当词汇阐明中国文物艺术图像的概念与内容，也能在很大程度上提升馆藏文物编目的质量和准确性，并促进数字图像信息的交换及获取。

3. 国内博物馆知识组织的初步实践

相较于欧美，国内博物馆在这方面还处于刚刚起步阶段。2014 年，我国科学技术部批复设立“文物数字化保护标准体系及关键标准研究与示范”项目，文博单位借助图情界的力量开始了文物元数据的研究。2017 年 5 月，文物数字化保护元数据标准规范各项征求意见稿陆续发布，迈出了文物数字资源标准化的重要一步。2017 年以来，山西博物院联合天津大学在国内文博领域开展文物知识图谱构建的探索性应用研究，首次提出了建立馆藏文物知识图谱知识表达的数据模型，初步开发了文物知识图谱的存储、检索功能，并结合智能问答、计算机语音等技术实现了语义搜索、文物知识问答机器人等知识图谱应用^[8]。2018 年底，上博配合“丹青宝筏——董其昌书画艺术大展”推出了“董其昌数字人文综合展示系统”（以下简称“董其昌展示系统”），在文博界产生了较大的影响力。董其昌展示系统依托上博丰富的藏品资源和雄厚的研究基础，参照 CIDOC CRM 等国际标准、基于关联数据（linked data）设计了董其昌的数据体系，并基本形成半结构化的明清文人书画本体^[9]。除了展示数据的组织之外，董其昌展示系统的开发以董其昌为基点、以古代文人活动为核心、以社会网络关系和历史地理信息技术为两个主要应用点，辅以人工智能技术，通过数据关联和量化分析，以可视化的形式呈现与董其昌相关的时、地、人、事。此次展示在业内引发的轰动效应，在一定程度上激发了国内文博界进行知识聚合与传播的热情。例如敦煌研究院与武汉大学联手，共同对敦煌壁画图像所蕴含的信息进行语义描述、揭示与标注，并提出层次性的图像语义描述框架（SDFDI）^[10]，在此基础上，还进行了敦煌壁画叙词表的构建和关联数据发布^[11]，以及利用元数据、语义标注、国际图像互操作框架（IIIF）等技术标准，构建文化遗产图像交互式数字叙事系统的开发^[12]；秦始皇帝陵博物院则搭建了“四海一”虚拟展览平台^[13]，这是一个利用数字网络技术和信息技术构建的大型秦文化虚拟展示，通过文物知识图谱等技术的运用，以文物知识服务系统的表达，揭示秦代的物质文明和精神文明；最近，江西省博物馆建成了“江西古代名人数字人文研究与服务平台”^[14]，这是依托江西古代历史文化遗产和名人文化资源，以馆藏文物和其他相关研究资料等数据为基础，以江西古代名人知识为主体，利用知识图谱技术建设的专题文化知识服务平台。

综上，虽然国内博物馆在知识组织和传播上已经有了一些成果，但都是个别的、零星的、自觉的行为，在文博界还没有形成较大规模的共识和共同的行动。在这方面，国内博物馆不用说与国外博物馆相比，就是与国内图情界相比也存在较大差距。比如国内图情界很多已采用联合目录等方法进行更关键的数据互通和共享，而博物馆界较多采用各自为战的方式进行数字化建设，很少有行业内各类型或区域内博物馆的联动和数据交换、共享机制以及建设实践；图情界对知识组织从上到下已形成共识并付诸于行动，博物馆界缺乏运用数字化方法进行知识生产的意识，导致出现知识生产效率低、知识表达缺少系统性、自建知识较为孤立等基础问题，以及知识交换机制不完善、公众参与度低等技术面和需求面的问题；在理论研究、数据标准、技术规范等方面，博物馆与图情界相比也差距很大。所以，在知识图谱与本体方法已经日益成为知识数据组织和表达的主要手段时，有志于文物数字化知识生产的博物馆界同仁，应该更多采用多维度、多层次、细颗粒度的领域数据的组织，来进行文物知识图谱及本体的构建，以实现本领域甚至是跨领域的知识融通，进而达成基于数字技术的、多学科知识整合的、创新型的知识生产模式，提高公众对文物知识生产与获取的体验感，全面提升博物馆文化创新传播能力。

三、知识图谱、本体与文物知识聚合与传播

在数字化领域，不论是开展文物知识聚合还是传播工作，都需要有相应的方法与技术的支撑。如果说，元数据的运用为我们提供了结构化数据基础，那么，随着语义网、深度学习等技术和方法的快速发展，可通过知识图谱和本体来达到建立数据或知识间的广泛联系，形成重要知识组织和利用的表达方式，并以此来支持新的研究范式、知识生产方法和知识传播模式。

