

# 发现、重构与故事化：数字人文视角下档案研究的路径与方法<sup>\*</sup>

牛力 高晨翔 张宇锋 闫石 徐拥军

**摘要** 数字人文研究体系对图书馆学、情报学与档案学的理念与模式产生了显著影响,给图情档学科带来新的发展空间。目前,数字人文对档案学研究的影响多体现在理论前瞻与对象选择层面。如何推动二者在研究路径与方法层面的深度交互,充分发挥档案学理论与方法的效用,发掘档案资源的多维价值,是数字人文视角下档案研究的核心问题。本研究首先分析数字人文环境下档案研究路径与方法变迁的依据及映射机制,在此基础上提出以“发现”“重构”和“故事化”三原语为主线档案研究路径基本模式,形成数字人文视角下档案研究的方法论体系。其次,基于档案价值视角将三原语操作化为具体的技术序列,形成数字人文视角下档案研究的微观技术系统。最后,以吴宝康档案数据为案例对象,对所提方法论及技术体系的有效性进行检验。数字人文视角下档案研究的路径与方法重组,将对档案学理论、实践方式以及学科间的深层交互产生一定影响。图 11。表 2。参考文献 40。

**关键词** 数字人文 档案数据 档案叙事 知识组织 知识图谱

**分类号** G254

## Discovering, Reorganizing and Storytelling: Paths and Methods of Archives Research on the Perspective of Digital Humanities

NIU Li, GAO Chenxiang, ZHANG Yufeng, YAN Shi & XU Yongjun

### ABSTRACT

Digital humanities (DH) aims to solve problems in humanities by using digital ideas and technologies in the fields of computers and the Internet. Nowadays, DH has a significant impact on the concepts and models of Library, Archival & Information Science (LIS). However, compared to Library & Information Science, Archival Science has neither comprehensively accepted the ideas and technologies from DH nor revealed its value in research paths and methods to DH. In some DH researches, archives are only the secondary resources or objects while the results or conclusions of these researches are not related to Archival Science. Thus, it is an essential to promote in-depth interaction between DH and Archival Science at the level of research methodology and to mine archives value from the multidimensional perspective.

By integrating the concepts and technologies, this paper puts forward a two-tiers framework in the paths

<sup>\*</sup> 本文系国家自然科学基金项目“价值导向的档案数据知识发现与服务模式研究”(编号:19BTQ098)的研究成果之一。(This article is an outcome of the project “Research on Value-Driven Knowledge Discovery and Service Model of Archives Data” (No. 19BTQ098) supported by the National Social Science Foundation of China.)

通信作者:高晨翔,Email: greene\_gao@ruc.edu.cn, ORCID: 0000-0002-6446-5822 (Correspondence should be addressed to GAO Chenxiang, Email: greene\_gao@ruc.edu.cn, ORCID: 0000-0002-6446-5822)

and methods of archives research on the perspective of DH. First, based on theories and practices of archival science, the authors analyze the theoretical basis and mapping mechanism of archives research in DH background including some core elements such as philosophy basis, research objects, service models and, etc. Second, through reviewing papers and ideas of multi-disciplines, a new model of archives research paths has been proposed. We define the archives research paths as the three core primitives called discovering, reorganizing and storytelling. Specifically, discovering is deconstructed to three sub-primitives containing positioning, contexts recognizing and mining in order to discover all the memory entities and their relationships in archives resource without any prejudice. Reorganizing includes static relating and dynamic clustering, which are implemented by building knowledge organization models such as Simple Event Ontology (SEM) and International Council on Archives Records in Contexts Ontology (ICA RiC-O) as well as adding semantic labels to represent an entity from different perspectives. User detecting, framework constructing and archives narrating are fused as storytelling to furthering the distant reading of archives resources. Third, correspondingly, the authors operate the three primitives to data processing technologies from the perspective of archives value: 1) Maintaining the value, aiming to transform multi-modal archives to archives data. 2) Mining the value, focusing on the dynamic and flexible knowledge organization of archives data. 3) Realizing the value, paying attention to the contextual and immersive narrative of archives data. Then, this paper selects archives data sets of archivist Wu Baokang as research case to validate the rationality and effectiveness of the paths and methods proposed above, and tries to tell some stories about Wu Baokang with different topics. Ultimately, we find that the new paths and methods of archives research on the perspective of DH could demonstrate the lofty beliefs and academic qualities of the older generation LIS scholars in New China.

In the follow-up research, we will focus on the limitations and problems, pay attention to the collection and organization of archives with multiple types and sources. We will also set some standards about user participation in archives data description and organization while asking users' opinions related to our knowledge discovery and service platform. In conclusion, we believe that the paths and methods mentioned above will induce more thinking about the archives resources development and archival science research. Scholars from library & information science and archival science should discover more consistent research methods and interests in DH environment. 11 figs. 2 tabs. 40 refs.

#### KEY WORDS

Digital humanities. Archives data. Archives narrative. Knowledge organization. Knowledge graph.

## 0 引言

数字人文理念、方法与技术对传统人文学科的研究路径产生了重要影响,引导相关领域学者从不同视角重新发现并认识人文研究的多维价值<sup>[1]</sup>。2009年,王晓光将“数字人文”概念引入我国的图书情报与档案管理领域,开启了

我国学者“数字人文与图情档融合研究”的求索之路<sup>[2]</sup>。2017年以来,图情档学界迎来了数字人文研究的“热浪潮”<sup>[3]</sup>。2019年1月10日《光明日报》理论部发布2018年度中国人文学术十大热点,“大数据视域下数字人文研究”位居其中<sup>[4]</sup>。2020年1月揭晓的“2019年度中国图情档学界十大学术热点”中,“图书情报与档案管理视野下的数字人文”位居第二<sup>[5]</sup>。目前,上海

图书馆、武汉大学数字人文研究中心、中国人民大学人文北京研究中心、北京大学信息管理系 KVision 实验室等研究机构纷纷在数字人文与图情档融合领域耕耘,数字人文视角下的图情档学术研究在理论与实践两个层面均取得了一定进展。

在图情档学科中,档案以其真实性、可靠性等独特属性受到数字人文学者的广泛关注。然而,在以档案为对象的数字人文研究中,档案学理论与方法的介入程度不高,未能发挥其应有的指导价值。数字人文领域的理念与方法在档案研究过程中的应用模式与其他类型文献没有明显差异,这种研究实践对档案学理论及方法的反馈效应有所不足,具有档案学学科特色的数字人文研究仍有待发展。当前,档案学学科内部诸要素及属性的演变为数字人文与档案学的深度交互式研究提供了前提条件。理论层面,在后现代主义哲学思潮的影响下,档案学者开始强调普通民众及特定社会群体在档案建构中的多元价值,反思传统档案学研究中的阙漏。实践层面,综合性档案馆更为侧重档案开发与利用服务,传统的档案创造者、所有者与管理者角色均得以重构,为数字人文视角下的档案学研究提供了必要空间。同时,随着馆藏档案数字化、数据化与文件管理“单轨制”的不断推进,“数据态”档案更加受到学者关注,档案基础数据的积累为数字人文研究提供了物质条件<sup>[6]</sup>。

基于以上背景,国内档案学学者已经展开了数字人文与档案学的思辨研究,主要表现为宏观层面的反思式、批判式研讨,预估数字人文对档案学学科带来的机遇与风险。然而中观的路径及方法层面仍处于起步阶段,缺乏明确的研究案例与成果。本文拟探讨数字人文环境下档案研究的路径、方法及其嬗变机制,着力构建数字人文视角下档案研究路径的基本模式和相应的技术体系。最后通过相应案例展示该模式下的实践成果,同时验证路径与方法的合理性,以期对数字人文与档案学在学科层面的交互模式加以梳理,为档案的数字人文研究提供理论

和方法。

## 1 国内外相关研究

在技术层面,研究者通常借助光学字符识别(OCR)、3D重建等技术将传统档案资源数字化,搭建数据挖掘系统对档案数据进行整理、统计、开发与展示,最终以多重感知检索结合动态可视化呈现的形式实现知识表达<sup>[7,8]</sup>。崔春等以普瑞特信息与图书馆科学学院的“关联爵士”项目为基础,从档案数据源、知识组织体系构建、众包分析、人才整合等角度考察了数字人文视角下档案文献知识组织与开发的基本模式<sup>[9]</sup>。张斌等从资源、过程和参与者视角梳理了档案馆与数字人文的关系脉络,同时就系统、平台、工具、用户界面等内容对数字人文环境下档案馆技术体系的影响进行了思考<sup>[10]</sup>。

