

建立国际中国少数民族古籍目录数据库的背景和设计

申晓亭

(国家图书馆善本特藏部少数民族语文组)

二十一世纪将是信息数字化的世纪。电讯技术的新发展、全球网络基础设施的扩大以及新的数字化服务的快速增长对传统的图书馆的功能和发展模式提出了挑战。放眼世界,数字化和网络化的图书目录已经具有了相当的规模并取得了很大的成就,现在,我们若不迎头赶上,就会被时代远远地抛在后边。

该论文谈及的是关于中国少数民族古籍¹目录的数字化和网络化问题。

一、建立国际中国少数民族古籍目录数据库的背景

中国的少数民族古籍文献包括回鹘文、西夏文、蒙古文、藏文、满文、纳西文、彝文、傣文、水文等近 20 种文字,是非常珍稀独特的、引人瞩目的、重要的文献资料。它们是我中华多民族大家庭文化的珍贵遗产,同时也是世界的人类优秀文化遗产。国内的少数民族古籍文献藏书约有数十万种,国外的藏书大约也有数万种。关于国外藏的中国少数民族古籍藏书,近几十年来已经出版了许多很好的书本目录。国内少数民族古籍目录工作近几年来也有很大的发展和成就。主要情况如下:

1、藏文藏书²

日本:《影印北京版西藏大藏经总目录》,四册,大谷大学藏书并监修,西藏大藏经研究会编辑出版,东京-京都,1961年。

《西藏大藏经总目录》,德格版大藏经目录,日本东北大学文学部编,1934年版。

法国:《国家图书馆藏敦煌藏文写本目录》,M. 拉露编,三卷本,巴黎,1939、1950、1961年。

苏联:《布里亚特综合科学研究所藏文手抄本和木刻本目录》,B. Д. 丹达龙编,二卷本,1960、1965年版。

中国:《民族图书馆藏文典籍目录文集类子目》,孙文景、木雅公布等编,第一辑,四川民族出版社,1981年;第二辑,民族出版社,1989年;第三辑,民族出版社,1997年。

2、蒙古文藏书

丹麦:《蒙文手抄本和木刻本目录》,W. 海西希编,哥本哈根,丹麦皇家图书馆,1971年。

德国:《德国东方写本目录:蒙文手抄本、木刻本卷》,W. 海西希编,威斯巴登,1961

① 在“中国少数民族古籍”中,有些民族古籍不仅仅是中国的,同时也是其它国家的,在这里,少数民族语也许就是这个国家的民族语,比如:蒙古文古籍和蒙古语等。

② 参见:冯蒸.国外西藏研究概况(1949-1978).中国社会科学出版社.1979。

年。

俄国：《苏联东方科学院蒙文手抄本木刻本目录》，A. Г. 卡泽金编，第一卷，莫斯科，1988。第二卷，最近。

美国：《华盛顿地区蒙古文手抄本木刻本目录》，D. M. 法夸尔编，《中亚杂志》第一卷第三期（1955）。

中国：《中国蒙古文古籍总目》，三册，乌林西拉，申晓亭等编，北京，北京图书馆出版社，1999年。

3、满文藏书

日本：《东洋文库满蒙文藏书目录》，N. 鲍培等编，东京，1964年。

俄国：《俄国科学院东方研究所圣彼得堡分所所藏满文手抄本木刻本目录》，T. A. 潘编，威斯巴登，2001年。

美国：《美国国会图书馆所藏满洲语文献目录》，松村润编，《东洋学报》，卷57, nos. 1-2，1976年1月。

中国：《全国满文图书资料联合目录》，黄润华、屈六生等编，北京，书目文献出版社，1991年。

4、西夏文藏书

俄国：《苏联东方科学院西夏文宗教文献目录》，E. И. 卡恰诺夫编，京都，1999年。

《俄藏黑水城文献》西夏世俗部分7-11卷，5册，史金波等编，中国上海古籍出版社，1997。

以上所列仅仅是众多目录中的一小部分。

这些书本目录的编辑特点是拉丁转写和本国文字相结合，即：标题、著者、引文用民族文字的拉丁转写记录，叙述文字用本国语言，如英语、德语、俄语、日语、汉语等。这些目录题录的数量多少不一，编辑内容的详略不一，但大多数目录的编辑质量都很高，记录准确，描述详细，有分类，有译名，编有各种索引。

例如：W. 海西希编的《（丹麦皇家图书馆）蒙文手抄本和木刻本目录》，除了题名的转写和英译之外，还有文章开头结尾的转写和英文翻译，除了常规著录之外，还有一些有关的参考说明。日本的《影印北京版西藏大藏经总目录》，有大藏经详细的子目，题名的藏文拉丁转写、梵文拉丁转写、汉文译名或日文译名，以及梵文藏文的著译者。E. И. 卡恰诺夫所编《西夏文宗教文献目录》，题名用的是西夏文原文，行文用俄文。题名有汉文和汉文的俄文拼音。西夏文引文有俄文翻译。这些都是水平都是很高的学术目录。

国内目录值得一提的是《中国蒙古文古籍总目》，它是一部综合性的大型全国联合目录，共题录各类蒙古文古籍文献13115条目。收藏范围涉及全国各类型图书馆、资料室、各地区档案馆、博物馆、寺院等，其中藏书单位180个，藏书个人80位。文献覆盖率约是全国藏书的90%以上，充分反映了国内蒙文古籍的藏书情况。它以蒙古文作为基本著录文字，题名有拉丁转写和汉译名，有详细的附注说明，采用ISBD标准书目著录。

