

出土文献数字化建设现状研究

王建欣

摘要：出土文献的数字化在推动出土文献的保护、研究和传播方面都有重要意义。当前出土文献数字化资源建设在重大项目的推进、多主体开发格局的形成、综合性及专题性数据库建设等方面都已经取得了显著成效，但还存在数据库资源重复建设、数据库不能相互兼容以及推介共享不足的问题，需要针对不同问题不断改进提升。

关键词：出土文献；数字化；建设现状

中图分类号：G230.7

文献标识码：A

文章编号：1009-1017(2023)01-0101-04

“所谓出土文献，是指考古遗存所得的甲骨文、金文、玉石文、陶文、玺文、货币文、简牍文、帛书等商周两汉文字资料，以及晋唐碑刻、敦煌写卷、吐鲁番文书等。”^①出土文献的整理与研究对于古文字系统的考证梳理，史实的考据印证以及历代名物典章制度的研究等方面都有重要意义。但与传世文献的整理研究相比，出土文献的整理研究存在搜集、保存困难，研究渠道闭塞，出版传播限制较多等不足，而出土文献的数字化则为我们提供了整理研究的新思路与新视角。近年国内出土文献数字化虽然已经有了较大发展，但在数字化推进的全面性以及系统性方面仍有待加强。因此，本文以出土文献数字化为切入点总结我国出土文献数字化的发展成果及存在不足，以期对出土文献数字化的产业发展提出切实的建议及启示。

一、出土文献数字化的现实意义

（一）数字化推进出土文献的全方位、多主体保护

随着国内外文献保护项目的整体推进，对文献遗产的保护逐渐由纸质载体向数字遗产的保护与开发转换^②。通过数字化方式对文献遗产进行保护成为当前以及未来文献保护的重要趋势。数字化的保护方式打破了传统单一的实物保护形式，缓解了出土文献保护与分享利用的矛盾，将线上、线下两种完全不同的保护形式有机结合起来，做到了物质实体保护与信息内容保护的并重统一，实现了出土文献的全方位保护。

出土文献的数字化可同时推动多元化保护主体的形成。出土文献一直以来的保存形式为分散保存在各博物馆、图书馆、文物局以及高等院校等收藏、科研机构，保护渠道以及保护主体相对单一与扁平化。而将出土文献数字化的保护方式除了可以调动常规组织机构参与以外，还可以将一些私人机构纳入文献保护体系，形成保护主体多元化的全方位保护模式。例如“龙语瀚堂典籍数据库”（2010年更名为“瀚堂典藏数据库”）即是由私人信息企业北京龙戴特信息技术有限公司与北京时代瀚堂科技公司联合开发的综合性数字化典籍数据库。

（二）数字化促进出土文献的系统、深入研究

出土文献具有重要的文字学、语言学及史学价值，无论是从文字形体层面还是从文献语料层面对其进行研究，都需要对出土文献材料进行细致的观察、整理与研究。但因具有重要的文物价值，出土文献需严密保存在不同入藏机构，这就导致一般科研工作者很难直接获取一手材料，在一定程度上造成了出土文献研究困难的局面。另外，出土文献的研究尤其是对于文字方面的研究，需要结合不同时期、不同器物、不

① 陈伟武：《出土文献之于古汉语研究十年回眸》，《古汉语研究》1998年第4期。

② 参见陈闽芳、李健：《“互联网+”环境下文化遗产“活化保护”模式研究》，《浙江档案》2019年第3期。

作者简介：王建欣（1985-），临沂大学文学院讲师。研究方向：古文字与古文献。

同文字材料中的文字形体,分析其源流以及发展演变,而研究所需的文字材料往往分散保存于各个不同机构,想要对材料进行系统深入的研究就需在各个机构间辗转往返,花费大量时间与精力。而出土文献纸质出版物一般较为滞后且缺乏整体系统性,这些原因都严重阻碍了对于出土文献的研究。出土文献数字化之后,研究者可通过已经数字化了的多路径系统材料检索系联不同时期、不同载体、不同出土地的文字材料,这都为对出土文献进行系统与深入的研究提供了更多基础与可能性。

(三) 数字化促进出土文献的传播弘扬

出土文献不仅具有重要的文物价值,同时也蕴含深厚的文化价值,大力促进出土文献的研究与传播符合当下传承、传播中华优秀传统文化的时代趋势。但传统的出土文献获取渠道狭窄且单一,对于普通社会大众而言尤难获取,且存在严重的“重藏轻用”情况。而出土文献在经过数字化之后,则可从一种不易寻觅和传播的珍贵文化遗产转化为能够面向社会大众且更易于获取的文献资料^①。经过数字化处理的出土文献材料在互联网环境下可以转变为一种易于流通、易于获取的信息资源,更易被社会大众认知并了解,从而实现社会化的资源共享与利用,并进一步提高出土文献在社会上的传播度与影响力。