1. 信息领域知识组织的发展路径

知识图谱、本体主要用于知识组织。所谓“知识组织”，一般泛指对知识数据进行描述、标引、分类和整序。但在大数据环境下，知识组织面临着新挑战与变革，需要结合海量的、多样性的、动态性的数据特征进行多维度、多层次的组织，同时知识组织也逐渐向开放化、语义化的方向发展^[15]。从在这方面发展得更为成熟的图情领域来看，引入语义网技术和本体方法后，由于语义网提供了一套比较完整的知识建模和知识表示规范，超越了元数据以资源对象为中心的知识组织方式，因此图书馆知识组织方法的适用范围得以进一步扩大。作为后来者的文博机构，图情界的做法值得借鉴。实际上，关联数据、知识图谱、本体等都是在语义网发展过程中产生的，是用于构建知识间语义关联的方法。1998年12月，蒂姆·伯纳斯-李（Tim Berners-Lee）正式提出了“语义网”（Semantic Web）的概念。语义网的目标是想用更丰富的方式表达数据背后的含义，让计算机能够理解数据。2006年，蒂姆·伯纳斯-李进一步提出了“关联数据”，这是一种用来定义在语义网上如何发布结构化数据的方法，以使得数据能够相互连接起来，便于得到更好的使用。2011年，英国大英博物馆（The British Museum）基于馆藏目录创建关联数据，建立了关联开放数据项目^[16]。大英博物馆为该项目提供了2500条关联开放数据，并利用这些开放数据陆续推出了多项语义检索、语义数据注释、语义图像注释等开放研究课题^[17]。2012年，美国谷歌公司（Google Inc.）提出“知识图谱”概念，初衷是让用户能够更快更精准地发现新的信息和知识^[18]。在它被成功应用于谷歌搜索引擎之后，知识图谱技术才真正得到了广泛的关注。在文博领域，最有代表性的应用应该是2012年开始的“欧洲数字图书馆项目”（Europeana），资源包括了来自15个欧洲国家的二百多家文化遗产机构的文物、书籍、图片等。开发者采用自主建立的欧洲数据模型（Europeana Data Model, EDM），在不同的知识实体之间构建语义链接，以实现欧洲博物馆、画廊、图书馆和档案馆资源的整合，成为一个欧洲数字文化资源聚合门户，同时也是一个知识聚合的交流平台^[19]。其中，作为数据组织核心的欧洲数据模型就是一个基于本体的跨领域、跨机构的复杂关联数据模型。本体以其跨领域融合的特性日益引起人们的重视，并开始不断地被各领域的研究者应用，成为知识组织方法的“新贵”。