人文学科学者也对数字人文环境下的档案研究进行了探索,他们深谙档案资源的基本特征,并在数字人文视域下进行了深入反思。Mendelsson、Falk 和 Olive 指出纸质档案数字化开发需要档案学家、档案馆员、信息科学家与 IT 从业者通力协作,将档案资源置于特定背景与环境中进行综合治理<sup>[11]</sup>。Jarlbrink 和 Snickars 以传统《瑞典晚报》的数字化过程为研究对象进行了民族志研究,认为图书馆、档案馆需要对数字人文技术进行深度考察,以避免失去源头追溯与质量控制的能力<sup>[12]</sup>。特别提出的是,Klein 利用命名实体识别与力导向图对美国前总统托马斯·杰斐逊的信件档案进行了分析,在还原档案中黑人奴隶活动的同时揭示了传统人文研究中的“惯性研究思路”与“选择性无视”等研究取向及价值观问题<sup>[13]</sup>,以档案数据为基础对被掩盖的历史进行了发掘与重构。

在档案学科的内部认知层面,冯惠玲于2015年率先提出“数字记忆”概念,首次将数字人文与档案相结合,指出档案是记忆建构过程中不可或缺的资源,应借助数字技术构建大众化的、资源内涵丰厚且组织合理的数字记忆<sup>[14]</sup>。

徐拥军和加小双在调研国内外记忆实践的基础上,从数字人文角度指出记忆实践数字化转型的必要性,强调这一过程在文化传承与保护中的重要价值<sup>[15]</sup>。曾蕾等界定了图档博智慧数据在数字人文应用中的角色,认为相较于体量庞大、异质程度较高的大数据,档案数据具有干净、明确、受信任的特征<sup>[16]</sup>。部分青年学者讨论了数字人文对档案学现行研究模式的冲击与影响,在分析数字人文与档案学交互方式的基础上提出了数字人文视域下档案学发展变革的策略<sup>[17,18]</sup>。

在数字人文领域,以数据收集、整理、挖掘与可视化服务为核心的档案研究路径基本形成,研究者对档案资源的独特性有所认知,并开始与其他类型文献资源加以区分。然而,研究者对档案研究路径及方法的归纳整合、规范阐释、步骤划分及其逻辑联系尚未进行有效界定,对档案资源基本属性的技术化表示尚未提出明

确见解,同时对服务层面档案价值的延伸式表达也没有给予相应关注。因此,本文拟结合案例对数字人文视角下档案研究的路径与方法进行归纳、界定与阐释,探究数字技术作用下档案资源组织、开发与服务的新模式,关注技术要素在档案资源基本属性表达与深化方面的实践效用,同时对档案服务中的叙事方式进行逻辑重组。

## 2 数字人文视角下的档案研究路径

### 2.1 研究路径与方法的变迁依据

档案研究思路与方法的变迁,根源于档案学理论、对象等学科本体的核心概念及其属性的演化。高层研究思维与底层研究对象的内涵更迭推动着中层研究方法的协同变化,从而以方法为桥梁建立起从理论到实践再回到理论的规范研究程序,图1揭示了这一协同演变的基本机制。



图1 档案研究的思路、方法与实践协同演变机制

20世纪80年代以来,档案学理论创新由后保管主义、社会记忆演化至近年来的档案多元论、社群建档理论,反映出档案管理主体的去中心化与分散化这一主线<sup>[19]</sup>。新晋理论引导档案学者从不同维度认识并挖掘档案价值,这成为了数字人文视角下档案研究的理论基础,同时也凸显了价值导向的档案内容发现与服务方式

转型的必要性。后现代主义倡导“破坏性的创造结合创造性的破坏”,反对宏大叙事结构,关注个人或集群的行为及意识特性。本文提出的路径与方法变革思路重点关注档案内容的实体挖掘、知识化组织以及在参与式服务中的自我呈现,使得关注亚文化或弱势群体记忆的后现代档案研究思潮在方法论层次得以延伸。实践

方面,对于学者而言,档案研究对象与技术不断经受社会意识的反作用,内涵日渐复杂多变,需要重新梳理既有的方法论;对于各级综合性档案馆而言,“十四五”规划下的“加大档案信息资源开发利用力度”的既定目标,亟需以新的技术方法体系为支撑,以档案资源的内容挖掘和参与式服务为核心,形成档案开发与服务的新模式;至于档案行政部门,需要有效挖掘档案价值,促进档案利用<sup>[20]</sup>。这样的顶层设计与总体布局需要对既有的档案开发与利用政策、标准

及模式进行调整迭代与空白填充,更与档案研究的新兴路径及方法不可分割。

## 2.2 档案研究路径的更迭与映射

### 2.2.1 数字人文视角下档案研究要素的更迭与映射

本研究对数字人文视角下档案研究要素的更迭与映射关系进行深入探讨,从学理层面重新审视与界定档案研究的诸要素,并阐明方法与路径变革的实际趋向(见表1)。

表1 数字人文视角下档案研究要素的更迭与映射

视角与条件 映射元素	传统视角	数字人文视角	映射条件
哲学依据	现代主义、历史主义	后现代主义、解构主义	社会存在变化驱动
研究对象	纸质档案	数字化档案、档案数据	科技创新与学者认知
研究重点	档案管理模式	档案内容分析	研究思潮与学者认知
研究视点	记忆技术(存与管)	技术记忆(治与用)	科技创新与学者认知
课题项目	档案实体控制 强调科层管理	关注档案内容 重视分众建构	政策驱动与学者认知
技术体系	分布离散式	创新协同式	技术重组与整合
开发路径	人工编纂与编研	专家、大众及机器 集成式挖掘	科技创新与领域借鉴
利用方案	单一线性的信息查阅	复合多维的价值发现	科技创新与政策驱动
服务模式	固化统一的信息服务	多元异构的知识服务	科技创新与政策驱动

在参考相关研究<sup>[21]</sup>的基础上,本文梳理了包括研究对象、哲学依据、技术体系等在内的9个档案研究核心要素,对其在传统研究范式下与数字人文视角下的内涵与形态分别作了界定,同时考察了研究要素更迭与嬗变的映射条件。研究发现,科技创新与学者认知在大多数要素变迁的过程中起到了关键作用。科学技术,尤其是基于计算机与互联网的信息处理技术不断发生革命性变化,首先作用于研究对象并引起其基本形态的变化,档案学及相关领域学者敏锐地观察到此现象,将其归纳总结并上升为研究对象本体及其属性的变化,经讨论、商榷后逐步成为学术共同体的普遍认识,进而催

动研究对象的中介效应,影响到研究重点、课题、技术与利用等一系列要素的发展变化。

### 2.2.2 基于数字人文与档案资源开发的方法论体系梳理

数字人文视角下档案研究新要素在学科研究中的常态化,仅靠外部技术引进与内部思路创新难以达到,需要在明确档案研究对象及其基本属性的前提下,尝试提出适应新环境的、相对稳定的研究路径与方法体系,借助数字人文的领域张力推进档案学科向新的研究范式过渡。

目前,在数字人文方法论层面,知名数字人文学者 Unsworth 和 Anderson 等分别提出了不同

的观点,包括发现(Discovering)、收集(Collecting)、注释(Annotating)、比较(Comparing)、发布(Delivering)等不同类型的原语,因此也产生了“五原语论”和“七原语论”<sup>[22,23]</sup>。刘炜和叶鹰进一步将“五原语论”中的五个原语分别与具体技术体系相对应,明确其逻辑关系,完善了数字人文研究的基本技术路线<sup>[24]</sup>。在档案资源开发利用的方法体系方面,牛力等创新提出了包括

数字化管护(Preservation)、知识化开发(Discovery)和可视化利用(Utilization)在内的“PDU”模型<sup>[25]</sup>。我们以学者归纳的数字人文研究和数字时代档案资源开发利用的一般方法论体系为基础,通过数字人文原语的映射、再阐释以及研究范围限定,结合档案资源的基本特征,提出数字人文视角下档案研究的方法体系,如图2所示。

