·前沿与热点·

文化记忆视域下家谱文献价值的再认识和内容的深开发 *

夏翠娟

(1.上海图书馆 上海 200031)

摘 要:家谱文献作为“史料”的价值常因陷入虚构与真实之争而受到质疑,但作为“文化记忆”媒介的价值得到历史人类学相关领域的重视。文章通过对“文化记忆”理论框架的梳理,对“研究者如何利用作为文化记忆媒介的家谱文献”这一问题进行文献调研,分析了家谱文献在文化记忆视域下,认为作为“记忆”媒介而非“历史”媒介、是“当下”而非“过去”记忆的见证、所反映出来的“为什么”比“是什么”更重要等特性。文章试图通过更为精准地定位家谱文献的目标用户和典型应用场景,为家谱文献收藏管理机构在整理研究和开发利用方面,提出重新认识家谱文献价值的思路和深度开发家谱文献内容的方法及路径,并以上海图书馆家谱知识服务平台中“湖北麻城”迁入迁出相关内容的深度开发为例,验证了所提出的思路、方法、路径的可行性。

关键词:文化记忆;数字人文;家谱;移民史

中图分类号:G256 **文献标识码**:A DOI:10.11968/tsyqb.1003-6938.2019077

The Value Re-recognization and Further Content Development of Genealogy in the Perspective of Cultural Memory

Abstract The value of genealogy documents as historical literature is controversial because of the fabricated information about ancestors and family history,but the value as cultural memory medium has been recognized in the area of Historical Anthropology. The author made investigation about how the researchers use the genealogy documents as cultural memory medium and found the difference between the research of genealogy documents for memory and for history. The author tries to locate more accurately the target users and typical application scenes of genealogy documents, and then put forward new ideas and ways for institutions to re-recognize the value of genealogy documents and develop the content deeply. The author provides an use case about the research of immigration history of Macheng, Hubei province by transforming the textual abstracts to semantic data which is easy to compute for statistics and analysis, and tries to find some new research questions or answers.

Key words cultural memory; digital humanities; genealogy; immigration history

家谱作为一种特殊的文献资源,经常作为与正史和方志并提的史料,但与正史和方志相比,家谱大多由民间组织或家族力量纂修,存在着攀附权贵、牵强附会的成分,故其作为史料的文献价值常常受到诸多诟病,尤其是历史学研究者,因为历史学强调的是通过对过去的理性审视来发掘沉埋在史料中的事实,注重真实性、客观性和批判性。然而,自20世纪初德国社会心理学家哈布瓦赫提出“集体记忆”理论以来,首次将“记忆”和“历史”分开讨论。自此,关于“历史”和“记忆”的研究层出不穷,到了20世纪80年代后达到高峰,其中尤以扬·阿斯曼夫妇构建的“文化记忆”理论为最,可为重新认识家谱的文献价值、促进家谱内容的深度开发和有效利用提供不一样的视角和思路。

在“文化记忆”视域下,“记忆”被从“历史”中剥离出来,不仅是个人的生理和心理机制,更是作为一种协助群体形成认同感和凝聚力的“集体记忆”,可被社会框架不断重构和选择性遗忘。本文在梳理“文化记忆”理论框架的基础上,试图在家谱文献价值的认识上,脱离真实与虚构的争论,分析家谱文献作为一种文化记忆媒介的特性,调研家谱文献利用的目标用户、研究的问题域和研究场景,重新认识作为

* 本文系国家哲学社会科学基金项目“支撑城市记忆项目的‘数据基础设施’理论建构与实践探索”(项目编号:19BTQ007)研究成果之一。
收稿日期:2019-10-06;**责任编辑**:柴若熔

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

“记忆”见证而非“历史”见证的家谱文献对于历史人类学等相关领域的研究价值。在此基础上,反思当前家谱文献的收藏、管理和服务机构在家谱数据库建设和新兴的数字人文项目建设上的不足,提出家谱文献内容深度开发的操作性建议、实施方法和路径,并从上海图书馆家谱知识服务平台中的6万9千余种家谱文献中,选取与被作为历史移民发源地的“湖北麻城”的迁入迁出相关的家谱文献1240种,复本去重后1162种,针对反映移民记忆的迁徙事件进行内容的深度开发,对所提出的思路、方法和路径的可行性做进一步验证。

1 “文化记忆”的理论框架

1.1 区别于“历史”的“记忆”

“记忆”理论系统性地将“历史”和“记忆”区分开来,认为“记忆”不仅是个人的生理和心理现象,而是一种社会建构,受制于社会框架,受到群体意识的影响和塑造。人群的许多社会活动如节日、仪式等,通常是为了强化群体的集体记忆,以增强群体成员间的认同感和凝聚力。尤瓦尔·赫拉利认为将一个超过150人的群体聚集在一起的,不是生物学意义上的亲缘关系或领袖的号召力,而是该群体一致相信的“故事”,如宗教团体关于人类起源的故事、家族中代际流传或家谱中记载的关于先祖迁徙的故事等,这里的“故事”即可理解为群体的集体记忆。王明珂也认为:决定一个族群边界的,不是语言、服饰、器物,而是共同的集体记忆。