2. 信息技术领域的本体及其应用

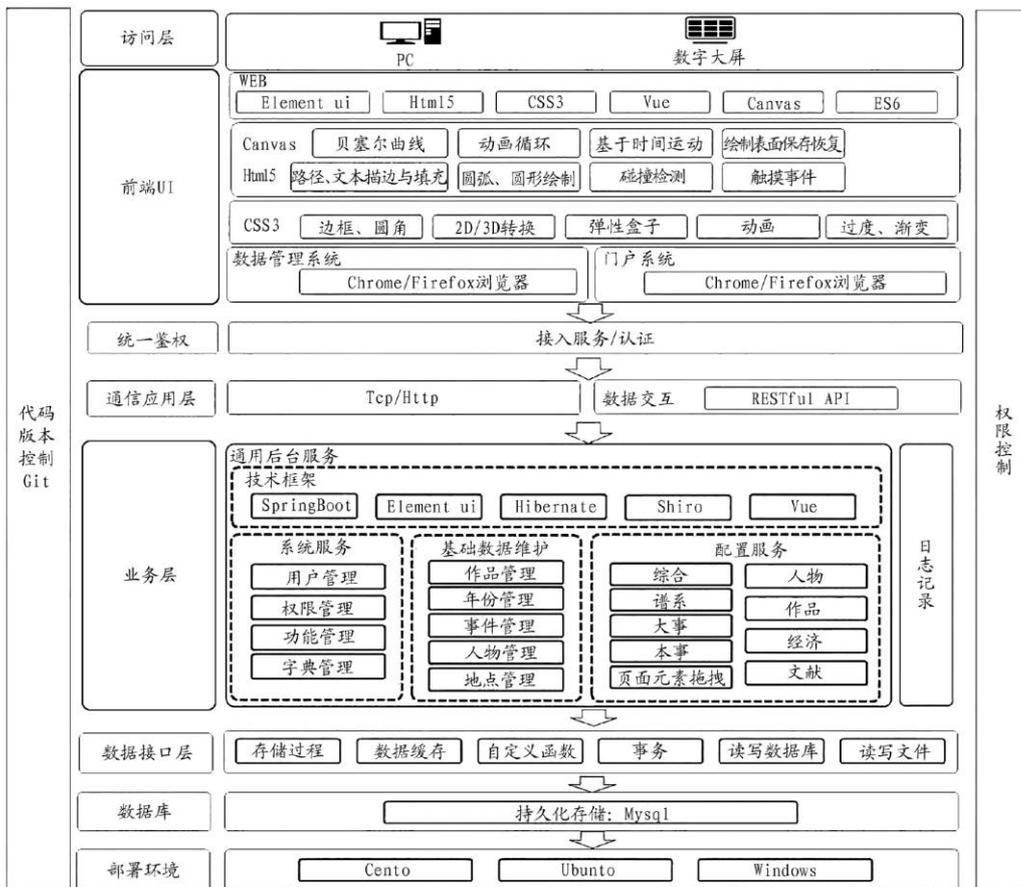
信息技术领域的本体是一种知识表示方法，它用计算机可处理的一种形式化表达方法，对特定领域的概念进行描述。文物数据信息处理的难点就是它有大量的非结构化数据及其多样性来源。因此要做知识组织，就必须首先对多源、不规范的信息进行综合处理，通过信息的相互补充、重组和优化，从而达到数据的规范化和语义化，这就是进行所谓的语义转化，而语义转化的基础就是需要有本体数据模型。建立本体是以知识共享为目的的，从大数据时代知识领域的角度来看，“机器世界中的‘知识融通’要求统一的知识建模和一致的知识表示，前者作为高层、抽象的知识共享模型，超越学科话语体系，在不同领域的术语间建立映射和关联；后者提供一致的知识编码规范，以支持与原生数据格式无关的数据传输、交换和融合”^[20]。在文博界，美国盖蒂基金会（Getty Trust）艺术信息工作组（Art Information Task Force）制定的用于描述艺术作品或作品集合的元数据标准——艺术作品描述类目（CDWA），在相当长的时间受到推崇^[21]。但它多是对某个藏品的个别描述，对数据间的关系定义较少，还构不成一个本体。而前述的CIDOC CRM则在《引言》中清楚地表示它就是要提供一个基于本体的元数据集成概念模型，形成促进多样化的文化遗产信息源的集成、转接和相互交换的形式本体设计^[22]。所以，从建立开始，CIDOC CRM就是一个遵循了知识本体原则的模型：它明确地定义了文化遗产领域的各种实体（类别）、属性（关系）。在CIDOC CRM模型中，“实体是通过属性连接的，属性两端的实体分别是属性的定义域和值域”^[23]。CIDOC CRM所制定的语义框架达成了文化遗产领域异构信息的共同理解，从而实现了异构文化遗产信息的整合和共享^[24]。也正是因为CIDOC CRM所具有的知识本体的特征，其在文化遗产及文博领域得到了广泛的运用。近年来，上博开始了将基于CIDOC CRM本体模型和词表的知识组织方法应用于数字人文实践中的初步尝试，即2018年

的“董其昌展示系统”；在此基础上，2021 年上博在此方面进行了更大规模的探索，即“宋徽宗数字人文专题”的开发制作，并以此作为上博在文物知识聚合与传播方面的一次新努力。

四、上博文物知识聚合与传播的再尝试

1. 项目的背景

在后疫情时代，博物馆的数字化传播和展示在线上线下都成为热点。但目前国内博物馆的数字展示都在往所谓的“沉浸式”方向倾斜，以视觉冲击、感官体验为特色，以赢得流量、吸引打卡为追求，而博物馆作为一个知识生产和传播的机构，其初衷有可能在无形中被消解了。因此，在沉浸式数字展示泛滥之余，我们应该也必须为博物馆的数字展示找到另外一条出路，那就是数字化的知识阐释。博物馆数字化的文化魅力在很大程度上也体现在通过阐释使知识得到传播，体现在观众对知识的发现和解读，体现在观众对知识的认识与体验的欢乐中。在这方面，数字人文给我们提供了一种很好的思路和方法。2018 年，上博的“董其昌展示系统”率先作了一些尝试，2021 年的“宋徽宗数字人文专题”是在这方面继续开拓的标志(图一)。从“董其昌展示系统”到“宋徽宗数字人文专题”，不单单是一个内容由点到面的扩展，更是力图在博物馆数字阐释上进行一些新的尝试，即在传统史学和器物学的基础上，融入政治史、社会学、经济史、文献学等多学科视角及语境，帮助观众对人、文化乃至背后的社会进行多元化解读，引起观众的发散性思考。通过数据的标引聚类、信息的整合呈现，对时、地、人、事、物全面关联：远观精读书画精品，多维度剖析赵佶的人生百态，可视化梳理人物关系，力求实现博物馆数字化展示见物见人、更要见社会的理想目标，成为知识聚类与传播的一个良好案例。