二、出土文献数字化的主要成就

(一) 重大项目稳步推进

随着计算机的普及应用以及国外文献资源数据库的推进,国内亦开始重视对古籍资源的数字化,一些大的数字化建设项目稳步推进^②,出土文献的数字化在其中占据重要地位。

“中华古籍数字资源库”是具有代表性的重大古籍数字化项目。国家图书馆一直以来非常重视古籍保护工作,90年代中期开始积极进行古籍资源的数字化开发,“中华古籍数字资源库”即为“中华古籍保护计划”的重要研究成果。此数据库立足于国家图书馆馆藏资源,并联合各地方图书馆,相继启动了“中华医药典籍资源库”“法藏敦煌遗书”“赵城金藏”“数字方志”等一系列数字化项目。其中,与出土文献相关的数据库主要有“甲骨世界”以及“碑帖菁华”。“甲骨世界”数据库依托国家图书馆所藏甲骨资源,共收录甲骨实物目录2964条,影像5932幅;甲骨拓片2975条,影像5932幅。“碑帖菁华”以馆内23万余件拓片为基础建设而成,现收录石刻拓片2.5万余幅,影像3.1万余幅。

一些专门的出土文献数字化项目也相继得到开发应用。华东师范大学中国文字研究与应用中心基于顺应出土文献数字化的大趋势以及全面构筑古文字数字化应用平台的目的,组织开发了内容丰富的出土文献数字化项目,包括商周金文检索系统、战国楚文字检索系统、花园庄东地甲骨检索系统等数据库系统。“商周金文检索系统”包括“金文字库”“金文输入法”“金楷对应转换程序”和“金文资料库”四部分,主要提供字头检索,使用时需安装该中心金文字库软件“jinwen.ttf”。“战国楚文字检索系统”为以字为单位的检索系统,收录字头1365个,原始字形4354个,检索系统主界面可显示原始字形图片。“花园庄东地甲骨检索系统”提供花园庄东地甲骨文的全文检索,同样需按字头进行检索。

(二) 多主体合作开发格局基本形成

依靠占有文献的便利,传统出土文献出版主体多为高等院校、公共图书馆等机构,但随着数字化事业版图的拓展,越来越多的社会企业开始加入并承担出土文献数字化的重任。

除国家图书馆外,各地方图书馆亦逐渐成为出土文献数字化的生力军。浙江省图书馆依据馆藏石刻拓片,建成“馆藏拓片数据库”。“秘籍琳琅——古文献资源库”是北京大学图书馆依托馆内金石拓片、地方志、舆图等丰富的文献资源建设成的包括善本古籍、普通古籍、地方志、拓片目录等一系列数据库在内的综合性数字资源库。

高校及科研机构亦紧跟数字化进程建设了一系列的出土文献数字化资料库。由陕西省考古研究所吴镇烽主持开发的“商周金文资料通鉴”收录传世、出土的商周青铜器资料,包括出土地、时代、器物图片、铭文释文及拓片等资料信息。台湾“中央研究院”历史语言研究所主持建设的“殷周金文暨青铜器资料库”

^① 参见刘志基:《简说古文字数字化工具书的编纂》,《2004年辞书与数字化研讨会论文集》2004年,第198-204页。

^② 参见冯国军:《我国文学古籍资源数字化开发现状及启示》,《出版广角》2021年第5期。

“先秦甲骨文金文简牍词汇资料库”“简帛金石资料库”“汉代简牍数位典藏”“甲骨文拓片资料库”等数字化资料库从不同的建设角度以及层面构建了全方位的出土文献数字化资源。

出土文献数字化资源建设不可忽略的重要力量是掌握先进数字化技术的社会企业。北京爱如生数字化技术研究中心自成立以来即专注于古籍文献的数字化事业，开发了出土文献数据库“中国金石库”和“敦煌文献库”。“中国金石库”收录历代金石文献，包括十万余件金石拓片以及两千部金石志书。“敦煌文献库”收录各地所藏敦煌汉文文献三万余件，全文超一亿字，影像超过三十余万幅。除出土文献外，爱如生数字技术研究中心还开发了“中国基本古籍库”“四库系列数据库”“中国方志库”“中国丛书库”“中国类书库”“中国史学库”“中国儒学库”“明清档案库”等数字资源库。经过一系列的开发建设，爱如生数字技术研究中心逐渐构筑起了类型多样、兼及传世文献与出土文献，基本涵盖中国历史与文化的数字化资源，成为数字化资源建设的重要力量。

相较于传统的出土文献出版方式，数字化的出版方式大大放宽了准入标准，可以让更多有志于投身文献数字化的企业和个人加入进来，形成全社会共同参与的多主体合作开发格局。不同的主体单位还可以互相合作、优势互补。高校、公共图书馆、科研机构拥有丰富的文献资源和知识力量；社会企业拥有能够将信息转化为数字资源进行存储、识别、输出的先进的技术能力，二者结合必定可以充分发挥自身长处，在各自擅长的领域发挥重要作用，从而形成更为优质、高效的数字化出版模式。

