

DOI:10.16104/j.issn.1673-1883.2024.02.004

# 彝文历史档案数字化开发研究

高建辉

**摘要:** 彝文历史档案是彝族历史文化遗产的精华,利用数字人文的理念和技术进行开发利用,可以有效保护和传承彝族历史文化。在分析彝文历史档案数字化开发现状及现实困境的基础上,结合数字人文与彝文历史档案数字化保护的逻辑关联,探讨数字人文视域下彝文历史档案数字化开发的思路以及具体实现路径,以期指导当前彝文历史档案数字化开发实践工作。

**关键词:** 数字人文; 彝文历史档案; 数字化开发

**中图分类号:** G270.7 **文献标志码:** A **文章编号:** 1673-1883(2024)02-0028-06

收稿日期: 2023-10-16

作者简介: 高建辉(1982—),男,云南楚雄人,楚雄师范学院档案馆副研究馆员,博士,硕士生导师,研究方向:图书馆管理,民族档案研究,文献数字化与数据库建设, E-mail: gao1026@126.com。

彝文历史档案是指我国各个历史时期的国家机构、社会组织和个人在社会发展进程中直接形成的反映彝族政治、经济、历史、军事、科技、文艺、伦理、宗教和民俗等情况,具有保存价值的彝文历史记录<sup>[1]</sup>,其类型包括彝文古籍、彝文文书和彝文金石铭文等。它是珍贵的民族文化遗产,具有丰富的历史文化价值,是研究彝族历史文化的第一手资料。2021年中共中央办公厅、国务院办公厅印发的《“十四五”全国档案事业发展规划》<sup>[2]</sup>中明确指出档案工作要提升档案服务能力,积极探索数字人文等技术在档案信息深层加工和利用中的应用。数字人文作为一种全新的研究范式,其跨界融合的理念使得其可以将数字技术与人文社科完美结合,改变当前彝文历史档案开发各自为政所导致的重复开发与资源浪费的混乱局面,更新彝文历史档案的开发理念,优化彝文历史档案的开发流程,挖掘彝文历史档案的双元价值。总之,在数字人文视域下进行彝文历史档案数字化开发势在必行。

## 一、彝文历史档案数字化开发现状及现实困境

### (一) 彝文历史档案数字化开发现状

当前彝文历史档案在数字化开发方面已经取得了一些成绩,主要形式是数据库平台建设和网站建设。

如楚雄州图书馆于2016年建设完成彝族文献数据库,分为彝族毕摩经典译资源库、彝族文献资源库、楚雄州地方文献数据库等子库,其中包含了大量彝族文献、毕摩经典和彝族地方文献等近160万条数字信息;楚雄师范学院图书馆2020年建设完成的彝文古籍数据库包含数百条彝文古籍目录信息和100多卷彝文古籍全文文献。目前西南民族大学、贵州民族大学、红河学院、西昌学院等彝文历史档案保管和研究机构也都在积极谋划,主动寻求资金和项目支持,开展了具有自身特色的彝族文献数字化开发和数据库建设工作。在网站的开发上,主要有彝学研究网、彝族人网、彝文化网、彝族毕摩网等,其中收录了大量的彝文古籍档案,包括彝语彝文教材、民风民俗文摘、彝医彝药文献、滇南古彝文文献等。虽然在数据库和网站建设方面取得了一定的成绩,但彝文历史档案数字化开发还面临着诸多困境,如数字化资源数量体量较小,开发的深度和广度明显不足,开发的政策和理念有待进一步完善,开发的机制尚未形成等。