图一// “宋徽宗数字人文专题”系统架构（图片来源：作者提供）

2. 项目的主要内容

“宋徽宗数字人文专题”项目由“综合主页”及“大事”“人物”“作品”等多个版块组成，从“人”“事”“物”三个维度实现数据的知识化、结构化和可视化呈现。

首先是“综合主页”。作为系统的总入口，采用动态时空地图的形式，将由数万条宋代文献、实物及相关学术研究成果资源转化而成的可读数据，根据数字化的原则结构整合，形成以时间为经、空间为纬，人物、事件、实物为坐标的整体知识呈现系统，意在整体呈现在宋徽宗 54 年人生中北宋末年历史文化的发展趋势、重大节点和实物例证，体现各孤立知识点之间显性与隐性的关系及彼此的层级，帮助研究与学习者透过真实的数据远（distant reading）主题，形成对北宋末代表人物和典型文化面貌的认知（彩插五：1）。

其次是“大事”版块。以事件为中心，采用“图谱化总览页面”“时空地图页面”两种形式，从综合、分类两个维度，通过和战、灾异、文艺三个角度，多样化聚类并深度分析北宋末年史事的政治、经济、文化内涵（彩插五：2）。“图谱化总览页面”将徽宗时期大事以散点形式组织于整体的可视化页面之中，以视觉逻辑呈现各事件看似散漫实则互具因果的关系，并提供时间、人物、地点元数据的图谱检阅功能，反映不同数据点在数据流中的位置与语义。“时空地图页面”以时间轴和地图相结合，并列对照北宋末年中国与欧洲的历史、文艺发展进程，展现中华文明在世界范围内的独特成就、位置以及前大航海时代文明交流的区域性特点。和战、灾异、文艺三个大事页面则从不同的横切面深入展现和探讨了北宋末年安全环境、自然环境、文化环境对历史进程的影响。

再次是“人物”版块。以人物为中心，展示宋徽宗本人、家庭、政治人物关系及其对北宋末期历史及文化走向的影响。由于宋徽宗在政治及文化上独特的中心地位，将其本人、家族及作品数据集以独立单元予以展示，并与“作品”等版块互相链接，形成既独立又整体的展现效果。“人物”版块以“变法人物”“文艺地图”“行迹图”为主要形式，分别展现徽宗时期政治人物、文艺人物及具有代表性政治文艺人物的结构化数据，从不同侧面和层次揭示其基本历史形态及产生的影响。“变法人物”以北宋著名的“元祐党争”为切入点，引用中国历代人物传记资料库（CBDB）的数据，对徽宗时期政治人物错综复杂的政治关系进行梳理并分层级、分类型展示，形成一对多、多对多的可视化关系描述，反映内在的关联规律，构成动态的关系图谱（彩插六：1）。

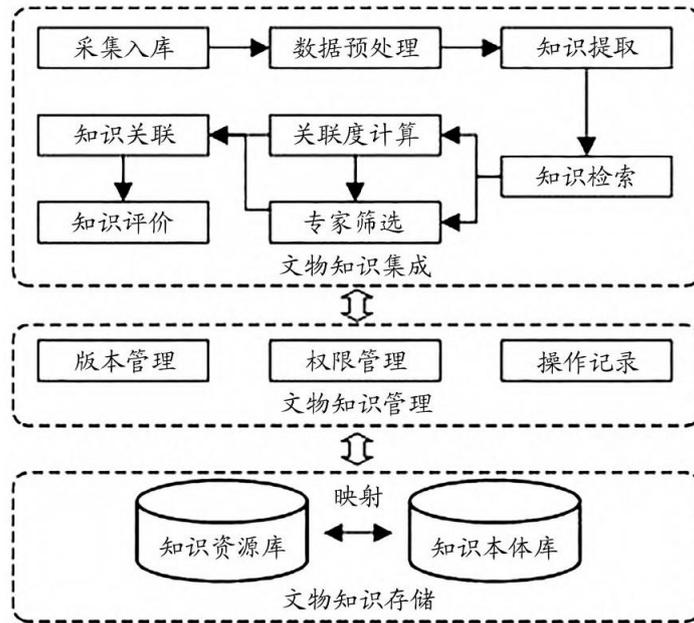
“文艺地图”依托年代、地点、人物基础数据库，结合历史地图、“搜韵”和 CBDB 等平台的数据，力图全面呈现北宋文学、书法、绘画等著名艺术家的时空分布规律、创作流行趋势及具体作品实例。“行迹图”分别挑选徽宗时期最具代表性的文艺人物——苏轼和李清照进行个案呈现。前者是宋代最为著名的政治家、文学家、艺术家，活跃于宋徽宗父兄在位期间及徽宗的青少年时期；后者的生卒年代贯穿反映徽宗人生的整个时期，是当时最具代表性的文艺人物，父族和夫族又深度参与当时政事。系统通过对二者生平经历的编年化梳理和空间定位，以及时空坐标与他们创作实例的精确关联，从侧面样本化地呈现了徽宗青少年时代北宋政治、文艺的特征及其对徽宗思想、艺术的影响，以及徽宗时期思想、艺术的发展流变和对南宋时期文艺发展的深远影响。

