

# 敦煌研究院资料中心信息资源的开发与利用

李鸿恩 李润渊  
(敦煌研究院)

提要：回顾了敦煌研究院文献信息资源数字化建设的发展历程，提出了四个阶段的实施步骤并对敦煌学特色文献数据库做了重点介绍。

敦煌研究院资料中心是为院内学者提供全方位的信息资料服务的部门。从 1944 年建立敦煌艺术研究所以来，历届领导都很重视敦煌学资料的积累。特别是在 1984 年扩建成为敦煌研究院以后，更是加大了敦煌学资料收集的力度。现在资料中心藏书，共计近 13 万册，其中单本书占 70%以上，与专业研究密切相关的资料占 50%以上。同时我们还与 100 多个敦煌学家和敦煌学研究单位建立了资料交换，几十年累计收集到敦煌学相关中外文论文 1 万多篇，专著 1 千多种。

在传统图书馆时代，我们只是被动的向读者提供所需的纸质文献资料。随着现在网络的发展、海量网络信息和数字图书馆浪潮的出现，敦煌学学者若没有各种信息资源作支撑，只靠传统的渠道去收集有限的纸制资料，其研究速度、形式和效果都会受到影响和制约。

为了加快敦煌学研究的进度、提高质量和效果，更好的为院内学者提供服务，我们制定了敦煌研究院资料中心信息资源的开发与利用计划。

这个计划的指导思想是：突出重点，坚持先易后难；先中文后外文、先普通图书后期刊资料；先二次文献后全文的原则，对能体现专业特色的资料做深做细，即打破常规，尽可能利用已有技术成果，坚持高起点、高标准、引进国内市场上比较成熟、商品化较高、售后服务较好的软件和信息产品；重视合作精神，在条件允许的情况下可与其他机构合作。

这个计划在实施步骤上分四个阶段：

第一阶段：实现传统图书资料工作的计算机管理，按出版物的印刷单元组织文献资料。

第二阶段：特色数据库建设，对文献资料内容进行深加工，按知识单元组织文献信息。

第三阶段：按信息单元组织文献资料，建设全文及多媒体数据库，实现敦煌研究文献信息全部数字化

第四阶段：对敦煌学资源更深层次的开发。

第一阶段，在院领导支持下，我们在 2000 年经过全国范围的调研和严格的论证选择了深圳大学开发的图书馆自动化集成系统。这个系统为客户/服务器体系结构，采用网络技术，可实现跨平台数据库管理，可按专题文献建立特色数据库。根据自身特点，我们采取了突出特色，走捷径的办法，建设好二次文献书目数据库。具体作法是：敦煌研究类及相关重点图书资料，如，丝绸之路、河西史地、佛教艺术及石窟研究的资料在 MARC 格式下要尽可能详尽著录，对其它资料如：小说类、文艺作品、政治、法律类则采取简要著录的方法。同时，我们还与兄弟图书馆积极协作，引进了一批基本数据，缩短了建库时间。经过全体人员的艰苦努力集中了五个多月时间完成了馆藏中文普通图书的建库工作，共收录 5 万余条记录，初步实现了传统图书馆的计算机管理，简化了借阅方式，方便读者上网查询。

第二阶段：在现有基础上对馆藏敦煌学专题文献进行深加工，建立特色数据库，利用网络技术，建设敦煌学研究文献信息检索系统，使之成为国内外有一定影响的专业资讯网。

在敦煌研究院及资料中心领导的支持下，组织了专人收集资料，参照国内外已有词表的编制体系，编制了《敦煌学检索词表(库)》它是一部规范敦煌学术语概念的词表(库)，是为敦煌学文献信息机构进行信息加工、整序、标引和信息资源数字化，建立计算机信息资源检索系统的工具性词表(库)。现在这项工作马上就要结束了，估计在今年年底将会公布。

我们与深圳大学图书馆联合开发了敦煌学专题文献数据库。这个数据库是客户/服务器型数据库，可以对图像、多媒体和各种格式的文件进行处理和检索。该项工作内容包括，按国家信息处理标准对 1 万多条专业研究文献逐一进行主题分析与标引，列出约定俗成的关键词，特别是洞窟编号与遗书编号等字段，再编制支持特色检索的数据库结构，将相应字段逐条录入。为保证标引质量，在开始阶段要求必须填写工作单，由校审与录入人员严格把关，分级负责，落实到人。现已开始从《敦煌研究》所包含内容作起，到年底可完成 80% 的工作量。在这个系统中，每个关键词特别是洞窟编号，遗书编号都可以为一条检索途径，利用软件功能将洞窟编号与相关文献的检索、标识链接起来，就能够回答某洞窟被研究和利用的全部资料，真正起到洞窟研究和遗书研究档案的作用。