## (二) 彝文历史档案数字化开发现实困境

### 1. 缺乏多方协作的开发机制

彝文历史档案除具有档案的原始属性外,还具有图书、古籍、文物和史料等多元属性,同时被为民委、档案馆、图书馆、博物馆、研究院等机构所收藏<sup>[9]</sup>。据笔者调研发现,目前的国内彝文历史档案开发工作大多由各保管部门独立完成,以自身的馆藏资源为开发基础,进行专题开发和数据库建设,过于依赖自身资源,缺乏与其他单位的协作开发与资源共享。彝文历史档案数字化是开发工作的前提,仅仅是由单个部门的资源数字化难以满足数字人文开发的海量资源要求。在档案数字化与数据化的处理过程中,只有各开发主体的标准一致才可实现数据的共享互通。目前缺乏统一性和全局性规划、彝文历史档案资源分散、各部门各自为政且欠缺协作机制,极易造成重复性开发和开发不完整的问题,也会导致人力物力等资源的浪费。

### 2. 缺少开发人才与技术支持

国内各彝文历史档案保存机构虽然在以往档案开发的基础上积累了部分经验,但存在档案开发人员数量不足、专业背景单一、技术能力有限的问题,工作人员的综合素质还达不到数字人文要求下深入开发彝文历史档案的要求。且彝文历史档案涵盖丰富,涉及历史地理、人文艺术、宗教礼仪等方面,想要对其进行深入开发不仅需要专业的档案知识,同时还要熟悉历史学、地理学、民族学、计算机等多种学科的知识。在数字时代,为满足社会大众多样化的档案利用需求,还需要掌握数字人文技术,如数据库技术、虚拟现实技术、三维建模、GIS等。而许多开发人员对于上述技术和知识的掌握还较为缺乏,开发深度和成果呈现较为有限,开发形式内容也较为单一。

### 3. 开发理念与政策有待完善

开发服务理念滞后主要体现在两个方面:一方面是彝文历史档案开发主体自身固有的落后认知,对于开发利用的价值体现认识不到位,对于数字技术的应用也偏向于基础层次,整体的开发利用工作存在被动开发的状态,主动参与感不强;另一方面是从彝文历史档案的需求者角度来说,目前彝文历史档案的开发工作公众参与度不高、参与意识不强,存在“重藏轻用”现象,开发工作比较保守。其次,在开发政策法规上也存在不够完善的问题。彝文历史档案的开发是一项长期且系统性的工作,在开发过程中如果没有科学合理的顶层设计和政策法规作为保障,势必影响到开发工作的顺利进行。具体表现为:

国家层面上,缺少针对档案开发的专项法律法规,尤其是在资源共享、档案信息发掘、公众参与开发等方面。在地方层面上,缺少专项政策保障和相关部门的统筹规划,制约了彝文历史档案开发工作的推进。

#### 4. 开发形式与内容创新不足

现阶段的彝文历史档案开发还无法实现深层次的内容挖掘与智慧化利用,也没有形成一套完整成熟的技术性开发体系,针对数字人文技术的整合工作开展得并不完善,无法应对彝文历史档案资源整合与系统开发的实际需要,也就难以实现档案开发的形式与内容创新。在彝文历史档案阅览与展览开发上,丰富的影像与音视频资源还没有及时适配,彝文历史档案的成果展现形式也较为单一。目前针对彝文历史档案的开发更多是集中在编研工作上,出版的开发成果也都是以档案资料汇编为主,缺少对彝文历史档案文化内涵的深入发掘,利用数字人文的理念与技术来指导开发工作,深入挖掘彝文历史档案的社会经济文化价值,从而更好地服务社会和公众。