另一方面,记忆常与遗忘相伴相生,正由于记忆的社会认同和群体凝聚功能,记住什么遗忘什么通常是群体成员有意识有目的选择。“对于过去发生的事实而言,记忆常是扭曲的或错误的,因为它是一种以组构过去使当前印象合理化的手段。”人类学家埃文斯·普里查德关于东非努尔人的研究表明:“忘记一些祖先或特别记得一些祖先,是他们家族发展与分化的原则。”同样的情况在英国人类学家古立佛对非洲杰族的研究中被发现,古立佛称其为“结构性失忆”,这种现象在中国的家谱中俯拾皆是,研究族谱的学者们也称其为“谱系性失忆”。清康熙时孔氏家族修谱,规定凡不孝、不悌、犯义、僧道、邪巫、优卒、贱役等,都被认为是“辱祖玷宗,丧名败节,皆不准入谱”。还因修谱都是集资,所以规定族人每丁交银八分,违者不准入谱。

简言之,“记忆”是为当下服务的,如果说“历史”是为了探求史实而对过去的理性审视,“记忆”则是为了当下的目的对过去的主观塑造。区分“记忆”与“历史”,不仅促进了对历史真实性和客观性的重新审视,也为图书馆、博物馆、档案馆这样的文化记忆机构如何保存、组织和利用文献资源提供了理论依据和新的视角。

1.2 什么是“文化记忆”?

哈布瓦赫的主要贡献在于提出“集体记忆”的概念和“记忆”的社会框架,认为“记忆”是群体的而非个体的,是对过去的社会性建构。这一理论后来在社会学、心理学、文学、历史学、历史地理学、人类学等领域得到进一步拓展和深化。首先,心理学家通过实验和观察发现,成人的记忆依赖图像图画、语言文字等文化工具来存储、表达和传递,而文化工具则会在社会框架下不断重建和再塑造。心理学家巴利特提出“心理构图”的概念,作为过去经验和印象的集结,人们通常依据自己的心理构图重建过去。而个人的心理构图深受社会群体及文化工具的影响。在此基础上,德国埃及学和古典考古学家扬·阿斯曼和文化人类学家阿莱达·阿斯曼发展出了“文化记忆”理论,将集体记忆分为“交往记忆”和“文化记忆”,探讨了何为“文化记忆”、“文化记忆”如何形成、如何发挥作用等根本性问题,系统性地梳理了“文化记忆”的媒介(载体),并进一步澄清了有关记忆与历史之间关系的模糊概念。

所谓“交往记忆”,是指在承载者之间代际传递的集体记忆,随着承载者的逝世而消亡。而“文化记忆”则是承载于一定的媒介(如节日、仪式、书籍、文物等)之上,可以大范围传播、长时间传承、不断被重构的集体记忆[1]。与交往记忆相比,文化记忆可在更长的时间段内为更大规模群体的认同感和凝聚力服务。

2 对家谱文献价值的再认识

关于家谱的文献价值,已有大量的分析归纳,陈建华和王鹤鸣[2]从“内容的独特性、资料的原始性、记载的连贯性”三个方面肯定了家谱有别于其它史料

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

的特点。对家谱作为史料的局限,其诟病通常集中在这几个方面:宋以后家谱多为私修,是出自民间的非正式文献,未经严格审校,错讹之处较多;家谱修纂过程中扬善隐恶的倾向违背史家秉笔直书的原则;为了攀附权贵而冒认祖先伪造世系,为后世研究者造成了不必要的困扰。最为人不齿的是,清代末期"谱匠""谱师"的盛行,致使同姓家谱之间互相抄袭,伪造名人序文题跋屡见不鲜,极大地影响了家谱文献整体的可信度。20世纪末,葛剑雄[3]从家谱对于移民史研究、历史人口研究、社会史研究三个方面较为客观地论述了家谱作为历史文献的价值和局限,为研究者们将家谱作为史料时的去芜存菁提供了方法上的参考。自20世纪末的华南研究到近期的徽州、江南、华北等区域社会史研究的兴起,受到了历史人类学、文化人类学的影响:重视民间文献和田野研究,注重文献的现场感,强调在历史现场解读文献背后蕴藏的意义[4]。家谱作为一种典型的民间文献,常被用做一种"记忆"而非"历史"的媒介,与其它类型的民间文献及田野调查相互印证,发现隐藏在文献背后的社会框架、文化规范和历史变迁。

2.1 家谱作为"记忆"而非"历史"的见证

王明珂[5]在研究西周历史时,针对结合西周铜器铭文与先秦文献来研究西周史的主流趋势,提出了不一样的思路。他认为"一篇文献不只是'过去事实的载体',也是人们对过去'记忆的载体'。"西周铜器铭文记载的是西周时人认为重要的社会记忆,而先秦文献对西周的描述是战国至汉初时人对西周的记忆。通过研究两者之间的差距,来研究西周至先秦时期华夏与华夏边缘形成的轨迹。这种将文献作为"记忆"载体的观点尤其适用于家谱。家谱作为一种可以长时间传承、大范围传播的"文化记忆媒介",对家族记忆有一种固化和强化的作用。对祖源和迁徙源流的追溯和认定,是修纂家谱的重要任务。重修家谱是重建家族记忆、重新定义核心和边缘、增强家族内部和家族间的认同感和凝聚力的基本途径[6]。