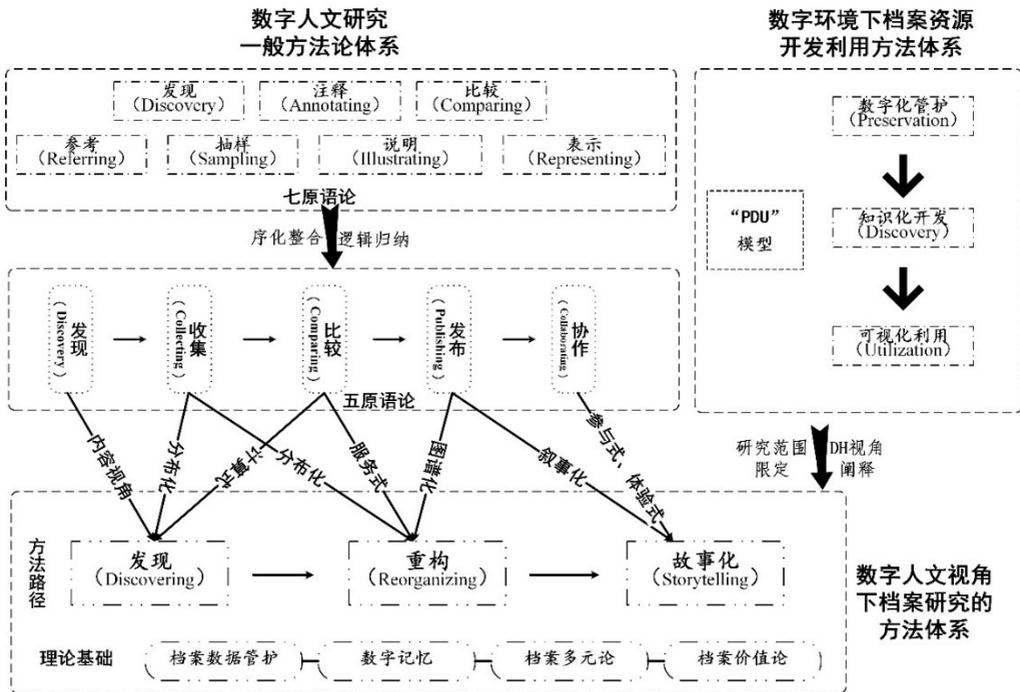


图2 基于数字人文与档案资源开发利用的方法论体系

鉴于档案学理论与方法在数字人文领域的应用尚有不足,需要找到既有的档案研究方法 与数字人文研究方法的对接域,以此为基础梳理 出具有档案学特色的数字人文研究方法体系。“PDU”模型在管护层面注重维系档案资源的 可信性,在开发与利用层面强调引入数字人文 技术对档案资源进行知识组织与可视化服务, 深入挖掘其多维价值,这与数字人文研究方法 的基本目的具有一致性。以档案开发利用环节 为切入点论述数字人文视角下的档案研究,

在理论与技术逻辑上都具有可行性。

数字人文领域一般研究方法的原语化表 达,为本研究提供了概念基础。学者提出的方 法体系旨在从整体上概括一切数字人文研究 的基本图式。相比之下,本研究避免构建宏大 的理论与技术体系,而是结合档案资源开发 利用的基本模式及其理论基础,构建符合档案 数据基本特征及开发需求的方法体系。因此, 我们以数字人文研究方法的“五原语”为基础, 将五个原语分散映射为发现(Discovering)、重 构(Re-

organizing)和故事化(Storytelling)三个原语,从数字人文视角,体现以档案内容为核心、以档案真实性为基础、呈现档案知识并挖掘档案价值

的研究路径,同时通过对以上三个原语的划分与再阐释,梳理出数字人文视角下档案研究的基本模式(见图3),并衍生出相应的技术体系。

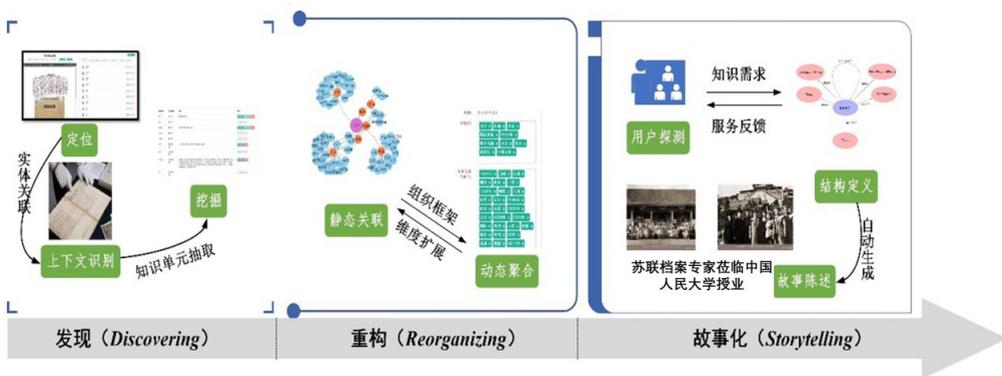


图3 数字人文视角下档案研究路径的基本模式

## 2.3 数字人文环境下档案研究路径的基本模式

### 2.3.1 发现:档案数据的细粒度治理

数字人文视角下档案研究的重点进一步向档案内容倾斜,全宗级、案卷级的研究粒度让位于档案单件与知识单元。细粒度的档案研究方法给予了档案内部诸要素自我呈现的契机,形成了比既往更为直接的档案解构模式。此时,“档案数据”概念的关注点也从档案单件的数据化保存与呈现转向记忆单元的组织与关联。

(1)定位。该过程意味着避免过分强调政治、种族、宗教与其他亚文化偏见,让档案数据中的各类命名实体(Named Entity)毫无保留地呈现在研究者和用户面前,同时建立起“实体—文本”或“实体—照片”一类的来源关系。传统视角下的历史学、档案学研究中,档案内部各类实体往往扮演沉默者与待筛选者的角色,由历史学家、档案学家依个人长期训练形成的经验、价值观、道德观与方法论,从中择取实体进行编纂与研究,在这一过程中,人为地添加实体间语义关系并预先确定叙事结论的情况时有发生。数字人文视角的最大价值在于它突出了研究者本人及其所研究档案的空间感与客观性,能够

借助相应技术对档案数据中蕴含的“记忆实体”进行锚定,在一定程度上规避了主观选择问题。

(2)上下文识别。在档案数据中,经过定位处理的记忆实体往往扮演着主语、宾语、表语等语法角色,镶嵌在特定的档案语篇或数据集中。实际情况下,档案数据中蕴含的记忆实体却不是单独存在的,而是与同语段、同文本或同案卷内的其他记忆实体紧密关联,相关系数又因研究视角与维度的变化而略有差异。因此,需要在一定程度上隐去其语法特性的前提下,将特定实体置于上下文环境中进行识别,如此才能加深对实体及其所处语义环境的理解,完善其作为记忆节点的基本属性。这样,通过数字人文方法来阐释档案数据,档案将更多地被理解作为一种活动空间,而非对确定性和缺失的记载<sup>[26]</sup>。

(3)挖掘。经过定位及上下文识别之后的记忆实体,能够形成基于特定上下文场景的档案记忆单元,反映着特定组织方式下由实体串联成的记忆片段。这种记忆单元的体量与维度并非固定,可能存在于同一个语句中,也可能是多个语段、篇章的集合。相同点在于,记忆单元普遍包含着记忆点、记忆对象、记忆场景等构成

记忆的实体要素,形成了各类记忆元素之间的弱关联与预聚类,此时需要将记忆点连同上下文一并挖掘、存储起来,在隐去偏见的基础上以档案资源内容为核心完成档案数据的“发现”过程,保留档案数据中“自我”(主流文化记忆中的实体指称)与“他者”(非主流文化记忆中的实体指称)的发言权。

### 2.3.2 重构:档案数据的多维度组织

本体、语义知识图谱等知识组织体系能够深入图书、档案等文献资料内部,链接同构或异构的知识单元,以形成特定主题下的知识网络<sup>[27]</sup>。然而,二维网状的知识组织形式无法形成档案数据完整性、可信性等基本属性的多维互证,难以从不同角度揭示档案资源特征。因此,数字人文视角下的“档案重构”需要采用静态关联与动态聚合相结合的多维组织模式。

(1)静态关联。采用领域本体等知识组织模型对档案数据进行描述、组织与规范,由此形成特定框架下可经过各类异构数据平台交换、映射与互操作的关联数据,是档案数据静态关联的主要方式。经过“发现”环节所得到的档案数据,仍然是相对孤立的记忆实体,需要通过领域本体进行关联化描述与结构化组织。档案数据中蕴含的人物、事件、时间、地点、物理实体以及相关的文献源等均应被看作有呈现意义的记忆实体,通过本体的对象属性以三元组形式建立实体之间的语义关系。由于领域本体的类目及属性往往由领域专家构建,一经确定其结构与应用方式便较为稳定,同时整个框架又遵循OWL语言的限定与规范,不易产生颠覆性变化。因此,我们将档案数据经由领域本体组织后产生的关系结构称为“静态关联”,以表现档案资源从“档案数据”向“档案关联数据”的跃迁,这一变化使得档案数据的非结构化存储、开放式组织与关联化发布成为可能。