最后是“作品”版块。通过对徽宗时期书画艺术品进行分类标注和聚类分析，以实例形式呈现北宋末年艺术发展的大致面貌（彩插六：2）。“作品”版块分别构建了针对书画类文物的本体，形成了书画文物的概念模型和知识图谱，并引入深度学习技术，对书画实物的视觉主题元素以及印章、题跋元素进行语义化聚类，最终形成智能化关联和呈现。

3. 项目的知识组织和表达

“宋徽宗数字人文专题”项目的一个关键难点就在于对来自不同渠道的、内容庞杂的资源数据进行聚合、梳理、组织和有序的可视化表达。比如以书画为主体的文物展示的组织，上博针对不同来源、不同层次、不同结构、不同内容的文物知识进行综合和集成，实施结构再建，使单一知识、零散知识、新旧知识、显性知识和隐性知识经过整合提升形成新的知识体系，同时不断迭代这一过程，实现文物知识的持续集成演化。图二为文物知识集成的模型，主要的处理流程包括采集入库、数据预处理、知识提取、知识检索、关联度计算、知识关联、专家筛选、知识评价；文物知识集成过程还需要有对应的管理策略，主要包括版本管理、

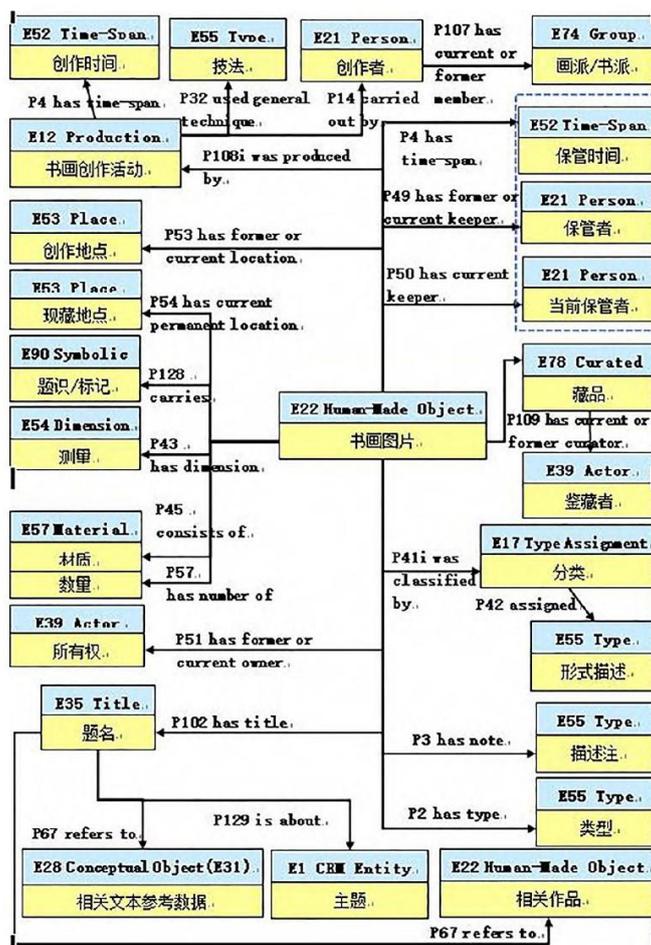
权限管理和操作记录。文物知识演化过程可分为知识协同、知识集成、知识评估、知识更新和知识应用五个过程。



图二//文物知识集成模型（图片来源：作者提供）

上博还对其现有书画文物数据结构和术语标准化情况进行梳理，以上博藏品管理系统中书画类文物采用的本地标准化术语，映射国际文物领域元数据标准《艺术品描述类目》(CDWA)，确认书画类文物元数据方案。同时，进一步定义核心元素作为发展书画本体之基础数据结构和内容，根据元数据元素和其他相关重点知识信息（如文物背景脉络、相关事件）与知识本体标准 CIDOC CRM 进行映射，研究确立映射路径，提出书画知识本体之实体与属性。再根据分析结果扩展书画类文物知识本体模型之实体，如画面元素、地点、主题类型、人物交往关系、相关事件等（图三）。此外，由于书画文物的创作者、创作时间、书派/画派、展览/借出史等亦需要进一步的实体和属性关系表述，故分别建立了相应模型。同时开展基于上博绘画藏品数据的主题词分析，研发书画类文物专题知识图谱，为知识聚合叙事传播提供支撑。书画文物知识图谱采用网络本体语言(ontology web language, OWL) 作为本体的形式化表达方法，主要包含书画文物本体、人物本体、年代本体、地点本体及事件本体。使用本体作为框架，利用语义网技术，通过知识抽取、融合及关联，对以书画为主体的文物数据、文献及数字资源进行组织和聚合，形成覆盖人、事、时、地、物的庞大信息网络，解决博物馆信息资源的多源异构、难以整合、知识服务水平较低的问题（彩插七：1）。