赵世瑜[7]在研究山西洪洞大槐树移民现象时,认为家谱中大量记载的"大槐树"和"老鸹窝"应被视为民众便于传承的历史记忆符号,而不必与历史真实联系起来。陈世松等[8]在研究"湖广填四川"的移民传说诞生及演变轨迹时,比较了明初到清末不同家族的家谱中关于祖源的文本书写,发现随着时间的推移,湖广入川移民后裔对于原祖籍地的记忆已经渐趋模糊,而家谱中的记述却渐趋具体而确切,表现出一致性的文法规则。说明从明代中晚期开始,特别是到了明代晚期,移民对于家世来源的记忆文本书写已经规范化模式化,具有很大的传说色彩,由家族记忆而成为一种族群记忆了。

基于此,笔者认为在文化记忆视域下,家谱文献也可以被作为家族、民族、乃至社会的集体记忆见证。

2.2 家谱作为"当下"而非"过去"的记忆

家谱在家族世系图和世系表中详细记载的家族成员个人基本信息以及繁衍、通婚的情况,是移民史、人口学、历史人类学研究的常用材料。值得重视的是,在这些领域的研究中,家谱常被作为一种反映家谱修纂的"当下"而非家谱中所反映的"过去"记忆。如众多民族学研究者发现,在历史的长程中,云南大理白族地区张氏家族的先辈对自己祖源的解释有一个不断变动的过程。从明代开始的家谱记载中,关于祖先来自南京的说法流传甚广,甚至"有意"抹杀或隐藏其元代以前的家族历史记忆[9]。经过调查研究,不少学者认为这是当地少数族裔为了对抗民族歧视而采取的"冒籍"行为,以获得汉人的身份和认同。从文化记忆的角度,通过对祖籍的重新认定来重新界定自己身份,"冒籍南京"是对家族记忆的集体性再造和结构性遗忘,是为了在"当下"获得更好的生存环境[10]。

在移民史研究中,常常利用家谱中记载的先祖迁徙信息。由于家谱是由迁入地(家谱的谱籍)的人们纂修并传播的,通常对"始迁祖"及其迁出地有明确的记载。赵世瑜对关于"大槐树"移民传说的研究发现:移民传说大多是在迁入地创造出来并传播开来的。这也是"若问老家在何处,山西洪洞大槐树"、"湖广填四川,麻城占一半"、鄂皖两地流行的江西瓦屑坝移民、珠江三角洲流行的南雄珠玑巷移民这些移民传说在迁入地,也即家谱的修纂地更为流行的原因。这说明,家谱修纂地的人们更重视将不同的群体通过移民传说凝聚在一起,获得一致的身份认同。如珠江三角洲很多地方的居民都认为自己的祖先是从南雄珠玑巷迁移过来的,研究者发现这个传说作为史实是不

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

可靠的,而与明初广东人的入籍问题有关,当地土著为了取得合法身份纳入政府户籍,从而采用了南雄珠玑巷的传说以证明其出自中原的正统身份[7]。

如果把家谱修纂的时间看作是时间意义上的"当下",家谱修撰地看作是空间意义上的"当下",那么,家谱的内容反映的是"当下"对于家族历史和先祖源流的看法、对群体认同和家园记忆的塑造的方式以及社会所流行的文本书写规范,而不仅仅是忠实地反映家族"过去"的历史。从这个意义上来说,家谱作为一种历史文献,除了其记录"当下"之前历史的功能,其反映出来的那个"当下"及隐于其后的深层原因对于研究者来说显得更为重要。因而,家谱的价值不仅在于对家谱所反映的"过去"历史的研究,更重要的是对纂修家谱时"当下"的研究。

2.3 "为什么"比"是什么"更重要

日本人类学家濑川昌久对家谱中的真实与虚构有着较为清醒的论述:家谱中真实的或虚构的记载之根源,在于人们对自身以及自身所属群体的认识和思考,重要的不是这种虚构有多少真实性,而在于这种虚构行为本身是有意义的[11]。编纂者所代表的群体意识结构是当时观念和思潮的剪影,若仅仅止步于将隐善扬恶、攀附权贵归结为对某种虚荣的满足,而不去探究其背后存在的社会性与文化性规范,则无法更好地发掘家谱文献的价值。

周建新[12]发现客家族谱有一个共同的突出特点是大量存在风水活动的记载和描写。这些客家族谱里风水活动方面的历史记载,不仅反映了客家宗族的活动以及隐藏其背后的父系理念和历史意识,而且可以从另一个角度来理解客家传统社会文化。该研究指出风水是客家人一种生存和发展的策略手段,还是常见于客家社会生活中的一种生活常识,是一套存留于客家普通百姓头脑中的历史记忆和社会规范,具有约束和教化功能。

钱晶晶[13]通过三门塘四大家族的族语材料、碑文与访谈材料,大致勾勒了谢、刘、王、吴四大家族的定居历史以及三门塘村落的形成过程。透过一系列难辨真假的关于先祖身份和迁徙源流的记载和传说,发现三门塘村落中地方宗族势力的成长与明清时期清水江一带木材贸易的兴起与繁盛相关,在当时的经济发展情况下,谱系的建构对于家族财产的确认与继嗣变得至关重要,这不仅仅是一个适应经济生活的变通方法,也是寻求国家正统话语的文化手段。