### 三、数字人文视域下彝文历史档案开发思路

#### (一) 科学构建协同开发机制

彝文历史档案的开发是一项长期的系统性工程,内容涉及历史学、档案学、艺术学等多个学科领域,开发主体包括档案馆、博物馆、研究院、图书馆等诸多机构,加之彝文历史档案保管和收集的特殊性,需要科学构建协同开发机制,鼓励各保存机构之间开展跨界联合,发挥各自的优势积极参与<sup>[4]</sup>。鉴于档案机构在档案开发中具有丰富的开发实践经验,在数据库、档案信息化建设方面有完善的标准与经验可以直接利用,可由档案机构作为开发的主导机构,承担开发工作的主要职责。主导机构在做好本职工作的同时应紧跟数字时代的发展趋势,熟练掌握当前的档案开发政策和技术标准,主动积极协调其他开发机构,共同打造彝文历史档案的协同开发体系,保证开发工作的持续推进。其他参与机构,如图书馆、研究院、彝族文化传承人、数字人文机构在彝族历史档案的开发中应积极响应主导部门的安排,按照预先制定的标准规范来开展彝文历史档案的开发工作,充分发挥各自优势,持续推进开发工作。此外,还可以引入专业的数字人文机构,采取技术外包的形式,减少开发的技术难度,不仅可以保证项目顺利进行,还可以确保专业开发质量,实现彝文历史档案的有效开发。

#### (二) 培养新型专业开发人才

将数字人文引入到彝文历史档案的开发工作中来,需要培养一批具有科技素养和人文情怀的专业开发人才。培养数字人文素养的档案专业开发人才需要从多维度进行。一方面,可以依托高校培养高级技术专业人才。彝文历史档案属于少数民族档案范畴,培养其所需的专业人才主要基地还要依靠高等院校档案学专业,可以联合档案管理机构建设档案专业高等院校实训基地,开设数字人文实践课程,联合培养专业人才。另一方面,可以建立多元化的继续教育培养体系,鼓励开发人员积极参与,提升数字人文素养,掌握数字人文理念与技术,提升自身综合素质;在培养内容上,开发主体可以通过对现有的数字人文项目进行深度分析,提取其中工作方法和成功经验,与当地高校和数字人文研究机构合作,通过统一学习、定期培训等方式促进产学研共同发展,以此更好地推动彝文历史档案的开发工作。

#### (三) 转变工作思想理念

数字人文视域下彝文历史档案的开发还需要引导及转变开发者的工作理念,从被动安排到主动开发,

寻求彝文历史档案开发方式的创新运用。随着数字技术与传统人文学科的融合研究不断深入,彝文历史档案的开发者应转变原先固有的传统理念,积极利用新技术主动开发馆藏的彝文历史档案。彝文历史档案不能只是要简单的征集保存,要在保护的基础上主动开发,积极利用数字化技术,加大彝文历史档案的资源开发力度,按照不同的彝文历史档案类型针对性地选择适宜的开发方式、应用技术、展示形式等,使彝文历史档案充分发挥其内在的深层价值,符合当前的时代需要。

#### (四) 引入先进数字技术

数字人文视域下的数字化开发并非对彝文历史档案进行简单的拍照扫描存档,而是在数字化的基础上,深度挖掘其内涵价值,实现有效开发,为用户提供更多元的知识服务。一是可以利用关联数据的知识组织功能,串联起凌乱的知识信息,提供概念检索服务。二是利用文本挖掘技术实现档案内容的深度编研,挖掘和提炼,整合档案内的潜在信息,构建特色专题知识库,优化档案开发的成果。三是利用GIS技术来实现档案信息的时空重组,对数量庞大分散分布的彝文历史档案进行空间集成和地图展示,构建可视化地图,优化利用体验等等。

#### (五) 完善法律法规与标准体系

首先,政府相关部门应根据现阶段彝文历史档案保护与开发现状和困难,融入数字人文的开发理念,制定具体详尽、便于实施的法律法规、工作方法和实施细则,填补彝文历史档案开发方面的法律缺失。其次,完善统一的数字标准是开发彝文历史档案的关键环节,为确保各彝文历史档案开发主体在进行数字化操作时有规可循、有例可依,更好地实现彝文历史档案信息资源的组织、整合、开发、利用。彝文历史档案开发应以政府机构主导,借助行政手段来规范各主体彝文历史档案信息资源的数字标准。如借鉴现有的《档案分类标引规则》《电子档案著录规章》《缩微影像数字化转换技术规范》《全国革命历史档案数据采集标准》等标准,同时参考其他少数民族历史档案数字信息资源的检索方法、存储格式等,不断完善彝文历史档案开发的元数据标准、网络基础设施标准、文档内容格式标准、文档内容表现标准、文档信息检索标准、文件存储标准,对不同载体形式彝文历史档案进行统一编码和格式转换,从而达到数字人文环境下的资源深度整合要求。