(2)动态聚合。相比依赖于本体模型构建的档案数据静态关联,动态聚合更侧重于从不同维度揭示档案资源的本质属性,凸显数字人文视角下档案数据研究与其他类型数据研究的

差异。本文所指档案数据“动态聚合”框架同样需要本体模型的介入,不同的是该框架需要采取自顶向下与自底向上相结合的构建方式。具体而言,首先需要自顶向下的知识模型构建,通过建立“维度型模型”(Dimension model)形成多维度的知识组织结构。此处的“维度型模型”主要指与记忆实体相关的多维度分类体系,将描述某一实体的离散的、定性的属性进行再组织。如“时间”概念即可以从公元纪年、封建朝代或其他历法等不同视角进行理解,具有相异的属性表达,而封建朝代的实例即是一些离散的、定性的字符值。其次需要自底向上的实例抽取与填充过程,即将已“发现”的记忆实体在不同维度与视角下的不同属性进行抽取并填充进某一概念直接关联的“维度集合”,这些维度的概念实例可优先采用来自不同领域的关联数据集、叙词表、主题词表或分类法<sup>[28]</sup>等规范性指称语。最终形成不同维度下的实体语义标签或实体画像,以此类画像标签为中心产生的多维核聚效应即是档案数据组织中的“动态聚合”过程,这一过程对档案数据凭证性、完整性、可信性等属性的表征具有重要意义。

### 2.3.3 故事化:档案数据的叙事化表现

相比“发现”与“重构”环节,档案数据的“故事化”属知识服务范畴,即将经过多维组织的档案数据梳理成册并映射至多元可视化空间,以主题故事的形式对档案数据中蕴含的记忆进行叙事化表征。这一过程具体包括“用户探测”“结构定义”与“故事陈述”。

(1)用户探测。传统数据或知识故事化可分为创作者驱动与受众驱动两种模式<sup>[29]</sup>。对于档案数据,尤其是目前已解密可供开发的档案数据而言,创作者与受众两种身份正在以一种渐进的方式产生融合关系。在后现代的档案学理念中,档案本身即由社会大众产生,代表着从个人记忆聚合为群体记忆,最终上升并凝练为国家记忆有机组成部分的全部历程。在一定程度上,档案用户既是档案数据的生产者,也是其管理者、组织者、传播者与消费者,用户在整个

档案数据的故事化过程中扮演着关键角色。目前,国内外档案学界在档案用户的研究视角层面稍显宏观,对用户意向、行为、心理、体验、情感等要素的探测与挖掘有所不足。因此,档案数据的故事化不仅需要注重数据中直接呈现的“能指”,更需深入挖掘并梳理出其中蕴含的“所指”,同时对档案数据的模板规则、故事主题、讲述形式、组织方案等内容进行进一步界定。

(2)结构定义。故事结构代表了档案数据叙事的基本框架与模式。在实现记忆实体的表

面揭示与深度计算后,需要考虑用户的知识需求与研究设计的叙事方案等主观性因素,为关联化的档案数据提供人文解释。在数据故事化的结构方面,朝乐门总结了 Aristotle 提出的五步叙述结构和成熟度曲线模型,将数据故事结构划分为引言、上升、高潮、下降及结局五个阶段<sup>[30]</sup>。这类结构整体上较为宏观,尚未讨论内嵌的要素组织和排列方式。本文从定域与结构两个视角对档案数据组织可采用的故事化结构进行了归纳(见图4)。

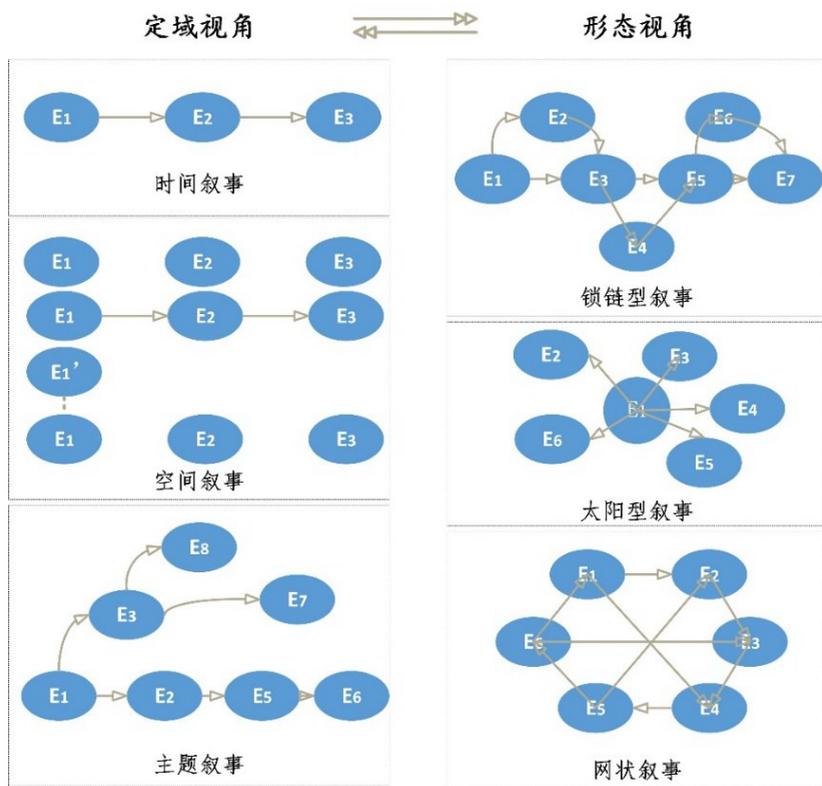


图4 档案数据应用的主流故事结构归纳

定域视角包含时间、空间与主题三种主流叙事结构。其中时间叙事可划分为“开端”“中间”“结尾”三个子事件,适用于情节简单且线性化的档案故事生成;空间叙事则包括从上至下的点结构、线结构以及分层结构,可用于同一空间内不同时间点或时间段的档案数据叙事组

织;主题叙事是一种反线性结构,以复杂事件的内部主题为划分依据梳理事件线,适合对记忆实体较多且实体能被聚类为不同主题的档案数据进行故事化操作。

相应地,在以形态视角划分的故事结构中,锁链型结构的线索与线索之间相互关联,不可

分割,适用于拥有同一条共通的情感主线索且多条线索交织并行的档案数据叙事;太阳型结构则强调事件集群中具有核心事件、人物或受限时空域,对其他事件起到支配作用,适合将档案数据中记载的重要事项或人物单独提取,结合上下文关系进行叙事安排;网状结构中的叙事关系较为松散,可类比于文学作品中的散文,更适用于某一类核心与主题均不够明确,但存在较密集实体关系的档案数据叙事组织,可看作简化的档案知识图谱。

(3) 故事陈述。在择取适当的档案数据故事化结构之后,我们得以触及数字人文视角下档案数据研究的高层次目标——以故事陈述为基础的档案知识服务。这一过程即将档案数据中蕴含的具备自然性(naturalness)、互相关联性(interrelatedness)、独特性(uniqueness)、真实性(authenticity)与客观性(impartiality)的故事进行可视化展示<sup>[31]</sup>,形成以研究者及系统为中介的档案创作者与档案用户间的双向知识通路,此时档案研究者的任务仍然聚焦于如何以多维度的表达形式将知识组织与知识发现之后的档案数据呈现给用户。在这一过程中,与用户直接交互的图形界面(GUI),其价值中立的必要性将

进一步凸显。交互界面中包含的时间性、空间性、触觉性与认知性元素均体现着研究者与开发者对系统前端与底层机制的综合性理解,同时反映着研究者对待档案数据的观念态度与价值取向,进而涉及档案数据中各类记忆实体能否以对等的条件进行表达,是否已经融入研究者想要呈现的知识架构中<sup>[32]</sup>。从用户视角来看,如今数字人文领域琳琅满目的交互界面能够有效塑造用户思维转变的方向和角度,研究者如何启发用户从不同视角进行自我观察与判断,是档案数据以可视化方式进行故事陈述时尤为重要的内容。

### 3 数字人文视角下档案研究的技术方法

本节以数字人文领域知识挖掘、知识组织与知识服务的具体技术为要素,在此前梳理的研究方法论基础上对档案数据研究技术体系进行整合,从而将“发现”“重构”与“故事化”过程一以贯之,阐明数字人文视角下档案数据研究的底层技术依托。下文分别从档案数据的价值保有、价值挖掘及价值实现的视角切入,构建档案数据研究的技术体系,如图5所示。