在文化记忆理论框架下,利用家谱文献时,不纠结于祖源和移民传说的真伪,而是去研究创造和传播这些传说的人群到底是为什么要创造和传播这些传说。这种对为什么的追问,是对另一种真实性和客观性的探求。正如金观涛和刘青峰[14]在对历史背后的观念史长期探究后得出的结论:"为了达到历史研究的真实性,并不是去排除历史记录中的主观性就可以做到的,而是去寻找支配该事件发生的真实观念,并进一步分析该事件发生后又对群体的观念产生了什么样的冲击,即它是如何改变支配该事件发生的普遍观念。"

3 数字时代的"家族记忆":家谱文献内容的深开发

近年来,家谱尤其是华人家谱因为数量大、覆盖人群广、内容自成体系而作为一种特殊的历史文献得到各收藏机构、宗亲会等民间团体、寻根问祖的个人的高度重视,如美国犹他家谱学会在全球范围内展开华人家谱的收集、整理、研究合作项目,成为世界上保存家谱(原件、缩微胶卷和数字化拷贝)种数最多的机构,并建成在线的家谱数据库(FamilySearch.com)在互联网上提供家谱目录和全文影像的查阅服务。上海图书馆作为全球保存家谱原件种数最多的机构,也投入了大量的人力物力资源,在对馆藏家谱进行整理、编目的同时,形成了一系列的研究成果,如《中国家谱总目》涵盖全球597家收藏机构的馆藏目录;《中国家谱资料选编》精选编校来自家谱文献中的凡例、序跋、传记、诗文、家规族约、礼仪风俗、经济、家族源流、教育、图录、移民等内容;《中国祠堂通论》对祠堂的起源、历史、历代祠堂形制、种类、建筑、祭祀礼仪、祠堂文化和功能等问题作出了全面的调研和论述。另一方面,上海图书馆还以数字人文的方法和技术对5万余种《中国家谱总目》和3万余种馆藏目录进行了数据格式、系统架构和服务模式上的重建,与上海图书馆的古籍、手稿档案等文献库和人名规范库等基础知识库形成一体化的数字人文服务平台。然而,上述的家谱编目、整理、研究和

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

基于此建设的家谱服务平台，仍然是将家谱作为一种特殊的历史文献来对待的，家谱作为家庭、家族、族群记忆媒介的属性，却没有得到足够的重视。本文试图探讨在文化记忆视域下，家谱文献该如何开发利用，才能充分发挥其作为“记忆的载体”而非仅是“历史的载体”作用。

3.1 对家谱数据库建设的审视

自1992年联合国教科文组织启动了“世界记忆(The World Memory)”项目以来，世界范围内兴起了大量的文化记忆项目，一般由图书馆、档案馆、博物馆、纪念馆等文化记忆机构和研究学会、基金会主导和推动。在数字时代，利用数字化技术来促进文化记忆相关资料和数据的采集、保存、服务、传播，成为共识。加小双和徐拥军[15]对集体记忆、文化记忆、社会记忆等相关概念作了系统性的梳理，并认为数字记忆是记忆实践的发展趋势，体现了人文、艺术和科技携手并进、融汇贯通，其本质是将现代信息技术和社会记忆建构有机地结合起来，利用数字技术以数字形式来捕获、记录、保存重现社会记忆，进而实现对文化的保护和传承。如果把建设家谱数据库，看作是基于家谱文献来构建数字记忆的过程，那么当我们去审视当前各机构以传统的特藏资源的建库方式，和新兴的数字人文方法的建库方式，就会发现如下问题：

传统的特藏资源建库的方式，以文献管理和文献服务为主要目标，将家谱文献扫描成全文影像，遵循一定的元数据著录规范对家谱文献的题名、纂修者、版刻(出版)时间、谱籍地(居地)、版本类型、尺寸、册数等文献特征进行标引后，形成书目记录(目录)，存储于关系型数据库或图书馆自动化系统中，便于业务管理和文献管理，同时开发面向读者的服务系统，于局域网内提供家谱文献的书目查询和全文阅览服务。这种建库方式的弊端在于：面向文献的元数据标引难以充分揭示与挖掘家谱文献作为“文化记忆媒介”的特质，家谱文献中大量体现“当下”家族记忆的内容在数字世界里被消减抽象为简单的元数据记录。单个机构的馆藏资源有限，只能满足文献检索和获取的需求，无法满足学术研究的需求。学术研究要求尽可能地穷尽与研究问题相关的所有相关文献。在互联网和云计算时代，建设分布式的全球华人家谱数据库已成为可能，而传统的特藏资源建库方式没有充分利用互联网和云计算技术的红利。

新兴的数字人文方法建库方式虽然试图解决上述两个弊端，但在如何深入理解研究者的需求、如何有效帮助研究人员解决具体的研究问题方面仍然捉襟见肘。上海图书馆的家谱知识服务平台将《中国家谱总目》《中国家谱资料选编》的家规族约卷、3位名人(胡适、盛宣怀、李鸿章)的世系表数据化，在内容范围上不再是上海图书馆的单馆馆藏目录，而是覆盖了597家家谱收藏机构的联合目录，在功能上由于采用关联数据、GIS、数据可视化等技术和众包思想，在原来的基于字段关键词匹配的文献查阅功能之外，还提供地图检索、分面统计、文献关系关联、用户贡献内容等功能。《中国家谱总目》虽然是一部提要式目录，但仍然以文献目录为主，家规族约是经过精心选择的303种家谱的相关内容，3位名人的家谱世系表更是九牛一毛。然而问题不在于数量的多寡，而在于看不到这些数据的应用场景，其目标用户人群是谁？他们是如何利用家谱文献的，一般用于解决哪些研究问题，是否存在一定的模式和规范？在文化记忆的理论框架下，本文尝试为这些问题找到部分答案。在过去的认识中，家谱文献的目标用户人群，除了寻根问祖的普通大众外，就是将家谱作为“史料”的历史研究者，而他们往往陷入真实与虚构的泥沼中。在文化记忆视域下，家谱作为一种民间文献，在解决文化人类学、历史人类学相关研究问题时，能够发挥其它文献难以替代的作用。当将家谱作为“文化记忆的媒介”时，我们认识到，在数字世界中忠实地再现每一部家谱所反映的“当下”，重建那个“当下”的数字记忆，应作为家谱数据库建设的另一个重要任务。

