

数字人文项目中的档案整理路径探索

贺钰钦

广西民族大学 广西南宁 530006

摘要：本文深入探讨了档案整理在数字人文项目中的重要性和实践路径。数字人文项目作为信息技术与人文学科交叉融合的新领域，其形成的海量数据对档案整理的需求不断增长，因此档案整理的科学性和系统性直接影响项目的质量和深度。文章分析了当前数字人文领域与档案管理领域的交叉融合，结合数字人文项目分析指出了优化项目中涉及到的档案整理逻辑，构建数字人文项目中档案整理路径，并提出数字人文项目中档案整理路径的优化建议。未来，档案整理的发展将更注重技术创新和跨领域协作，以服务于数字人文项目的需要。本文的研究成果和建议可为数字人文项目的深入研究和实践提供理论与实践的参考。

关键词：数字人文；档案整理；自由来源原则

1 数字人文与档案整理的交叉点

1.1 数字人文项目中的档案角色重要性

在数字人文产品开发进程中，档案充当着至关重要的信息资源提供者角色。其蕴含的丰富历史文献与资料，构成数字人文研究的基石。经数字化保存，档案资源获取与研究更为便捷，有力推动数字人文项目发展，提升项目可靠性与权威性。同时，档案对数字人文有别于其他实践活动，数字人文项目多借助档案资源开展，档案管理思维能对数字资源进行系统规范化组织，在资源管理、成果维护、合作关系、风险评估等数字人文项目全过程管理方面发挥关键作用。此外，数字人文项目需处理海量文本、图像、音频、视频数据，档案作为关键数据来源，整理质量直接关乎数字人文项目的深度与质量。

1.2 数字人文项目中档案整理理论可行性

在数字人文背景下，图情档领域基础理论发展研究主要集中于学科发展与理论构建。有学者认为诸多档案学经典理论需进一步分析或重新定义，如马仁杰提出要突破原有理论局限以建立适应性更强的新理论，然而笔者认为档案学经典理论仍具普遍性。以自由来源原则为例，其在数字人文项目档案整理模式归纳中的可行性可从多方面考量：一是尊重档案材料的历史背景和原始属性，不随意改变或篡改内容和形式，以此保障档案真实性与准确性；二是保留相关档案材料的原有组织结构，同时进行优化升级以契合现代管理和利用需求；三是注重档案材料价值，挖掘其多层次价值与意义，

整理者需系统分类、标注，并结合文献资料进行数据挖掘和深度解读，提升其应用价值。

2 国内数字人文项目档案整理实践分析

2.1 样本选取

本研究主要通过网络调查法来分析数字人文项目中档案整理的相关情况，对我国现有的数字人文项目中相关档案项目的整理进行分析，结合 2022 年中国数字人文年会优秀项目名单中的 26 个数字人文项目，排除网站平台无法访问、网站数据长时间未更新、网站资料比较偏离档案范畴等情况现选取了我的北京记忆、多维度图像智慧系统等 6 个数字人文项目为研究对象进行分析。

2.2 项目分析

对选择出来的数字人文项目样本网站所呈现出的分类效果进行分析，以自由来源原则中强调的将来源和事由按照比例形成新综合体以实现档案资源的系统化呈现为指导档案分类整理分类为原则。如图 1 对数字人文项目网站中相关档案资料进行整理时的分类分析。

项目名称	来源	事由		
	来源主体	事由时空	事由客体	
我的北京记忆	大众记忆板块：普通公民精选记忆板块：名人 / 机构	地点：北京范围时间：无固定范围	砖瓦之间五行八作	北京故事采风问俗
多维度图像智慧系统	博物馆 / 图书馆 / 私人藏家…	藏品名称藏品作者藏品现所屈机构	藏品中心法书字帖	
盛宣怀档案知识库	盛宣怀及其相关人物资料	地图浏览：盛宣怀相关 – 人物行为轨迹时间轴：盛宣怀相关 – 人物活动时间	关系人物	时空公司
中国家谱知识服务平台	私人 / 图书馆	地点：中华人民共和国境内时间：清代 1691– 至今	家谱文化家规家训迁移图	
宋元学案知识图谱	中华书局 / 中国历代人物传记数据库 / world map	地点：南北宋境内 时间：建隆 960– 祥兴 1278	关系图谱 学术流变师承关系 学派传承	
中国古代历史人物迁徙可视化	中国历代人物传记数据库…	地点：中华人民共和国境内时间：唐朝 – 清朝	生死地迁徙路径城市生死人数差异	

图 1 档案资源整理

经对我国六个数字人文项目网站分析，在档案数据挖掘工作上呈现多维度特点。来源主体维度，各项目均对档案资料来源主体的“来源信息点”挖掘，如“我的北京记忆”标注提供者姓名，宋元学案知识图谱项目标注来源机构并设友情链接。事由时空维度，网站重视时空环境点挖掘，时间上多数项目明确起始时间。事由客体维度，各项目依研究主题挖掘档案内容，经归类、整理等呈现主题，且侧重点不同，“我的北京记忆”与“多维度图像智慧系统”侧重资源内容整理，“盛宣怀档案知识库”与“宋元学案知识图谱”聚焦历史人物关系流派。

2.3 档案整理路径探索

基于前文理论，数字人文项目里档案资源整理，在自由来源原则指引下与数字技术深度融合。为清晰呈现这一过程，归纳前述要点，构建出图 2 所示适用于数字人文项目的通用整理逻辑。