### 四、数字人文视域下彝文历史档案数字化开发的实现路径

当前数字人文视域下档案开发技术研究主要是针对某一流程或者环节,如陈涛等<sup>[5]</sup>利用关联数据构建方志档案的知识图谱,位通等<sup>[6]</sup>利用GIS技术开发年谱档案的可视化平台,但很少有针对性针对档案开发全流程的技术路径研究。笔者结合当前的彝文历史档案开发情况与数字人文技术研究,将数字人文视域下彝文历史档案开发技术路径划分为数字化、文本化、知识化和图谱化4个阶段。

#### (一) 彝文历史档案文献数字化

目前收藏的彝文历史档案大多为实体档案,受外部环境和内部环境的变化难以永久地留存下去。通过图像扫描设备,将彝文古籍、金石铭文、彝文文书等彝文历史档案转化为数字形式存储,实现电子阅读,避免在开发过程对档案本身造成二次损毁,同时也方便后续对其加工、标引、挖掘和利用等。在对彝文历史档案进行图像扫描中,依据档案特性选择不同的图像扫描设备。以古籍档案为例,若经卷破损严重,在进行图像信息采集时必须注意保护经卷的安全,采用专用书刊扫描仪,图像分辨率不低于

400dpi。若毕摩经卷保存状态较好,可以轻微接触,则尽量进行高精度信息采集,采用彩色扫描仪,将输出分辨率和光学分辨率调到最高。此外针对碑刻等特殊载体的彝文历史档案可以采用数码相机对碑刻的前后左右等方向进行全方位的拍摄,获取该实物的一组图片集,也可以利用3D扫描设备进行扫描以得到该实物的三维立体模型,光学分辨率不低于600×1200dpi。

## (二) 彝文历史档案图像文本化

经过图像扫描后的彝文历史档案已经极大地降低了利用的难度,可提供在线阅读功能,满足一定的服务需求,但是还存在不足,比如无法实现文本的检索、定位等功能。为了实现数字人文理念下彝文历史档案资源的深度开发,还需要进行文本化处理,对扫描后的彝文历史档案图像中的文字进行识别,转化为机器可读的全文本数据。彝文文字的识别不同于汉文,彝文字体字形变化较多,且无统一标准,识别难度较大。西南大学陈善雄教授依托于贵州工程应用技术学院人工收集整理于《西南彝志》的古彝文字体,提出了一种古彝文手写体的自动采样方法并邀请彝族学生进行临摹获取到大量的手写体古彝文字符,汇总为古彝文数据集(HWAYI),彝文手写体的识别已有较高的精准度。上海大学的古彝文专家们探索利用AI技术来实现古彝文原籍数字化,设计出四字节编码系统,在对拓片、竹简等非常规的载体彝文文字识别方面已取得重大突破,目前已经建立起了精确的彝文古籍电子数据库。此外,基于词典和语法规则的方法和基于统计机器学习的方法在古彝文字的识别方面都有较大优势。在OCR识别技术中,当前市场上也已经出现了较为成熟的彝文识别技术,如“AI+OCR”融合下的智能文字识别技术,在解决古籍版式检测、图像处理和文字识别难题上取得了突破性进展,识别准确率已经能基本满足大规模彝文历史档案文本实际应用的需求。

