

# 联合建设旧方志数字资源的构想

吴澍时

中国地方志是一个地区综合性的百科全书，内容涵盖了历史、地理、文化、风俗等各方面。作为重要的地方文献资料，地方志不仅是了解一个地区历史与现实的珍贵参考书，对于地方史、社会学、民俗学、民间文学等学术研究也具有重要的意义。中国地方志历史悠久，现存数量较大，分藏于全国乃至世界各地。多年来，图书馆、档案馆、博物馆及其它研究机构在地方志的收藏、保存、整理、研究等方面做了大量的工作，也取得了显著的成绩。但是旧方志分散收藏的局面也给研究和利用带来了不便，在信息时代和网络环境下，旧方志的整理应向数字化、网络化方向发展，要通过网络渠道，实现方志资源的网络传输和共建共享。本文介绍目前旧方志收藏及整理的现状，汇报国家图书馆旧方志数字化工作的实践和经验，提出合作建设的几点思路。

## 一、旧方志收藏及整理概况

据《中国地方志联合目录》(以下简称《联合目录》)统计现存旧方志 8200 多种，分藏于全国 192 个图书馆、博物馆、档案馆及其各大学图书馆、研究机构。集中收藏的单位有国家图书馆、上海图书馆、南京图书馆、天津图书馆、中国第一历史档案馆、故宫博物院等。由于各收藏部门的数字化、网络化建设程度不一，地方志资源没有充分揭示出来，这在很大程度上影响了地方志的利用。笔者通过网络查询和电话咨询了解到一些单位旧方志资源的整理和利用情况。在许多图书馆、档案馆和收藏机构，旧方志的网络化工作已经起步，也取得一定的成果，如天津图书馆完成馆藏旧方志的影像扫描工作，在馆域网上提供浏览。甘肃图书馆编辑完成馆藏 2200 余种旧方志提要数据库。山东图书馆完成 500 种地方志的书目数据库。湖南图书馆网页上整合了收藏于不同机构的 401 种湖南旧方志目录提要，其中 308 种收藏在湖南图书馆，其余 93 种在博物馆、档案馆和其它单位收藏。但也有些部门由于资金短缺和工作重点不同，地方志的数字化、网络化工作还没有开始。由于地方志收藏比较分散，地方志的数字化合作工作需要全国性的权威机构来统一协调完成。

另外，全国范围旧方志的收藏情况还需要做进一步的清理，《联合目录》是目前比较全面反映全国旧方志收藏状况的重要工具书，但档案馆旧方志在该书中收录很少，只有中国第一历史档案馆 1384 种、广西档案馆 98 种、河南档案局 105 种。现阶段，全国各省市档案馆的旧方志目录大部分都没有整理出来，笔者通过查阅各省市档案馆指南，了解部分地区档案馆旧方志的收藏情况。如在四川省档案馆收藏清代志书 104 部，涉及 81 个县，收藏民国时期方志近百部<sup>1</sup>。吉林档案馆收藏清代、民国及东北沦陷时期官方部门编印的官报、年鉴、通志、调查报告等 2。上海档案馆、广西档案馆、青海档案馆、云南档案馆、福建档案馆等各地档案馆都收藏多部清代至民国时期本省及全国各地的方志资料。清理档案馆收藏的旧方志及相关文献，应是今后方志整理工作的主要内容。

旧方志收藏整理工作的状况基本上是有共藏而无共建，即不同部门都收藏大量的方志文献，但在文献整理、利用以及数字化、网络化建设方面合作的力度还不够。主要有以下几方面的原因：

第一、分散收藏的局面以及部门隶属关系的不同给文献数字化工作带来很大的困难。旧方志现存数量大，但比较分散，除集中收藏在几个较大的机构外，其余则分散于各省、市、县的图书馆、博物馆、档案馆以及大学图书馆、研究机构等。不同的收藏单位隶属于不同的部门，多部门之间的合作不容易实现。又由于资金和人力的原因以及工作重点不同，各部门都把有限的人力和资金投入各自馆藏的重点，如中国第一历史档案馆数字化的重点在于馆藏清代的全宗档案，故宫博物院的“数字故宫”项目，重点在于它的珍贵藏品和建筑。

第二、由于缺乏统一协调的部门，各收藏单位地方文献工作的规范和标准不同，书目数据格式不一，影响了地方文献的共建和整合。如旧方志目录，国家图书馆使用的是 CNMARC 格式，科学图书馆准备使用元数据格式，