在构建书画文物专题知识图谱过程中，需要整理书画中的花鸟、人物等元素，同时需要提取印章、题跋等信息，从而丰富书画文物知识。由于一件书画文物往往有着大量主题、印章、题跋等信息，依靠传统人工提取虽然精度较高，但是需要花费大量时间，无法满足知识图谱应用需要。因此，需要研究特色文物的识别技术，针对人物、动物、山头、树木、桥梁、房屋等绘画元素以及印章、题跋等信息，利用人工智能技术(artificial intelligence, AI)，让计算机辅助专家进行信息提取（彩插七：2）。上博采用深度学习的方法，以卷积神经网络提取图像特征为基础，应用监督/无监督方法学习特征分布到类别的映射，实现对绘画元素的自动识别、标注，并构建基于构图元素的绘画内容的索引，查询比对与分析系统，通过 k 均值聚类算法(k-means clustering algorithm)实现图像的聚类，以此辅助绘画的研究和鉴赏工作。目前，从实现的效果看，在绘画元素的提取方面取得了令人满意的效果；在印章和题跋信息的提取上，由于古文字本身的识别难度大和易受干扰，效果还不是十分理想，将来随着书法文字识别技术(OCR)的提高，相信还会有很大的提升空间。



图三//书画文物知识本体基本数据模型截图（图片来源：作者提供）

五、结语

无论是博物馆数字化还是智慧化建设，本质上是大同小异的，最终目的都是推进、革新博物馆的业务模式，更好地履行博物馆作为知识生产、传播机构的职能。因此，利用新的数字化知识生产、组织工具，探索一条属于博物馆自己的知识聚合、传播之路，应该是未来博物馆改革发展的题中应有之意。当博物馆数字化基本实现了传统资源向数字资源的转化，以及能够被计算机存储、处理和展示，并有了一定的“数据化”基础时，基于数据的知识组织和传播就是对博物馆数字化的一种拓展与推进。从上博的实践来看，如果说数字化改变了资源储存和展示形态，那么数据知识组织就是建立了数字资源之间的联系，形成了一个供博物馆开展各种知识传播应用的知识环境。

参考文献：

- [1] 宋新潮：《关于智慧博物馆体系建设的思考》，《中国博物馆》2015年第2期。
- [2] 同[1]。
- [3] 《什么是CIDOC CRM?》，[EB/OL][2022-03-11]<https://www.cidoc-crm.org/>

-
- [4]王绍平:《FRBR 与面向对象模型》,《新世纪图书馆》2007 年第 2 期。
- [5]Jane Alexander, Jake Barton, Caroline Goeser. Transforming the Art Museum Experience: Gallery One. [DB/OL] [2022-03-11] <https://mw2013.museumsandtheweb.com/paper/transforming-the-art-museum-experience-gallery-one-2/>
- [6]“中国图像志索引典”, [EB/OL] [2022-03-11] <https://chineseiconography.org/>
- [7]张茜:《专访张弘星 | 在数字时代, 为中国传统图像建立知识图谱》, [EB/OL] [2020-04-20] https://mp.weixin.qq.com/s/r_FSCy4u8z_XR3HAfe_Ca_XXHw.
- [8]王晓丽:《数字化展示在博物馆的应用和研究——以山西博物院为例》,《文物鉴定与鉴赏》2020 年第 22 期。
- [9]童茵、张彬:《董其昌数字人文项目的探索与实践》,《中国博物馆》2018 年第 4 期。
- [10]王晓光、江彦斌、张璐:《敦煌壁画图像语义描述层次模型实证研究》,《图书情报工作》2015 年第 19 期。
- [11]王晓光、侯西龙、程航航、夏生平:《敦煌壁画叙词表构建与关联数据发布》,《中国图书馆学报》2020 年第 4 期。
- [12]王晓光、谭旭、夏生平:《敦煌智慧数据研究与实践》,《数字人文》2020 年第 4 期。
- [13]“四海一”虚拟展览平台, [EB/OL] [2022-03-11] <http://124.115.173.6:8079/>
- [14]江西古代名人数字人文研究与服务平台, [EB/OL] [2022-03-11] <http://www.jxmuseum.cn/zstp/index.html>.
- [15]李亚婷:《知识聚合研究述评》,《图书情报工作》2016 年第 21 期。
- [16]Connect, Communicate and Represent Knowledge with Context. [EB/OL] [2022-03-11] <https://researchspace.org/>
- [17]刘绍南、杨鸿波、侯霞:《文物知识图谱的构建与应用探讨》,《中国博物馆》2019 年第 4 期。
- [18]杨延云、杜建强、熊旺平:《本体和知识图谱的比较研究》,《江西中医药大学学报》2021 年第 4 期。
- [19]贾君枝、史璇:《面向博物馆关联数据的语义转换综述研究》,《晋图学刊》2016 年第 2 期。
- [20]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》,《图书情报知识》2021 年第 1 期。
- [21]石秀敏:《CDWA 和 DC 元数据标准与故宫博物院绘画类藏品信息指标体系的对照研究》,《故宫学刊》2016 年第 2 期。
- [22]同[3]。
- [23]翟琼、刘宏哲:《CIDOC CRM 在数字博物馆中的应用》,《中国博物馆》2015 年第 2 期。
- [24]毛雅君:《国际标准〈信息与文献文化遗产信息交换的参考本体〉(ISO 21127:2014) 解读》,《国家图书馆学刊》2018 年