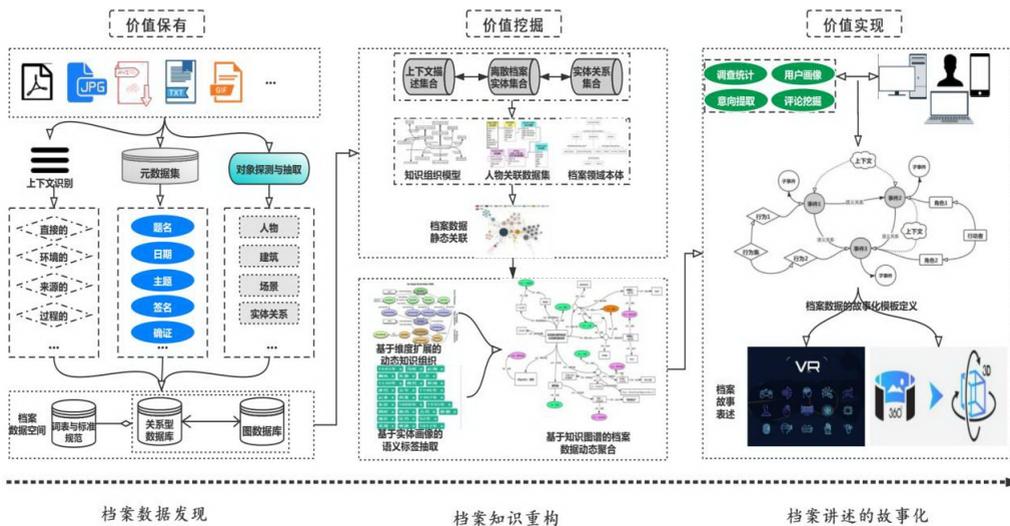


图5 数字人文视角下档案数据研究的技术体系

### 3.1 价值保有视角下的档案数据处理技术

保值视角下的档案数据处理技术用于支撑档案数据研究方法论中的“发现”环节。该环节在档案数据要素的提取方面分为“元数据标注”“对象探测与抽取”与“上下文识别”三个主要部分。“元数据标注”层面,应考虑在传统的档案元数据类目体系中嵌入“签名”“确证”“格式永久性”等电子文件元数据新标准<sup>[33]</sup>,使得数字人文视角下数字档案向档案数据转化时,在粒度分割与实体重组的过程中仍保有其来源关系及证据特性,并将其作为相关档案资源的属性进行处理。

“对象探测与抽取”部分重在应用基于深度学习的自然语言处理及图像识别框架,从规范文本、图像及视频中抽取人物、建筑、时间等关键实体,以客观算法结合人工识别的形式将档案资源中的实体及其相关关系统一提取,在避免主观偏见与客观技术缺陷的基础上补充实体间的语义联系。

“上下文识别”强调对档案创作者、用户、要素结构、档案功能及活动、业务场景与机构职能等要素的识别与关联,档案资源与其他信息资源的本质区别之一即在于保留上下文<sup>[34]</sup>。目前,档案资源的上下文识别过程仍依赖于人文研究的基本方法与技能,这要求档案学者在占有档案资料且已知对象抽取结果的前提下,结合数字方法带来的分析结果,将档案上下文内容补充并保存进相应数据库,重在营造与档案或档案联(Archives Bond)本身相关的事件与空间环境,从而维系解构之前档案的固有价值。

### 3.2 价值挖掘视角下的档案数据组织技术

价值挖掘视角与档案数据研究方法论中的“重构”环节紧密对应,旨在通过多维度的知识组织模型与技术对档案数据进行描述、串联与聚合,形成数字人文视角下动态的档案语义知识图谱,在此前凭证价值、参考价值的基础上,进一步挖掘其存史价值与文化价值,形成明确

的档案数据“增值”路径。经“发现”环节所得的人物、时间、地点等记忆实体、实体语义关系及上下文内容被弱关联化地存储在不同类型的数据库中,并统一链接到档案数据原文,这些要素在原文中被锚定的位置与排列关系仍有记录。

如前所述,档案数据组织分为“静态关联”和“动态聚合”两个步骤。“静态关联”通过本体模型对抽取所得实例及其语义关系进行规范组织,形成档案数据知识图谱的基本框架并导入图数据库存储,建立融合知识图谱技术的档案知识语义检索与问答机制。“动态聚合”则侧重于底层资源中实体的主题聚类与划分,通过知识计算得出实体间的语义及上下文相似度,将围绕同一实体的描述性概念予以聚集。与此同时,借助如“事件本体”(Event Ontology)或“简单事件模型”(Simple Event Model, SEM)等内部带有“实体类型”(sem:Entity Type)概念体系的动态知识组织模型(如“事件”(sem:Event)概念具有“事件类型”(sem:Event Type)),对实体的认知维度进行扩展,从多个侧面切入,并利用规范分类体系对概念实例进行多维度组织<sup>[27]</sup>。

本文以 SEM 模型为基础,以前苏联档案学家谢列兹涅夫与中国人民大学档案系创建者吴宝康间的关系为例,构建基于多维认知的知识组织模型。在图 6 中,“谢列兹涅夫”与“吴宝康”两个实例节点间的语义关系“师生关系”以空节点形式出现在知识图谱中,充当事件中的一个特定角色,其角色类型由《中国图书馆分类法》进行规范,规范概念采用其中编码为“G456”的“师生关系”。而谢列兹涅夫与吴宝康之间的“师生关系”并非永久不变的,而是发生于“1952—1955 年”这个时间段内,因此通过“sem:时间限定”概念对此关系合理的时间段进行描述,这样就形成了与档案数据实例密切相关的高维知识表征空间,同样也可以看作是与实例有关的离散上下文集合,这为此后的多视角档案数据叙事建立了基础。

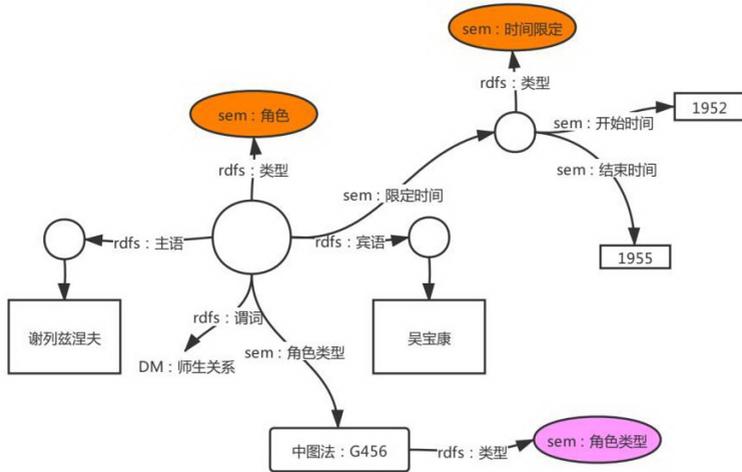


图6 多维认知视角下的档案数据动态组织框架

### 3.3 价值实现视角下的档案数据表达技术

叙事法是历史学、文学等学科常用的资料组织与知识呈现方法。研究者通常在占有有一定数量且相互关联的档案资料的前提下,通过个人训练习得的“映射法则”以陈述形式将事件要素进行逻辑融合,最后完成“档案故事”的原始梳理与再造过程。在数字人文视角下的档案数据研究中,经过价值挖掘所得的档案关联数据及其上下文信息,需要在融合用户意愿的基础上以多维视域进行价值表达与知识服务。因此该环节的首要任务即在于主导者的视角切换,应梳理出用户在档案数据利用方面相关行为与意识的理论框架,结合多维用户画像形成全面的档案用户意向知识集,进而由研究者设计并形成特定的档案数据故事模板,确定不同档案数据集故事化的特异性视角。

在档案数据故事化的呈现层面,数字人文技术的广泛介入能够有效补充传统的档案编纂、主题展览等单一静态方式。本文提到的档案数据表达技术以VR、全息投影等数据可视化与情景仿真技术为依托,以图谱叙事(Graphical Storytelling)结合视觉叙事(Visual Storytelling)为基础,以档案数据流或图像流合并其上下文特征作为输入要素,以此前构建的知识图谱为辅

助的档案数据自动叙事机制<sup>[35]</sup>,形成多维度、沉浸式的档案数据故事环境。不同维度的档案叙事会呈现不同类型的故事表征路径与实体内涵,但每一侧面和角度均以原初的、客观的过程本身为基底进行还原重组,在弥合档案内部实体间自然断层与消减档案本体外部影响的基础上,将史学的态度与文学的形式相结合,能动地讲述档案故事。

## 4 基于数字人文路径与方法的案例研究——以吴宝康档案数据为例

本文以著名档案学家吴宝康先生的档案数据为研究对象,在基于数字人文的档案研究方法论指导下,对档案数据进行分析与可视化,形成以“发现”“重构”及“故事化”为主线的档案数据数字人文研究案例,同时以融合性的知识组织、发现及服务模式为依据,对本研究提出的中观路径与微观方法进行实证检验。