山东图书馆用的是电子表格 3。在古旧方志数字资源的建设中，更需要成立一个部门来协调管理，制定数字化的规范和标准，制定出可操作性的合作方案和协议。

第三、资金不足是影响古籍数字化建设的主要因素之一。古籍数字化工作需要投入大量的资金和人力，但是目前各收藏单位经费有限，只能维持正常的业务工作，很难有多余的资金投入古籍数字化中。

其实，在学术领域，关于不同收藏单位文化资源整合的必要性、可能性，关于整合的方法及模式早有文章论及，比较认同的观点是要通过数字化、网络化形式实现资源的整合。这是一种比较可行的方式，既可以避免重复建设，避免资金和人力的浪费，又可以实现网络传输、共建共享。有学者认为：“数字图书馆工程应当成为整合中华文化资源的平台。”<sup>4</sup>并进一步提出中国国家图书馆倡议的中国数字图书馆联盟应吸纳档案馆和博物馆参加，全国联合书目数据库应纳入档案馆和博物馆的相关馆藏目录，形成全国中文文化资源目录数据库。

对分藏全国各地的旧方志文献进行数字化加工与合作具备一定的基础。许多省市图书馆、档案馆和博物馆都开始了数字化工作，只是根据馆藏的特点，数字化工作的重点不同，但各自在数字化方面都积累了一定的经验。国家图书馆对旧方志书目数据的整理和正在进行“数字方志”项目为古籍数字化、网络化工作提供了实践的经验，奠定了一定的基础。由全国多家图书馆和信息情报单位联合进行的数字图书馆相关标准规范研究也取得了一定的成果，国家图书馆负责地方志相关标准和规范的制定，已完成《地方志描述元数据规范》和《地方志描述元数据著录规则》，目前正在进行数据测试。这为不同收藏单位的地方志数字化工作提供了统一的规范、标准和理论基础，有利于全国范围内地方志资源的揭示和利用。

## 二、地方文献数字化建设的实践——“数字方志”工程

中国国家图书馆地方志家谱文献中心收藏明至民国间旧方志 6,013 种，近 12 万册。计划用 5 年的时间，对馆藏的旧方志（1949 前所修）进行数字化加工，即“数字方志”工程。“数字方志”以全文影像资源库和全文文本资源库为基础，在影像、文本之间实现全文、单项、复合检索，同时也实现简繁、异体、避讳字之间的关联检索。截止 2007 年底，馆藏旧方志已全部完成影像扫描，旧方志数字化的部分成果已在国家图书馆主页“数字方志”栏目中发布，目前应该尽快策划发布系统，要把数字化的旧方志资源全面呈现给公众。

“数字方志”是中文古籍数字化的一个实验项目，在该项目实施的过程中吸收了其它古籍数字化项目的优点，如要实现全文检索，实现简繁、异体、避讳字之间的关联检索等。“数字方志”在古籍数字化工作方面也有一些突破，并为其它古籍数字化工作提供了经验。

从目录标引上看，该项目根据旧方志的特点做了三级目录的指引，根据书目、卷次、卷名、篇名分别指引到原文，这样细致的工作更方便的读者的查找和利用。如[道光]《广东通志》共三百三十四卷，卷一百五十八至卷一百八十为《经政略》，包括铨选、恤助、经费附、禄饷、恤政、田赋、盐法、榷税、积贮、祀典、土贡、诸夷入贡事例、学制、兵志、邮政、鼓铸、船政、市舶等内容。目录指引到铨选、田赋、盐法等篇目。从使用的字符集来看，“数字方志”在全文转换中使用的 Unicode4.0 大字符集，包含 7.1 万个汉字，提高了全文转换的质量，使检索更加准确。另外在文本转换过程，数字化加工公司实现了异体字、避讳字的造字功能，避免了文本中的贴图字。在文本转换过程中，收集整理旧方志中出现的表外字，国家图书馆数字化管理部门对这些表外字进行比对，并请古文字学专家鉴定，准备提交给国际标准化组织表意文字工作组（IRG）。