3.2 重建家谱文献“当下”的数字记忆

重建家谱文献所反映的“当下”数字记忆，目的是为了在数字世界中还原家谱文献的历史现场。对于达成此目的途径，本文提出以下几点操作建议：

(1)在数字世界中如实反映家谱文献记载的内容：从数字化到文本化到数据化。将家谱文献扫描成影像文件，使纸质文献具有数字化的载体形态，虽然也是如实反映文献记载内容的一种形式，但只是完成了基础性的工作，因为扫描影像文件仍然需要人

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

用肉眼来阅读,在数字时代显得效率极为低下。还需要将内容文本化，以便于支持数字人文方法和技术对内容的深度挖掘，如全文检索或利用文本分析软件进一步处理。虽然目前基于机器学习的古籍文献OCR技术进展迅速,识别精度和速度大为提高。但如若将家谱文献内容全部文本化的同时还保留内容之间的逻辑联系，尤其是世系图和世系表的人物关系结构,仍然是一项需要人力参与的巨大工程,非短时间内可以完成。可以借鉴“威尼斯时间机器项目”的做法,先以家谱的“种”为单位,将重要的内容如人物(始祖、始迁祖、名人)、地点(谱籍地、原居地、迁居地、散居地)、重要事件(迁徙)等提取出来。始祖、始迁祖、名人、谱籍地可以利用已有的结构化元数据,而迁徙事件相关的时间和地点可用名称实体识别(NER)或机器学习等现代数字技术从《中国家谱总目》的提要或扫描影像文件中提取。而后将提出取来的人、地、时、事及其相关信息结构化处理,并基于关联数据技术进行语义编码后作为知识节点,建立彼此之间的关联关系,形成某一种家谱的单一知识图谱(见图1)。

(2)置万物于时空:为所有知识节点添加时间和空间属性。时间和空间属性对还原历史现场至关重要,因此,在提取家谱文献中重要内容时,需要将所有提取出来的内容置于其所在的时间和空间之中。简言之，也就是为提取出来的知识点添加时间和空间属性。如人物的生卒年、大事年表、籍贯、活动空间等,迁徙事件中的迁徙时间、原居地、迁居地等。虽然

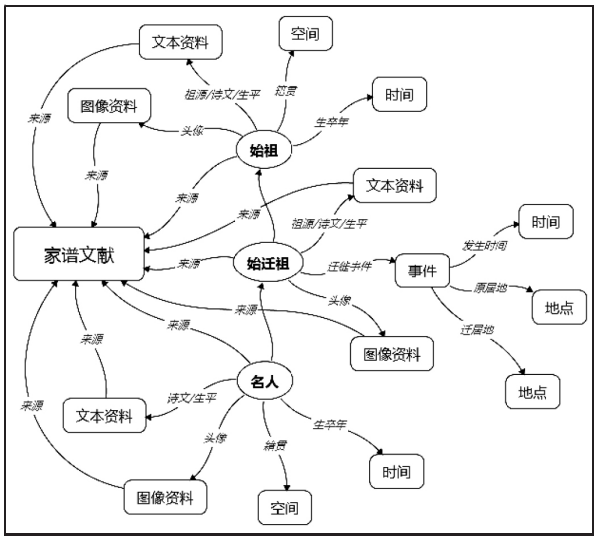


图1　单一家谱文献的知识图谱模型

家谱文献中对先祖的生平评价和祖源说明存在着文过饰非、夸功耀绩的现象,对于时间的表述大多采用标准不一的中国历史纪年法,地点的表述又大多为当时使用而非现在已经改名的地名。在家谱数据库中,仅需按照原文的表述忠实记录,不作主观性判断。

(3)建立家谱文献之外的多重证据参照体系。家谱作为一种“纸上之材料”,在由王国维创立的、而后对历史学、考古学乃至中国的近现代学术研究影响极为深远的“二重证据法”研究方法体系中可与其他证据相互印证,以便于对“记忆”抽丝剥茧、为“历史”去芜存菁。

二重证据法经过几代学人的不断探索，有了新的发展。如20世纪80年代的三重证据法:即在二重证据法的基础上,或是结合调查资料或材料中的“口述史料”研究历史学、民族学;或将考古材料又分为两部分——考古资料和古文字资料；或是再加上文化人类学的资料与方法的运用。除此之外,甚至可将历史文献与自然科学研究的成果互相印证，如复旦历史学和人类学联合课题组通过安徽亳州市曹操祖辈墓发掘的曹鼎牙齿获得其DNA数据,与曹操家谱文献互相印证,试图解开曹操的身世之谜并发现当今分布于各地的曹操后人[16]。到了信息技术高度发展的数字人文时代,除了传统意义上的历史文献和文物外,还可以加上基于大规模文献的统计分析数据、SNS可视化呈现图景、GIS时空模型、3D空间模型等。