### 4.1 档案数据收集与整理

吴宝康先生是新中国档案教育事业的奠基者,也是中国人民大学档案系(中国人民大学信息资源管理学院前身)的创立者,终其一生为新

中国档案学及档案事业的发展作出了重大贡献<sup>[36]</sup>。在此过程中,吴宝康先生投身革命与兴办教育的足迹也深深拓印在了档案文献之中,形成了一系列珍贵的历史记录。

研究团队以南浔区档案馆、中国人民大学信息资源管理学院老照片网站作为吴宝康档案数据的主要来源,将收集到的全部吴宝康档案数据统一导入团队自行开发的文档知识组织与知识发现平台中,对以照片和文本为主要类型的档案数据进行基础排列与预整理。在数据类型层面,本次实验收集到的档案数据包括图像与文本两类,其中图像 381 张,以吴宝康数字化照片档案为主;文本共 6 卷,主要取材于徐拥军教授编著的《吴宝康学术年谱》。在数据内容层面,图像主要为吴宝康在不同时间、地点的留影,而文字则从叙事角度对吴宝康的成长、革命及学术经历进行了完整而全面的描述与概括。在数据管理层面,南浔区档案馆和人大信管学院老照片网站采用“时间为经,事件为纬”的整理方式对吴宝康档案数据进行了细致且专业的保管与整理,对吴宝康的革命与学术历程进行了阶段划分与主题标注,为本研究的数据组织与开发研究奠定了一定的基础。

## 4.2 档案数据处理与可视化

### 4.2.1 档案数据的分层式“发现”

以方法体系部分的“发现”环节为指导,吴宝康档案数据的处理过程具体划分为“元数据标注”“基于深度学习的对象探测与抽取”以及“上下文识别”三个层次,下面分别从不同角度对档案数据的内容及形式特征进行关联化的锚定与挖掘。

首先,在元数据标注层面,我们在参考国家档案局发布的《照片类电子档案元数据方案》<sup>[37]</sup>(DA/T-54-2014)基础上,考虑到照片档案元数据标注的基本要求和后续研究需要,自行设计了包括了“物理层”“内容层”及“元素层”在内的照片档案元数据标注体系,主要结构如表 2 所示。经过系统化的档案数据标注与整

理,形成了吴宝康档案元数据与原始档案资料的对应陈列结构,建立了从元数据体系到档案单件(Item)之间的粗粒度关联,以满足此后档案数据的批量处理需求。

表 2 吴宝康照片档案元数据标注体系

物理层	内容层	元素层
图像 ID	题名	人物
保管场所	描述	组织机构
图像大小	版权	物理实体
图像格式	拍摄时间	时间
归档时间	地理位置	地点
标注者	照片类型	事件
...	...	...

其次,在档案数据对象的探测与抽取层面,本研究主要应用基于深度学习的图像识别与自然语言处理技术完成这一任务。在图像识别任务中,对吴宝康各个时期的照片进行图像实体探测,对照片中的实体类型进行判定。此部分着重关注照片中吴宝康及其合影人物的识别,对其中的重要人物如吴宝康、谢列兹涅夫等进行再标注。此外,图像中具有叙事意义的代表性物理实体如上海汇丰银行大楼、中国人民大学信息楼等也是图像探测与实体标注任务所关注的内容。

在档案数据自然语言处理方面,我们基于 Bert 模型对《吴宝康学术年谱》各卷逐一进行命名实体识别,重点发掘吴宝康与其他人物、时间、地点、文献等各种类型的命名实体之间的共现状况,以实体间动词为核心抽取语义关系并进行关联式存储。同时以时间节点为依据进行文本与照片内容间的互证,通过文本数据中蕴藏的语义关系丰富照片的语义内涵,形成双模态数据的知识级融合,从而为照片档案的深度解读及上下文识别提供必要支持,如图 7 所示。



图7 《吴宝康学术年谱》的实体识别结果

最后,在吴宝康档案数据的上下文识别层面,先期研究主要采用人工识别方法,分内外两个层次界定照片档案的相关上下文内容。具体而言,团队选派5位具有档案学研究背景、掌握档案学研究方法的博硕士研究生,在参考国际档案理事会档案描述专家组(International Council on Archival Description, ICAEGAD)提出的档案文件上下文本体框架(International Council on Archives Records in Contexts Ontology, ICA RiC-O)<sup>[38]</sup>的基础上,结合此前的元数据标注与实体识别成果,完成照片档案上下文关系的界定与标注。其中,在照

片档案的内部上下文层面,基于 RiC-O 提到的关键实体如对象、人物、时间、地点、事件等内容,我们在元数据与实体识别层面已完成了大部分工作,随后即结合《吴宝康学术年谱》的分析与档案馆实地调研、查阅档案等方式补足每张照片中缺失的具体实例,并对其关系及相应的关系视角进行考证。在外部上下文即照片单件的关系方面,我们在尊重档案整理既有逻辑的前提下<sup>[39]</sup>,分别从时间、主题等不同维度切入,将照片序列中的几张进行上下文聚类,构建特定的叙事框架以形成档案数据故事化的基础结构,过程如图8所示。



内部上下文识别

外部上下文识别

图8 吴宝康照片档案的分层上下文识别

## 4.2.2 档案数据的级联式“重构”

在档案数据“重构”部分,我们参考 SEM 模型设计了包括“实体集”“类型”和“限定”三级概念集合在内的本体模型。其中,在“实体集”层面,将“sem:行动者”概念具体化为“foaf:人物”和“foaf:组织机构”两个概念,新增了“crm:

文献”与“crm:物理实体”概念,用于体现照片或文本提及的论文著作或物理实体,并在 SEM 模型既定对象属性的基础上补充新增概念间的语义关联关系,通过“类型”及“限定”集合对照片档案数据进行维度扩展化的知识组织,体现档案资源的基本特性,其结构见图 9。

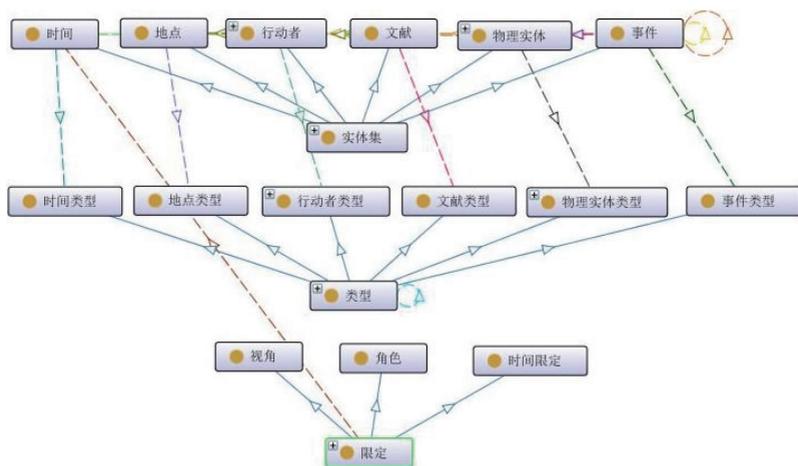


图9 数字人文视角下档案数据的知识组织模型

以本体模型为框架,将所得实例录入系统及图数据库,可得到“静态关联”状态下的吴宝康照片档案数据知识图谱。在此基础上,对《吴宝康学术年谱》进行聚类分析,尤其对吴宝康等核心人物进行实体画像,得到了不同事件或主题中相异视角下的语义标签,再结合《中国图书

分类法》《中国档案主题词表》等规范分类体系或词表对语义标签进行对齐操作,借助开源可视化库 Apache ECharts 生成了具有“动态聚合”效应的档案数据知识图谱,从而以具体事件及人物为核心体现出档案数据的多维知识组织模式,如图 10 所示。



图10 “动态聚合”的吴宝康照片档案数据知识图谱

#### 4.2.3 档案数据的融合式“故事化”

在吴宝康档案数据的“故事化”实现层面,团队自行开发了基于B/S架构的知识服务系统作为用户交互平台,该平台设置了用户意见调查与用户留言模块,同时给予了用户根据不同主题照片的呈现内容及方式自行完成吴宝康照片档案叙事的权限。

在吴宝康照片档案故事化的先期研究中,我们尚未应用基于深度学习模型的全自动化叙事机制,而是秉承后人根据一手史料重构前人记忆的人文研究理念,采用了众包的叙事方法,基于档案数据的对象检测结果、上下文信息及