苏轍
1039—1112
苏轍(1039—1112)字子由，一字同叔，号颍滨遗老，眉州眉山(今属四川)人。苏洵子。嘉祐二年(1055)进士。神宗时为王安石属官，反对新法，受其兄苏轼清议影响，遭贬谪。哲宗时官至尚书右丞、门下侍郎，亦屡受贬滴。居许州，乃颍滨遗老，著文汪洋恣肆，为一时名家。唐宋八大家之一。有《栞城集》、《春秋集解》等。

1087丁卯北宋哲宗元祐二年
苏轍、刘敞编写神宗御制文集(汴京)

1089己巳北宋哲宗元祐四年
苏轍献《神宗御集》

当前时间
公元纪年:1097
帝王纪年:绍圣四年
干支纪年:丁丑

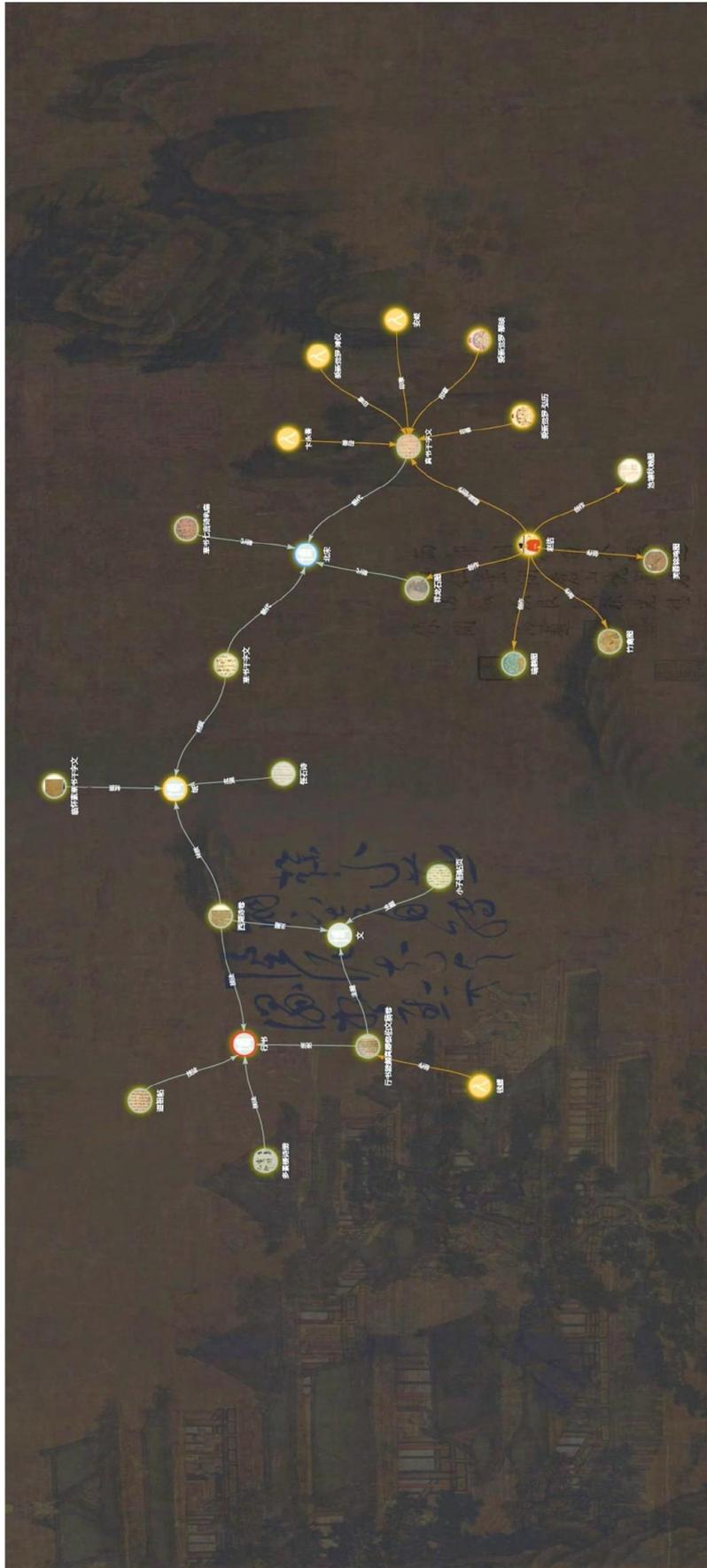
重要人物