在为每一种家谱建立忠实于原文内容并带有时空属性的单一知识图谱后，如能在知识节点的层面建立与之相关的多重证据参考体系，包括其它多种家谱文献的知识图谱,可灵活组配,动态地根据同谱籍、或同时代、或同姓氏、或同祖源等条件形成更大的复合型知识图谱。单一家谱文献知识图谱中的诗文、家规家训与其它类型的文献库如古籍、方志、档案等建立关联,人物、地点、迁徙事件等可与通用的人物资料库如哈佛燕京历代人物传记资料库(CBDB)、历史地理信息系统如CHGIS、历史事件库建立关联。这样基于结构化语义数据建立的大量关联关系,可方便地用于聚类和比较分析,使家谱文献中的内容与各种证据互为印证,在此过程中,可借助数字人文工具和方法所提供的聚类和分面统计分析数

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

据、社会网络关系可视化、地理信息系统可视化以及时空关系可视化图表(见图 2)。

3.3 以家谱文献中的人物迁徙事件为例

湖北麻城孝感乡、山西洪洞大槐树、南雄珠玑巷、宁化石壁、苏州阊门、江西瓦屑坝常被作为中国历史上的大规模移民发源地或中转站,但相关记载却少见于正史和方志,而往往作为移民传说和祖先记忆被记载在家谱文献中。在历史人口学和移民史研究中,家谱中记载的始祖或始迁祖的迁徙信息是非常宝贵的资料。在《中国家谱总目》的提要中,编目人员简明扼要地撰写了关于始祖和始迁祖的迁徙信息,基本包含了迁徙人、迁徙时间、原居地和迁居地信息。

本文在所提出的操作建议基础上,以“湖北麻城”的迁入迁出数据为例,将《中国家谱总目》提要中的迁徙信息转换为结构化、语义化的数据。试图探索“湖北麻城”作为移民群体记忆中的迁徙中转地,在家谱文献中是如何记载的,通过可操作的数据将静态的文字记忆变成动态可展演的记忆,以观察家谱文献中“湖北麻城”相关的迁徙事件,在时间和空间上呈现出何种趋势或规律、可以了解何种细节、以及是否能发现新的研究问题。本研究所利用的数据基础是上海图书馆家谱知识服务平台来自全球 597 家收藏机构的 6 万 9 千余种家谱,从中选取提及湖北麻城相关迁徙事件的家谱 1240 种,复本去重后 1162 种,提取先祖 1042 人,涉及迁徙事件 1837 条。每条迁徙事件由人物、迁徙时间、原居地、迁居地 4 个属性组成(见表 1)。

从时间属性来看,家谱的纂修时间有清代、民国和当代,以民国为最。根据文化记忆理论,这并不能如实地反映有关湖北麻城迁徙情况的史实。只能说可以从一个侧面反映出四川人民对于湖北麻城的祖源认同自清代开始,在民国时期达到高峰,并一直延续到当代。从迁徙事件统计数据来看,迁入到湖北麻城的事件数量大于从湖北麻城迁出的事件数量。从迁徙时间来看,迁入到湖北麻城的时间主要集中在元末明初之际,其中明洪武二年是一个重要的时间点。从湖北麻城迁出的时间则主要有两次,元末明初和明末清初,其中明洪武二年和清康熙年间是一个重要的时间点。结合家谱的纂修时间和家谱中记载的迁徙事件发生的时间,可以提出这样一个问题,为什么迁徙事件始于元末明初,但祖源记忆却是从清代才开始流行呢?这个问题有待历史人类学家进一步研究。

从空间属性来看,这 1162 种提及湖北麻城相关迁徙事件的家谱,谱籍地涉及 11 省。其中,湖北共 750 种,麻城占 558 种;其次是四川和重庆,共 358 种,其中四川宣汉、重庆丰都、四川简阳超过半数之

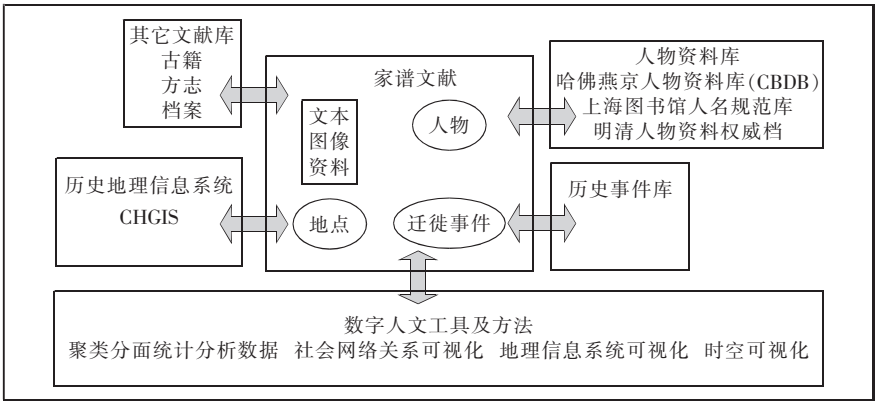


图 2 家谱文献的多重证据参考体系

表 1 1162 种家谱中湖北麻城迁入迁出数据统计

家谱种数	1162	谱籍地	共 11 省:其中湖北 750(含麻城 558)、四川 248、重庆 110	纂修时间	清 260;民国 508;当代 383
迁徙人数	1042	迁徙次数	1837	迁徙地名	迁入地 206;迁出地 206
迁入到麻城	737	迁徙原居地峰值	江西 494、湖北 111	迁徙时间峰值	元末明初 354(含明洪武二年 48)
自麻城迁出	587	迁徙目的地峰值	四川(含重庆)371、湖北 149	迁徙时间峰值	元末明初 246(含明洪武二年 50);明末清初(至康熙)120