以“将来源和事由按比例形成新综合体”为导向，从“来源信息点”“时空环境点”“主题描述点”挖掘数据。借数字技术完成收集后，串联形成来源主题轴、事由客体轴、事由时空轴。来源主体与事由客体的时空“环境点”结合成基本要素面，便于检索关键信息；来源主体轴和事由客体轴依原则结合，可按需整理档案；事由时空轴与事由客体轴组成事由全貌面，提供完整细节。在数字人文环境下，各元素多维组合成“由面到体”的数据展示层，按“来源”和“事由”需求高效整理信息。这种从点到线、线到面、面到体的逻辑，经挖掘、串联、整合、展示，直观呈现档案整理内涵，为不同层次需求提供档案利用支持。

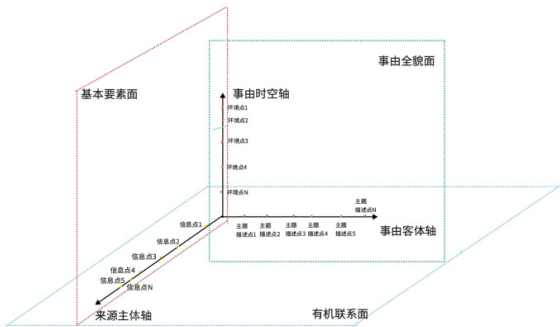


图 2 档案资源整理逻辑归纳

2.3.1 数据挖掘层面的整理逻辑归纳

在数字人文环境下，运用机器学习等技术分析档案数据，“来源信息点”“时空环境点”“主题描述点”是基本数据收集节点，挖掘的数据类型各不相同。“来源信息点”分个人与机构，个人挖掘名字等信息；“时空环境点”含时间与空间，时间点挖掘起始时间等；“主题描述点”挖掘档案形成目的等内容，针对特定主题整理分析，实现档案深度挖掘。

2.3.2 数据串联层面的整理逻辑归纳

来源信息点串联形成来源主题轴，其本质是人际关系网。时空环境点串联形成事由时空轴，其本质是档案记录事件发生的时间线与空间线的时空交织网。主题描述点串联形成事由客体轴，其本质是档案记录事件的主体关系网。围绕着人物与地点等角度的数据挖掘的结果进行整合，人员汇集面与地点汇集面，从而建立起丰富的文本语料库，为进一步探索群体记忆做好数据串联准备工作。

2.3.3 数据整合层面的整理逻辑归纳

在档案数据组织与展示中，整合来源主体轴、事由客

体轴、事由时空轴能形成不同层面。来源主体轴关注事件来源及参与主体，事由客体轴聚焦事件所涉实体与概念，二者整合形成有机联系面，凸显档案对不同主体的价值，如中国家谱知识服务平台“姓氏浏览”将来源与主题描述关联，助力数字人文研究。事由时空轴记录事件时空信息，与事由客体轴融合构成事由全貌面，体现档案凭证价值。来源主体轴和事由时空轴整合的基本要素面，彰显档案参考价值，方便研究者深入分析。

3 档案整理路径的优化策略

数字人文的兴起为档案管理提供了创新视角与技术路径。在此背景下，档案整理作为基础性工作，其方法革新对提升服务质效具有双重价值：既需要完成实体档案的系统化分类，更需通过数字化转换、语义化标注和关联数据构建，形成支持人文研究与教育创新的知识底座。

档案部门在数字人文项目中承担着立体化职能：作为资源中枢，需持续供给优质档案资源；作为技术适配者，应强化数字档案的跨库整合与智能检索能力；作为协同节点，须深度参与项目设计，通过动态反馈机制优化档案内容的知识挖掘维度。实践层面的路径优化可聚焦三个进阶方向：①标准建设（建立元数据规范与质量控制体系，保障数据结构化程度）；②技术赋能（融合OCR、NLP、知识图谱等技术提升数据处理与知识关联效率）；③生态构建（创建分级开放共享平台，激发跨学科知识再生产）。这种深度融合不仅推动档案管理从保管导向转向知识服务范式，更通过激活档案的叙事价值与证据链价值，为人文研究提供多维度的学术支撑，最终实现档案资源从馆舍存储到社会智识生产的价值

跃迁。

4 结语

通过对国内6个数字人文项目的深入实践分析，我们归纳出一套涵盖数据挖掘、串联、整合及展示的系统性整理逻辑。各项目在多维度数据挖掘方面积累的丰富经验，为后续研究提供了极具价值的参考。然而，当前信息技术与人文科学的融合仍面临诸多挑战，诸如档案资源整合与共享程度不足、利用效率有待提升等问题较为突出。针对这些问题，本文提出推进档案资源整合共享、明确档案部门角色定位等优化策略，以期推动二者深度融合。

展望未来，随着技术创新的不断推进以及跨领域协作的持续深入，档案整理在数字人文项目中将继续发挥关键作用，助力知识挖掘与文化遗产，推动学术研究与社会发展迈向新高度。期待各方携手努力，进一步完善档案整理路径，实现信息技术与人文科学的协同共进，开创数字人文领域发展的新局面。

参考文献：

- [1] 任红娟,秦勇豪,宋京京,等.档案领域数字人文研究热点主题发现[J].档案管理,2024,(06):5-9.
- [2] 张新芳.数字人文视角下的档案数据化与国家记忆保护和传承[J].档案管理,2024,(01):69-72.
- [3] 马岩龙,胡可杨,朱佳焯,等.数字人文视域下新疆红色文化资源开发与利用[J].兰台世界,2024,(04):125-129.

作者简介：

贺钰钦(1999—),女,汉族,四川省达州市,硕士研究生,研究方向为档案管理。