刘季 林希 吕公著 吕大防 文彦博 苏轍 苏轼
司马光 赵鼎

事发地点
汴京 徐州 宿州 雷州 潮州 昌化军

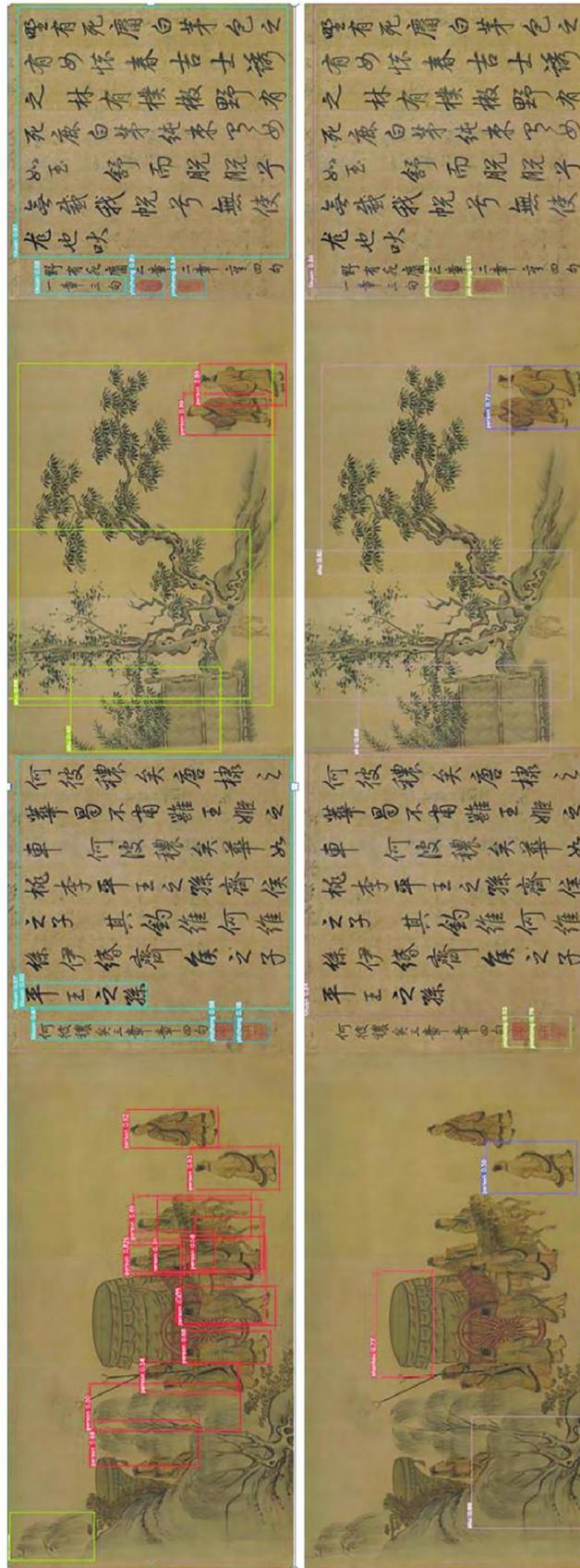
1085 赵鼎封...
1090 苏頌进...
1098 章惇等...
1109 何执中...
1125 内祥款...

绍圣追贬(循州...)
文彦博去世
苏轍罢官(雷州)
苏轼贬谪海南
林希进宰相(汴京)
夏元丰维茶法
旱灾:夏,两浙...

2.“宋徽宗数字人文专题”的“大事”版块展示界面



1.上博书画展示图谱可视化界面



2. 古代绘画元素在识别初始阶段与算法调整后的效果对比图(下为识别初始阶段与算法调整后的效果对比图,上为算法调整后的效果图)