（三）综合性与专题性兼顾的开发模式逐渐形成

综合性的数字化资源指包括传世文献、出土文献等多种文献形式的数据库，出土文献数据库只作为其中重要的组成部分。专题性的出土文献数字化资源则专门针对一类或几类出土文献资料，不涉及传世文献部分。随着数字化事业的推进，出土文献数字化事业已经形成了综合性数据库和专题性数据库兼顾并举的开发模式。

“汉达文库”是由香港中文大学中国文化研究所中国古籍研究中心开发的综合性数字资源库，包括“甲骨文资料库”“金文资料库”“先秦两汉资料库”“竹简帛书资料库”“魏晋南北朝资料库”“类书资料库”等出土、传世文献数据库。北京时代瀚堂科技公司与北京龙戴特信息技术公司联合开发的“瀚堂典藏数据库”亦为较具代表性的综合性数据库，既有出土文献资料也有传世文献资料，分“经部集成”“史部集成”“子部集成”“集部集成”“专题文献”“近代报刊”六个子库。

专题性数据库同样成果显著。除前文提及的陕西省考古研究所及台湾“中央研究院”史语所开发的专题性数据库外，另外一些较有代表性的数据库，如香港中文大学图书馆建设的“郭店楚简资料库”，武汉大学简帛研究中心开发的“中国古代简帛字形、辞例数据库”，复旦大学出土文献与古文字研究中心的“上博简字词全编资料库”“清华简字形辞例数据库”，四川天元文化产业公司和成都商网信息科技有限公司联合开发的“中国甲骨文文献库”都属此类数据库。

三、出土文献数字化建设存在问题及建议

（一）数字化资源的重复开发建设问题

“由于目前还没有一个统一的组织和机构去管理和协调甲骨文数据库的建设，加之甲骨文收藏单位和研究单位又十分复杂，因此，国内外重复开发和建设了大小不一的数据库。”^①不仅甲骨文数据库建设存在这样的问题，金文、简帛等出土文献数字化资源建设亦面临重复开发与建设的问题，浪费了大量的社会资源。而且，重复性的开发也会给读者带来检索、利用方面的困难。针对此问题，我们认为可以从两个方面解决：一，建立统一调度机构对出土文献数字化工作进行管理协调，有规划地将出土文献数字化工程建设成为涵盖所有出土文献类型，开发角度多样化、实用化的数字化资源系统。二，各单位在建设数字化资源库之前应进行细致的前期调研，了解数字化资源开发情况，以长远的建设眼光，努力将出土文献数字化建设成一个统一的国家性的整体性项目，而不应只考虑本单位及眼前情况。

（二）数据库兼容问题

^① 毛建军：《甲骨文全文数据库的建设与思考》，《图书馆学研究》2010年第12期。

当前出土文献数字化资源的开发多为图书馆、高校、科研单位等主体机构进行封闭式的开发，彼此之间除了在内容方面外，在技术层面同样缺乏沟通与共享。这就导致数据库在技术应用层面表现出排他的特质，数据库与数据库之间彼此不能够兼容，用户往往不能实现跨平台检索与利用。仅就字库一项而言，即使是同一古文字类别的字库，不同的数据库也不能彼此够兼容，需要用户重复安装。针对这一问题，我们认为出土文献数字化资源的开发应建立一个在相同的技术支持下，包括应用技术、搜索方式、应用界面等在内的，涵盖不同文献材料的，可以多层次、多角度进行检索和利用的标准规范体系。这不仅可以方便用户对数据库的使用，同样可以防止技术重复开发带来的资源浪费。

（三）数字化平台的分享与推广问题

当前国内出土文献数字化工作已经进入稳健发展阶段，不同类型、不同层次的文献数字化工作都已经取得显著成果，但数字化资源在分享与推广上的相对薄弱仍值得我们重视。现阶段出土文献数字化资源库的使用者主要为相关领域的科研工作者，但即使对科研工作者而言，数字化资源的传播度与获取渠道也是远远不够的，相当一部分研究人员对已经建成使用的资源库不够了解，并且因为很多数据库主要针对高等院校等机构单位，如果高校不购买该数据库，那么个人则会缺乏相应获取渠道。对于专业的研究人员尚且如此，那么仅仅是因对传统文化感兴趣而希望获得相关资料的普通人，想要了解数据库信息从而获取相关资料就显得更为困难。这些问题不仅在一定程度上阻碍了数字资源的推广与分享，同时也有违数字化资源开发的初衷。因此，出土文献的数字化平台不应只将市场局限于高校、图书馆等国家科研机构，而是应有步骤、有系统地逐步向公众开放，加大对公众的宣传力度，做到使公众方便易得地获取出土文献材料。

（责任编辑：闫丽）

欢迎订阅《古籍整理研究学刊》

《古籍整理研究学刊》（双月刊）为非邮发期刊，全年订费¥72元整，订阅方式：

户名：天津市河西区联合征订服务部

帐号：0302060509104619603

开户银行：天津市大寺泉集北里别墅17号

电话：022-23973378，23962479

Email：LHZD@public.tpt.tj.cn wms@lhzd.com

订阅时请标明：（2839号）《古籍整理研究学刊》×份即可，并可向其索要发票。

编辑部电话：0431-85099501