知识图谱引导用户实现档案故事的多视角讲述<sup>[40]</sup>。在图11中,我们搭建了半自动化的档案数据故事化框架,选取5张20世纪50年代初吴宝康奉调入京创建中国人民大学档案系时期的照片档案,逐一进行对象检测处理,进而以以前构建并存储的人物语义知识图谱为关键上下文信息,为用户营造基本的叙事空间并提供必要元素,同时也完成了对叙事主题和范围的有限界定,以便在遵循基本客观事实的前提下容纳用户的主观思维,对同一历史事实进行多维描述。

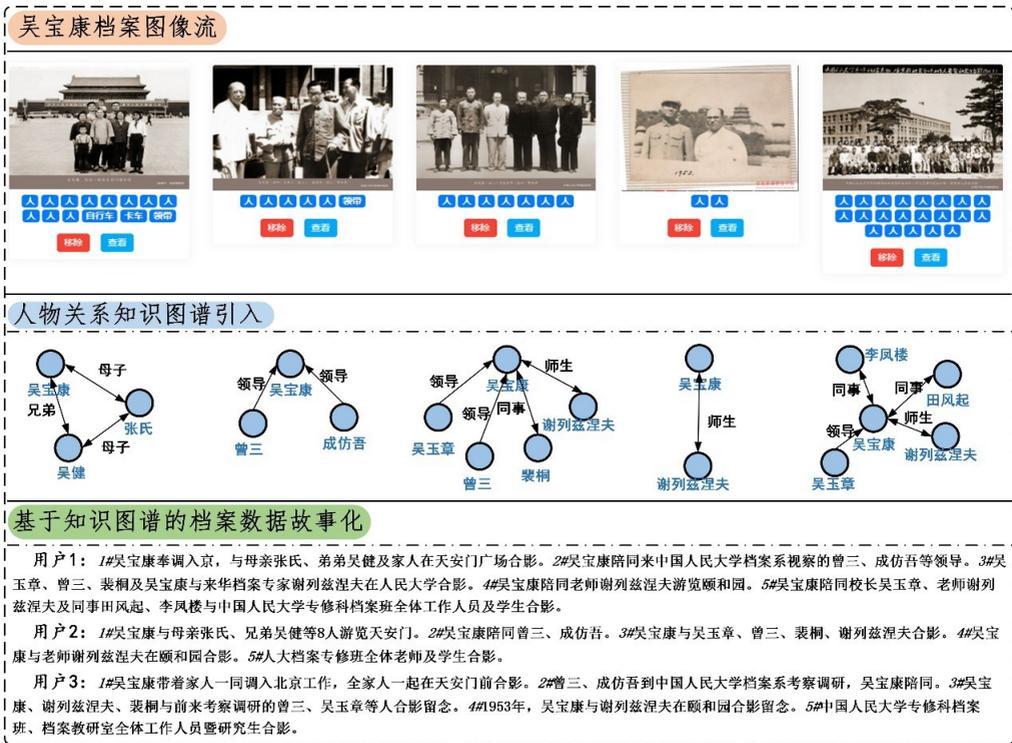


图 11 基于知识图谱的吴宝康照片档案故事化

## 5 结语

本研究分析并界定了档案研究路径与方法在数字人文环境下的变迁依据与要素更迭机

制,在此基础上结合数字人文研究的核心理念将档案研究路径概括为“发现”“重构”与“故事化”三层原语,同时形成了包括“定位”“上下文识别”“动态聚合”等八个子原语在内的档案数字人文研究方法论体系,将档案学研究理念融

入数字人文研究领域。相应地,以“价值保有”“价值挖掘”与“价值实现”为主线梳理出数字人文视角下档案研究的技术体系,以体现数字人文方法的介入对档案价值表现的推动作用。实证部分,我们以吴宝康档案数据为研究对象进行案例研究,综合应用元数据标注、基于深度学习的数据处理、人文方法主导的档案上下文识别、嵌入档案元素的知识组织与动态知识图谱、融合领域知识进行叙事表现等宽领域的数字人文技术方法,再借助知识工程技术体系实现以上方法系统化、平台化的整合与开发,构建了吴宝康档案知识库,生动地展现了以吴宝康为代表的老一辈图情档案学者的崇高信仰与学术品格,体现了吴宝康档案数据的价值内涵,也在一定程度上验证了本文所提研究路径与方法的有效性。

与此同时,本研究也存在一定的提升空间。应用层面,以本研究所提方法论为导向构建的档案知识库在形式与功能上还不够完善,我们将进一步探索动态知识图谱的生成、叙事框架及模板的搭建与档案数据故事化的全自动化实现,以及档案故事的沉浸式体验与利用。理论层面,本文的研究范围限定于档案数据的开发利用与学术研究,提出的原语在数量、系统性与

阐释深度方面尚有不足,数字人文视角下完整的档案研究方法论体系仍处于动态构建之中。

以本文所提研究路径与方法及其背后支撑性的理论与观念为切入点,可提出更多有关数字人文与档案研究的交互性问题。目前,国内外的档案研究逐渐由封闭式材料整合与推断走向开放式的数据关联及开发。数字人文理念及技术与后现代主义思维逻辑相结合,其固有的反思性与批判性使得学科原有的部分研究取向、方法及结果面临修正与重组,给予了档案学研究充分的调整空间与拓展维度,日渐中和此前理论研究中的内卷效应。此外,在学科内部层面,数字人文视角下档案研究新路径与方法的提出,能够连接当前学界关于档案数字人文研究的宏观探讨与底层实践,弥合既有的研究断层;而在同一级学科之间,数字人文环境下的档案学与图书馆学、情报学理应在研究路径及方法层面找到更多的契合点,发现研究对象特征相似性之上的学科研究逻辑与研究旨趣的相似性。

致谢:浙江省湖州市南浔区档案馆为本研究提供了300余张珍贵的吴宝康照片,在此表示衷心的感谢!

## 参考文献

- [1] Webster J W. Digital collaborations: a survey analysis of Digital Humanities partnerships between librarians and other academics [EB/OL]. [2020-05-20]. <http://www.digitalhumanities.org/dhq/vol/13/4/000441/000441.html>.
- [2] 王晓光.“数字人文”的产生、发展与前沿[G]//全国高校社会科学科研管理研究会组.方法创新与哲学社会发展.武汉:武汉大学出版社,2010:207-221.(Wang Xiaoguang. The origin, development and frontier of Digital Humanities[G]//National Institute of Scientific Research Management for Social Science of Universities. Methodological innovation and development of Philosophy and Social Sciences. Wuhan: Wuhan University Press, 2010: 207-221.)
- [3] 夏翠娟.数字人文之热浪潮与冷思考[J].图书情报知识,2019,188(2):2.(Xia Cuijuan. Digital Humanities: hot waves and cool thoughts[J]. Documentation, Information & Knowledge, 2019, 188(2): 2.)
- [4] 《光明日报》理论部.2018年度中国十大学术热点[EB/OL]. [2021-01-15]. <https://news.gmw.cn/2019->