## (三) 彝文历史档案文本知识化

数字化、文本化的彝文历史档案已经能够极大地促进档案的广泛传播、阅读和开展利用。但是对于读者来说,只是阅读的载体发生了改变,实质上还是一种提供全文本式的简单利用方式。而真正要想能够提高档案资源的利用效率,则需要将彝文历史档案中的隐性知识显性化,从而使其发挥最大效用。“文本知识化”正是其中关键的一环,即利用数字人文和关联数据技术对彝文历史档案文本数据中的知识单元进行组织,将文献的利用方式从“只读”拓展为“可分析”,重组文献的内容,建立语义关联,便于文本检索、知识关联,构建知识子库等。从而解决历史档案语境不适用于当前现代叙事的问题,实现历史档案在不同时空的价值转换。

## (四) 彝文历史档案知识图谱化

图谱化是知识化的最终阶段,是指将知识化后的彝文历史档案与数字地图结合,将档案内容的时间元素和空间分布数据一体化展现,实现碎片化档案信息的系统整合,增强用户的历史空间认知感。同时,借助数据库系统实现统计分析、知识地图展示等功能,使彝文历史档案重新焕发时代活力。而在图谱化的这一过程中,主要应用到的技术为虚拟现实技术、GIS技术等。如上海图书馆与上海戏剧学院重点实验室联合开发的“灵境石语——碑帖建筑VR沉浸项目”,就是利用VR技术将《九成宫醴泉铭》《城隍庙碑》《龙藏寺碑》三部碑帖作品中“涌泉”“求雨”“点彩”的碑文故事复原到中国青绿山水风格的虚拟场景中,再现唐朝人民的世俗生活和精神世界,观众可以通过沉浸式交互游戏,全景体验唐太宗在九成宫发现醴泉的历史故事。将VR技术应用于彝文历史档案资源的开发中,不仅可以解决珍贵孤本难以开发等问

题,还可以满足外部用户多样化的档案需求,搭建虚拟场景还原历史片段,再现历史场景。

## 五、结语

彝文历史档案是我国宝贵的历史文化财富,需要持续地保护好、开发好、利用好。“数字人文”这一新兴跨学科应运而生并渐趋成熟,为档案资源的开发利用提供了新视角、新思路和新渠道,产生了近乎变革性的影响,借助数字人文技术对彝文历史档案数字化开发进行转型升级。

### 参考文献:

- [1] 华林.西南彝族历史档案[M].昆明:云南大学出版社.1999.
- [2] 国家档案局.“十四五”全国档案事业发展规划[EB/OL].(2021-06-09)[2023-10-11].<https://www.saac.gov.cn/daj/toutiao/202106/ecca2de5bc44a0eb55c89076286883.shtml>.
- [3] 吴蝶,曹如中,熊鸿军,等.新文科背景下数字化赋能图情档案学科向数字人文转型发展研究[J].图书馆,2022(8):44-48+86.
- [4] 金波,杨鹏,王毅,等.“十四五”图书馆、情报与文献学学科发展态势与前瞻[J].图书馆杂志,2022(1):4-16.
- [5] 陈涛,张靖,赵宇翔,等.数字人文实践中特藏资源的关联数据实现机制探索——以方志资源为例[J].情报理论与实践,2022(7):180-187+147.
- [6] 位通,桑宇辰,史睿,等.基于知识重构的年谱时空可视化呈现:以《朱熹年谱长编》为例[J].中国图书馆学报,2022(2):62-75.

# Research on the Digital Development of Yi Language Historical Archives from the Perspective of Digital Humanity

GAO Jianhui

**Abstract:** The historical archives in Yi language are the essence of the historical and cultural heritage of the Yi people. The development and utilization of the digital humanities can effectively protect and inherit the historical culture of the Yi people. Based on the analysis of the current situation and practical difficulties in the digital development of Yi language historical archives, this paper combines the logical relationship between digital humanities and the digital protection of Yi language historical archives, and explores the ideas and specific implementation paths for the digital development of Yi language historical archives from the perspective of digital humanities, in order to guide the current practical work of digital development of Yi language historical archives.

**Keywords:** digital humanities; Yi language historical archives; digital development