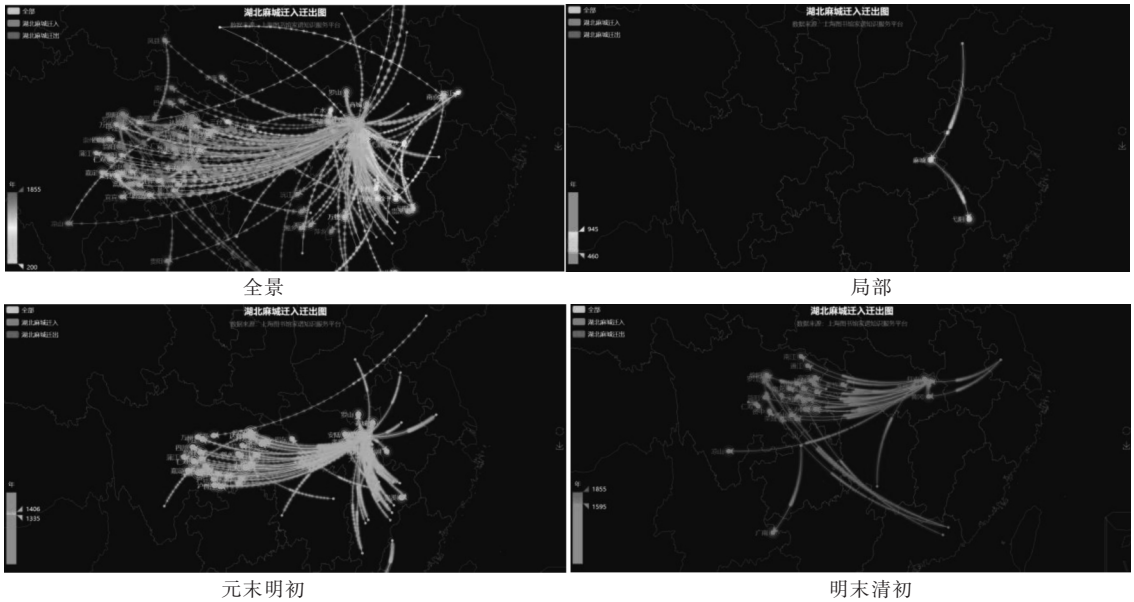
(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

多。这里是否可以提出一个这样的问题,是什么样的原因使得湖北麻城的祖源记忆在这几个地区更为流行?从迁徙事件的原居地和目的地来看,迁入到湖北麻城的人主要来自江西省,自湖北麻城迁出的目的地主要是四川(含重庆)和湖北省内。将迁徙数据在地图上呈现,可看出迁徙事件在时间和空间中的分布情况(见图3)。利用GIS系统,可对事件范围、迁徙人、迁徙事件、原居地、迁居地进行选择性操控和交互,既可俯视全貌,又可察看细节,还可链接至外部知识库进行多重参照研究。图中的地点和迁徙路径均可点击,点击地点,可链接至历史地名库,点击迁徙路径,可链接至原家谱,时间段可任意拖动,地点和迁徙人物可查询、筛选、过滤,若是始祖和名人,则可链接至人名规范库,查看更多家谱文献之外的其他相关文献。

从该例子中可以看出,当对家谱文献的内容进行深度开发,将文本转为结构化、语义化的数据后,利用时空可视化技术,可以针对具体的问题进行全景透视式的观察和剖析,所生成的统计分析数据和可视化图表,可作为多重证据参照体系中的一环。对研究人员经过资料搜集、遴选、研读、比较、解释、演绎过程的个案解析式研究形成了一种方法论上的补充和参照。以此类推,当所有的家谱文献内容中的迁徙信息都经过了这样的处理,变成了可以操作的数据,系统就可以在前端服务界面提供这样的服务:迁徙时间段、迁徙人群、迁出地、迁入地,作为可选择可操控的条件,系统根据研究人员的需求动态生成任意指定时间段中指定人群从指定地点到另一地点的迁徙图和聚类分面统计分析数据,或者生成所有家谱在不限定的时间段和地域中的迁徙全景图。而当研究人员需要了解某一次具体的迁徙情况和原始文献记载时,只需要点击该迁徙线段,就可以进入相关的家谱文献一探究竟,当要了解迁徙人的背景和生平,则链入该人物在人物资料库中的个人页面,当要了解迁徙时间的历史背景,则链入事件库了解此一时期的历史大事或其他类型的相关文献库如古籍、方志、墓志等。这样就将家谱文献置于多重证据参照体系之中,以弥补家谱作为史料的不足,同时帮助研究者追问和解释记忆背后的"为什么"。