- 01/11/content\_32334236.htm. (Theory department of *Guangming Daily*. Top ten academic hotspots of China in 2018[EB/OL]. [2021-01-15]. [https://news.gmw.cn/2019-01/11/content\\_32334236.htm](https://news.gmw.cn/2019-01/11/content_32334236.htm).)
- [5] 《情报资料工作》编辑部. 2019年度中国图情档案学界十大学术热点[J]. 情报资料工作, 2020, 41(1): 5-12. (Editorial department of *Information and Documentation Services*. Top ten academic hotspots in library, archives & information science of China in 2019[J]. *Information and Documentation Services*, 2020, 41(1): 5-12.)
- [6] 钱毅. 从“数字化”到“数据化”——新技术环境下文件管理若干问题再认识[J]. 档案学通讯, 2018(5): 42-45. (Qian Yi. Digitalization to datalization; several questions on records management under new technology environment[J]. *Archives Science Bulletin*, 2018(5): 42-45.)
- [7] Chen Chih-Ming, Chang Chung. A Chinese ancient book digital humanities research platform to support digital humanities research[J]. *The Electronic Library*, 2019, 37(2): 314-336.
- [8] Hendry R, Burrell A. Playful interfaces to the archive and the embodied experience of data[J]. *Journal of Documentation*, 2020, 76(2): 484-501.
- [9] 崔春, 毕强. 关联数据在数字人文领域中的应用剖析——以关联爵士项目为例[J]. 图书情报工作, 2014, 58(24): 99-105. (Cui Chun, Bi Qiang. Analysis on application of linked open data to Digital Humanities: a case study on the Linked Jazz Project[J]. *Library and Information Service*, 2014, 58(24): 99-105.)
- [10] 张斌, 李子林. 数字人文背景下档案馆发展的新思考[J]. 图书情报知识, 2019(6): 68-76. (Zhang Bin, Li Zilin. New thoughts on the development of archives in the context of Digital Humanities[J]. *Documentation, Information & Knowledge*, 2019(6): 68-76.)
- [11] Mendelsson D, Falk E, Oliver A L. The Albert Einstein archives digitization project: opening hidden treasures[J]. *Library Hi Tech*, 2014, 32(2): 318-335.
- [12] Jarlbrink J, Snickars P. Cultural heritage as digital noise; nineteenth century newspapers in the digital archive[J]. *Journal of Documentation*, 2017, 73(6): 1228-1243.
- [13] Klein L. The image of absence; archival silence, data visualization, and James Hemmings[J]. *American Literature*, 2013, 85(4): 661-688.
- [14] 冯惠玲. 数字时代的记忆风景[N]. 中国档案报, 2015-11-19(003). (Feng Huiling. The memory landscape of digital age[N]. *China Archives News*, 2015-11-19(003).)
- [15] 加小双, 徐拥军. 国内外记忆实践的发展现状及趋势研究[J]. 图书情报知识, 2019(1): 60-66. (Jia Xiaoshuang, Xu Yongjun, Study on the status and trends of memory practice at home and abroad[J]. *Documentation, Information & Knowledge*, 2019(1): 60-66.)
- [16] 曾蕾, 王晓光, 范炜. 图档博领域的智慧数据及其在数字人文研究中的角色[J]. 中国图书馆学报, 2018, 44(1): 17-34. (Zeng Marcia Lei, Wang Xiaoguang, Fan Wei. Smart data from libraries, archives and museums and its role in the digital humanity researches[J]. *Journal of Library Science in China*, 2018, 44(1): 17-34.)
- [17] 加小双. 档案学与数字人文: 档案观的脱节与共生[J]. 图书馆论坛, 2019(5): 10-16. (Jia Xiaoshuang. Archive science and Digital Humanities; disconnection and commensalism in archive concept[J]. *Library Tribune*,

- 2019(5):10-16.)
- [18] 左娜,张卫东. 数字人文视角下的档案学研究[J]. 图书与情报,2019(6):94-100. (Zuo Na,Zhang Weidong. Archival science research from the perspective of Digital Humanities[J]. Library & Information,2019(6):94-100.)
- [19] Tom N. Seeing archives;postmodernism and the changing intellectual place of archives[J]. The American Archivist. 2002,65(1):24-41.
- [20] 丁薛祥同志在中央档案馆国家档案局调研时的讲话[EB/OL]. [2020-05-20]. <http://www.saac.gov.cn/daj/yaow/201805/d7ff928a8ceb40da9d7410fe90cdb522.shtml>. (Speech by Comrade Ding Xuexiang during the investigation of the SAAC [EB/OL]. [2020-05-20]. <http://www.saac.gov.cn/daj/yaow/201805/d7ff928a8ceb40da9d7410fe90cdb522.shtml>.)
- [21] 牛力,刘慧琳,曾静怡,等. 数字时代档案资源开发利用的重新审视[J]. 档案学研究,2019(5):67-71. (Niu Li,Liu Huilin,Zeng Jingyi,et al. Re-examining the development and utilization of archival resources in the digital age[J]. Archives Science Study,2019(5):67-71.)
- [22] Unsworth J. Scholarly primitives; what methods do humanities researchers have in common, and how might our tools reflect this? [J]. Humanities Computer: Formal Methods, Experimental Practice, 2000; 12.
- [23] Anderson S, Blanke T, Dunn S. Methodological commons; arts and humanities e-Science fundamentals[J]. Philosophical Transactions, 2010, 368(1925): 3779.
- [24] 刘炜,叶鹰. 数字人文的技术体系与理论结构探讨[J]. 中国图书馆学报,2017,43(5):32-41. (Liu Keven W, Ye Fred Y. Exploring technical system and theoretical structure of digital humanities[J]. Journal of Library Science in China, 2017, 43(5): 32-41.)
- [25] 牛力,曾静怡,刘丁君. 数字记忆视角下档案创新开发利用“PDU”模型探析[J]. 档案学通讯,2019,5(1):65-72. (Niu Li,Zeng Jingyi,Liu Dingjun. “PDU” model: research on the digital development and utilization of archives from the perspective of digital memory[J]. Archives Science Bulletin, 2019, 5(1): 65-72.)
- [26] Klein L. 缺席的图像:档案沉默、数据视觉化和詹姆斯·赫明斯[J]. 宋迎春,译. 山东社会科学,2018(9):65-77. (Klein L. The image of absence; archival silence, data visualization, and James Hemmings[J]. Song Yingchun, Trans. Shandong Social Sciences, 2018(9): 65-77.)
- [27] 陈涛,刘炜,单蓉蓉,等. 知识图谱在数字人文中的应用研究[J]. 中国图书馆学报,2019,45(6):34-49. (Chen Tao,Liu Keven W,Shan Rongrong,et al. Application of knowledge graph in Digital Humanities[J]. Journal of Library Science in China, 2019, 45(6): 34-49.)
- [28] Hage W R V, Malaisé V, Segers R, et al. Design and use of the Simple Event Model (SEM)[J]. Journal of Web Semantics, 2011, 9(2): 128-136.
- [29] Segel E, Heer J. Narrative visualization; telling stories with data[J]. IEEE Transactions on Visualization and Computer Graphics, 2010, 16(6): 1139-1148.
- [30] 朝乐门,张晨. 数据故事化:从数据感知到数据认知[J]. 中国图书馆学报,2019,45(5):61-78. (Chao Le-men, Zhang Chen. Data storytelling; from data perception to data cognition[J]. Journal of Library Science in Chi-

- na, 2019, 45(5): 61-78. )
- [31] Duranti L, Patricia C F. Encyclopedia of archival science[M]. Lanham, Maryland: Rowman & Littlefield, 2015.
- [32] 贝里, 费格约德. 数字人文: 数字时代的知识与批判[M]. 王晓光, 等, 译. 大连: 东北财经大学出版社, 2019: 158-173. (Berry D M, Fagerjord A. Digital Humanities: knowledge and critique in a digital age[M]. Wang Xiaoguang, et al, Trans. Dalian: Dongbei University of Finance and Economics Press, 2019: 158-173. )
- [33] InterPARES Trust. IP ontologies [EB/OL]. [2020-05-20]. [http://www.interpares.org/ip2/display\\_file.cfm?doc=ip2\\_ontology.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2_ontology.pdf).
- [34] Theimer K. Archives in context and as context[J]. Journal of Digital Humanities, 2012, 1(2): 1-8.
- [35] Hsu Chao-Chun, Chen Zi-Yuan, Hsu Chi-Yang. Knowledge-enriched visual storytelling [EB/OL]. [2020-05-20]. <https://arxiv.org/pdf/1611.01567.pdf>.
- [36] 徐拥军. 吴宝康学术年谱[M]. 北京: 中国人民大学出版社, 2018: 258-260. (Xu Yongjun. The academic chronicle of Wu Baokang[M]. Beijing: China Renmin University Press, 2018: 258-260. )
- [37] 国家档案局. 照片类电子档案元数据方案 (DA/T-54-2014) [EB/OL]. [2020-05-20]. <http://www.jnuca.cn/6contents-15-5730.html>. (SAAC. Metadata standard for digital photographic records (DA/T-54-2014) [EB/OL]. [2020-05-20]. <http://www.jnuca.cn/6contents-15-5730.html>. )
- [38] ICA EGAD. International council on archives records in contexts ontology (ICA RiC-O) version 0.1 [EB/OL]. [2020-05-20]. <https://www.ica.org/standards/RiC/ontology>.
- [39] 宋宁远, 王晓光. 基于情节本体的叙事性文本语义结构化表示方法研究[J]. 中国图书馆学报, 2020, 46(2): 96-113. (Song Ningyuan, Wang Xiaoguang. Semantic structured representation method of narrative text based on plot ontology[J]. Journal of Library Science in China, 2020, 46(2): 96-113. )
- [40] Yang P, Luo F, Chen P, et al. Knowledgeable storyteller: a commonsense-driven generative model for visual storytelling[C]//Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19, 2019: 5356-5362.

牛力 中国人民大学信息资源管理学院副教授, 博士生导师。北京 100872。

高晨翔 中国人民大学信息资源管理学院博士研究生。北京 100872。

张宇锋 浙江省湖州市南浔区档案馆副馆长。浙江湖州 313009。

闫石 武警北京总队执勤第七支队副连职排长。北京 100012。

徐拥军 中国人民大学信息资源管理学院教授, 博士生导师。北京 100872。

(收稿日期: 2020-05-22; 修回日期: 2020-07-19)