同时与深圳大学图书馆联合开发了敦煌学研究文献信息检索系统。它是围绕着“以读者为中心，以学科典籍为基础，人性化设计理念，规范化著录”四个方面来设计的。除具有题名、著者、分类、主题等常规检索外，提供了按洞窟号、遗书号检索，可将该标目下的题要、全文、综述及摄影作品、临品等全部相关信息提供给用户。检索点全面，检索方式多样，逻辑检索功能所有字段均可任意匹配检索，作到查询无盲点，可定义输出检索结果的限制条件。数据库模型设计上突破了传统的方法，在关系型数据库中首次将事物的数量词作为叙词处理（如洞窟号 m. 275, 遗书号 p. 2005）将它们作为主标目。可将石窟和遗书研究的全部信息集中整合处理，体现了学科特点，起到了石窟研究与遗书研究档案的作用，使数据库功能大为扩展。在信息处理中严格遵守国际标准和国家标准，采取了将 MARC 机器可读目录）格式与元数据相结合的方法，提供了对主题、分类、题名、著者等十多种规范文档的管理。以敦煌学研究成果为对象，对文献中的检索点深入分析，按元数据语义标准准确选择标引词，大大加快了数据库建设进度，提高了检索质量。

第三阶段，在这个阶段我们计划通过独自或者合作的方式，将我们特色资料全部数字化后添加到敦煌学专题文献数据库中。

与此同时，我们要经常征求用户意见做到认真服务主动回访，跟踪服务，提高敦煌学专题文献数据库的利用率。经过人工干预，系统合理地备份网上信息，收集新敦煌学和相关文献，把这些文献经过充分细致的分析研究，严格加工、筛选、归类和整理后不断的添加到数据库中。使敦煌研究院资料中心，不仅成为研究敦煌学信息服务中心、专业资料中心，而且真正成为国际敦煌学信息搜集和研究中心。

第四阶段，在做好上述三个阶段的工作后我们还要对敦煌学资源更深层次的开发。

敦煌学内容十分广泛，主要包括敦煌遗书、敦煌石窟考古、敦煌石窟艺术、敦煌史地、敦煌科技、敦煌旅游、敦煌学理论等。我们仅仅做敦煌学专题文献还是不能够满足读者的需

求。我们计划通过敦煌研究院院内合作以及与国内外机构实现资源共享和数据库之间的互相链接的方式，来扩大我们的馆藏资料（虚拟馆藏）。

敦煌研究院院内其它部门也在计划或正在进行一些敦煌资源的数字化工作。

敦煌研究院摄录部正在对几十年来拍摄的敦煌石窟群的几万张底片、大量的录像带以及相关的多媒体资源进行数字化和计划建立数据库。

敦煌研究院保护所正和美国梅隆基金会联合开发的数字化洞窟项目，已经拍摄了 30 多个洞窟，在敦煌研究院的网站上可以看到敦煌莫高窟第 323 窟虚拟漫游 (<http://www.dha.ac.cn>)。

敦煌研究院文献所和兰州大学图书馆联合研制开发了《敦煌遗书数据库》它是为《敦煌遗书总目索引新编》（施萍婷主撰稿 邵惠莉协编 北京 中华书局 2000 年 7 月出版）研制的计算机查询检索程序。

敦煌研究院敦煌石窟文物保护研究陈列中心计划对保存的敦煌石窟的各种出土文物建立数字化档案。

敦煌研究院美术研究所计划对从 1944 年建立敦煌艺术研究所以来所有壁画临摹品建立数据库。

我们也计划对敦煌研究院收藏的约 800 敦煌遗书（包括写本碎片）进行数字化。

以上所列的各种计划包括了所有敦煌研究院藏敦煌学资源，这些资源大部分都是我院独有的。如果在我院领导的统一协调和安排下，由我们资料中心搭建一个数据库平台，把这些资源合理地利用起来。对学者的使用提供很大的便利条件。

在与国内外机构实现资源共享和数据库之间的互相链接方面，我们希望能用购买使用权限，资源共享等各种方式，在遵守《版权法》、互利互惠的条件下尽可能多的与这些机构合作。来扩大我们的馆藏资料（虚拟馆藏）。

上述四个阶段的工作，第一阶段已经完成，第 2 阶段工作正在进行，第 3 阶段工作我们正在进行先期准备工作。在开展第三和第四项工作的过程中我们会遇到许多的困难和技术难关。比如：标准的统一、信息存储的格式、硬件配置、专业人员的培养等等。我们希望能有一些机构能给予我们支持和合作，共同为敦煌学的发展作出贡献。

参考文献：

本文主要参考了敦煌研究院网站 (<http://www.dha.ac.cn>) 中刊登的以下文章。

- 1、《中美合作“数字化洞窟”的拍摄进入新阶段》
- 2、邵惠莉 沈子君《敦煌遗书数据库简介》
- 3、李鸿恩《图书资料数字化与我们的设想》
- 4、李鸿恩《网络环境下院资料中心信息资源的开发与利用》