由于旧方志文献的多样性、特殊性，该项目面临很大的困难，尤其在文本转换过程中会遇到许多问题。如版式还原问题，旧方志版式不完全统一，有抄本、稿本，又有眉批和注释，在文本转换的过程中既要客观反映原貌，又要考虑版式的美观。又如旧方志包含的内容较多，举凡一地的历史、地理、文化、艺术等等无所不包，在文本转换中会遇到大量的异体字、生僻字，特别是在地名、金石、古迹里较多。文本转换是一项需要投入时间、人力和资金的工作，需要工作人员具备的一定古文基础，需要细致、重复的校对，必要时还要请古籍方面、文字学方面的专家来认定，这样才能保证文本转换的质量。但目前由于资金问题，有些工作还无法做到。

### 三、联合建设旧方志资源库的构想

旧方志文献分藏于各地区不同单位，在合作整理方面已有《联合目录》和《中国地方志总目提要》的出版，已经为研究和利用提供很大的方便。但是各收藏机构在数字化、网络化方面合作建设的力度还不够，在全国范围内还没有实现资源共建和网络传输。旧方志的整理和数字化的合作包括以下几个方面内容：

第一，以《联合目录》和各省市的地方志目录为基础，对现存旧方志收藏情况进行彻底调查。《联合目录》出版至今已二十余年，为读者研究和利用方志文献提供了很大的方便。随着方志文献不断被发现，在学术领域不断有对《联合目录》的补充和修订，山西、黑龙江、山东、内蒙古等地都有方志文献需要重新著录。另外《联合目录》中反映档案馆收藏很少，调查工作要补充档案馆的收藏。从方志文献的整理到学术研究领域，都需要一部全面反映方志收藏状况的工具书。

第二，在调查工作进行的同时，建设现存旧方志的书目数据库，为公众网上查阅目录提供方便。调查工作和文献著录要借助网络来实现，随着调查工作的全面和深入，最终也自然形成了方志书目数据的总库。旧方志分藏于全国各地，许多收藏单位还没有开始进行目录整理。在调查和书目数据的建设中，可以利用已经完成的数据。如国家图书馆的旧方志占现存总量的 70%以上，并已完成所有馆藏旧方志的机读目录，在国家图书馆网页上发布，供读者查询，甘肃图书馆已完成馆藏 2200 种方志的提要目录。没有制作书目数据的单位可以利用已建资源，根据特殊的需要对数据进行再加工。这样既避免重复建设，也可以通过利用使已有的书目数据更加完善。

第三，合作建设旧方志影像资源库和文本资源库。据《联合目录》统计，现存旧方志 8200 多种，分藏于全国各地 192 个单位，其中单本只占总量的 10%左右，其它都有两个或两个以上复本，分藏于不同的单位，一种方志在十几个单位同时收藏的也不占少数。从全国范围考虑，旧方志数字资源建设的目标就是要完成现存旧方志的数字化。旧方志影像资源库、文本资源库的建设需要投入大量的资金和人力，其成本要远远高于书目数据的制作成本，也要高于其它二次文献的制作成本。因此要避免多个收藏单位同时建设、重复建设的情况。如上文所述，国家图书馆所藏古旧地方志的数量占现存总量的 70%以上，国家图书馆已完成数字化的部分，其它收藏单位则没有必要再进行数字化加工。以已完成的数字资源为基础，补充其它单位收藏的部分，最终形成中国现存古旧方志的数字资源总库。朱士嘉先生早在上个世纪八十年代对旧方志的收藏情况做全面的调查，遗憾地指出，目前全国没有一个单位能够收全现存所有旧方志，他建议组织一个调查团对散见海外的地方志做全面、彻底的调查，把所有的孤本全部复印回来。朱士嘉先生的愿望应该在数字时代得以实现。

联合建设旧方志数字资源需要几个必备条件：首先这是多个单位合作完成的项目，需要成立一个机构统一协调，要制定出可操作性的合作方案和协议，包括项目管理、资金筹集、资源利用、权利和义务等等内容。其次，要制定出旧方志数字化工作的相关标准和规范，确定一些常见问题的解决原则和方法，在统一的标准和协议条件下互通有无，共建共享。最后，完成旧方志数字化的计划需要财政的支持，这是一个基本的前提。旧方志数字化是对古籍资源的保护与再生，数字化的资源实现网络传输、共建共享，为专家学者利用方志进行学术研究提供便利的条件，对于弘扬中华传统文化更具有重要的意义，这是一项值得资助的工作。

地方文献的收藏保护和整理利用需要不同的收藏部门合作完成，在数字化、网络化环境下，各单位应该发挥优势，联合起来建设中文古籍数字资源。这样既能节省资金和人力，又能使分布不同地区、不同单位的文献资源得以整合和利用。