4 结语

家谱文献的价值由于时常陷入虚构与真实之争而受到质疑,但在文化记忆视域下,其作为"记忆"媒介而非"历史"媒介、是"当下"而非"过去"记忆的见证、所反映出来的"为什么"比"是什么"更重要的特性,已经得到了历史人类学等相关学科的重视,并基于家谱内容的充分研读和解析,结合其他文献和田野调查,发表了大量的研究成果。而对于家谱文献的收藏管理服务机构,却仍然将家谱作为与正史和方志并提的史料来看待,在整理研究和开发利用上,难



全景　局部

元末明初　明末清初

图3　湖北麻城迁入迁出图

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

以充分发挥家谱文献的价值。厘清文化记忆理论框架，分析历史人类学相关领域在文化记忆视域下如何利用家谱文献，有助于从另一个角度来重新认识其对于人们寻根问祖和人文社会科学研究工作的价值，重新理顺家谱文献开发利用和提供服务的思路，寻求新的方法和路径。

目前的家谱数据库建设仍然停留在文献层面的描述、揭示和查阅服务，上海图书馆家谱知识服务平台在建设过程中已经认识到家谱内容层面揭示的重要性，如对人、地、时等知识节点的初步描述和关联，提供家规家训全文检索阅览、极少量世系图的数据化和可视化的尝试等。并使用了一些典型的数字人文技术和方法，如聚类分析和统计，基于地图圈画多边形的查询和时空浏览，但对除了提供查阅服务外还能如何更有针对性地精准服务于目标用户的研究需求，还没有清晰和充分的研究和认识。本研究在调研家谱文献对于人文研究的典型应用场景和精准目标用户的基础上，提出重新认识家谱文献价值的新思路，和深度开发家谱文献内容的方法和路径，是为进一步充分利用家谱文献和提升家谱知识服务平台的服务而进行的尝试，并以“湖北麻城”相关迁徙事件的深度开发为例来验证其可行性。

由于本研究所用到的迁徙数据源于对1162种家谱元数据的提要文字进行名称实体识别，进而提取其中的迁徙事件，形成结构化、语义化的知识图谱。将来若在家谱知识服务平台上提供单一家谱的知识图谱浏览和多种相关家谱的复合知识图谱的浏览，或所有家谱迁徙事件的时空分布图，面临的主要问题是如何将数据处理的数量从1162种家谱提要扩展到6万9千余种家谱提要，而非直接处理目前还没有完成从数字化到文本化转型的6万9千种家谱文献。后者在目前看起来尚属于不可能完成的任务，有待于人工智能技术的进一步发展，而前者在当前的技术条件下却是可行的。另外，本研究尚未及探讨的另一个问题是如何通过“用户贡献内容(UGC)”来保存当代人的家族记忆，虽然家谱知识服务平台在功能上已经支持专家用户修改部分数据、在线上传家谱并编目，将世系图数据化的在线识谱、以及在线修谱和续谱等功能，但仍可以在文化记忆视域下重新思考并作进一步改进，这将是本研究未来的研究方向之一。

参考文献：

[1] (德)扬·阿斯曼.金寿福，黄晓晨，译.文化记忆:早期高级文化中的文字回忆和政治身份[M].北京：北京大学出版社，2015.

[2] 陈建华，王鹤鸣.中国家谱资料选编第1册，总序[M].上海：上海古籍出版社，2013.

[3] 葛剑雄.家谱:作为历史文献的价值和局限[J].历史教学问题，1997(6):5-8.

[4] 常建华.历史人类学应从日常生活史出发[J].青海民族研究，2013，24(4):17-22.

[5] 王明珂.华夏边缘：历史记忆与族群认同[M].北京：社会科学文献出版社，2006.

[6] 薛柏成，朱文婷.叶赫那拉氏族谱与满族集体历史记忆研究[J].吉林师范大学学报(人文社会科学版)，2014(6):8-14.

[7] 赵世瑜.说不尽的大槐树[M].北京：北京师范大学出版社，2018:16，96-97.

[8] 陈世松.大移民：“湖广填四川”故乡记忆[M].成都：四川人民出版社，2015:272-282.

[9] 赵玉中.祖源记忆:社会情境与民族认同——《大理喜洲翔龙村张氏族谱》的个案研究[J].西南民族大学学报(人文社科版)，2011，32(8):23-27.

[10] 张海超.祖籍、记忆与群体认同的变迁——大理白族古代家谱的历史人类学释读[J].北方民族大学学报，2011(1):109-114.

[11] 胡鸿保，定宜庄.虚构与真实之间——就家谱和族群认同问题与《福建族谱》作者商榷[J].中南民族学院学报(人文社会科学版)，2001，21(1):42-45.

[12] 周建新.走向历史的田野:客家地区碑刻、族谱与文书的历史人类学研究[J].赣南师范大学学报，2009，30(1):1-2.

[13] 钱晶晶.历史人类学视角下的村落空间——三门塘人的谱系建构与姓氏空间[J].青海民族研究，2013(2):32-36.

[14] 金观涛，刘青峰.历史的真实性：试论数据库新方法在历史研究的应用[J].清史研究，2008(1)：96-114.

[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.

[16] 韩昇.曹魏皇室世系考述[J].复旦学报(社会科学版)，2010(3):11-19.

作者简介：夏翠娟，女，上海图书馆研究员，研究方向：元数据、知识本体、知识组织、关联数据、数字人文、文化记忆。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net