另外，目前，在国家民委的领导下，国内正在开展大规模的《中国少数民族古籍总目提要》工作，要对所有的少数民族古籍文献进行编目，并且做出提要。这是一个很大的工程，得到国家的大力资助，计划 2008 年完成。《总目提要》采用汉文作为基本著录文字，题名附有民族文字及拉丁转写。

以上众多目录的编辑，说明了中国少数民族古籍文献在国际上是很受重视，具有很大影响的，同时也为进一步的目录数字化工作打下了很好的基础。仅目前我们从这些目录中所得知，国外中国少数民族古籍文献的藏书已达 3 万余种，而且，还会有许多藏书散落各处，没有专业人员，没有目录，等待我们去开展协作工作。今天，已经是信息数字化网络化的时代，已经不能仅仅满足书本目录了，面对着那些规模庞大的，现代化的、拥有上千万数据的西文、中文网络目录，中国的少数民族文献信息的数字化不能再等待了，一个国际性的中国少数民族古籍文献数据库应该被提到日程上来了。我希望：与中国少数民族古籍文献工作有关的国家联合起来，共同建立一个国际的中国少数民族古籍文献数据库。

二、建立国际中国少数民族古籍目录数据库的条件

1、中国少数民族文字的计算机平台问题。

近十几年来，国际、国内少数民族语文计算机技术的开发都取得了很大的成就。在中国，于 90 年代初陆续推出了蒙文、藏文、维文、朝文、彝文、壮文、傣文的文字处理系统，到现在许多文种已经有了 Windows2000 的文字处理系统，而且，蒙文、藏文、维文、朝文、彝文等还建立了 ISO 10646 国际标准编码。在国际上，许多国家也都开发了蒙文、藏文、维文、哈文、朝文的文字处理系统。但是，文字处理系统和文字平台还有一定距离。目前，在国内适用于建立中国少数民族古籍目录数据库的多文种平台还不具备，比如：藏文，虽然有了藏文的 ISO 10646 国际标准编码，但是实现它的软件还没有；蒙文虽然有了实现蒙文 ISO 10646 的软件（WPS 蒙文 office），但是，WPS 蒙文 office 在微软 Win2000unicode 平台上的运行是横向，不能竖向运行。国际上，据我们所知，微软 Win2000 系统上没有开发和设置中国少数民族文字，目前国际上也不具备适用于建立中国少数民族古籍目录数据库的多文种平台。

那么，中国少数民族古籍目录数据库的问题怎么解决呢？是要等到具备这些文字平台之后再去做吗？我的回答是：不用等。根据国外编辑中国少数民族古籍目录的传统，可以用民族文字的拉丁转写建库。

2、拉丁转写建库的可行性和科学性。

中国的少数民族文字有三种类型：①拼音文字，比如藏文、蒙古文、满文、傣文等。②形声字：比如彝文、西夏文等。③象形文字：东巴文。

对拼音文字可以用相应的拉丁文字去进行转写，这种转写在字母上是一一对一的，是可以逆行，可以还原的，而且它具有阅读功能。所以现在用拉丁转写建库，待条件成熟时可以自动添加或转换成民族文字，这是很科学的。国外有关中国少数民族文字的研究，通常都是使用它的拉丁转写。国内，随着科学的发展，也有越来越多的人认识并掌握了这一工具。而且我认为即使将来有了多文种少数民族文字平台，作为文字的识读和研究工具，拉丁转写也会有与民族文字共同存在的必要。比如，现在国际上用汉语拼音的地方就很多。拼音文字约占

中国少数民族古籍的 80%以上，采用拉丁转写基本可以解决中国少数民族古籍的建库问题。

对于形声字，可以采用拉丁拼音（或国际音标）和原文同时使用方法。因为象西夏文、彝文这样的形声字，仅用拉丁拼音不容易解决还原问题，所以最好是同时使用原文。没有文字平台，是否可以用扫描粘贴等方法去解决呢？有的古文字无法解决它的读音问题，如契丹文等，只好用汉文或英文译名替代。

象形文字东巴文可以用国际音标来记录它的语音。

对于某些比较特殊而罕见的文字，比如：西夏、东巴、回鹘、契丹、女真等，不仅考虑做目录数据库，还可以考虑做全文的数据库。

3、国际上中国少数民族古籍目录数字化的情况。

目前，除了各国、各馆有自己的网站之外，还有成为世界编目中心的 OCLC、RLIN 等网络系统。系统的网上中国少数民族古籍目录虽然我还没有看到，但是新近出版的民族图书目录及零散的民族古籍目录在网上是可以找得到的。比如美国国会图书馆网上可以看到藏文、蒙古文的目录；日本东洋文库的网上有藏文的数据库。而美国哥伦比亚大学东亚图书馆与我馆最近合作的哥伦比亚大学东亚图书馆所藏的中国少数民族古籍目录将在 RLN 和 CLOT 网上发布。该目录有 100 余条目，文种包括：满文、蒙古文、西夏文、苗文、女真文。该目录采用 USMARC 格式，著录文字全部是民族文字的拉丁转写，同时再用汉文重复一遍。有主题词，有责任者文档。另外值得一提的与中国少数民族古籍有关的是正在进行的“国际敦煌项目（IDP）”。这是一个由英国大英图书馆主持的国际合作项目，它将把世界上所有的敦煌文献数字化，其中包括西夏文、藏文等中国少数民族文献。“国际敦煌项目（IDP）”的开展给了我们许多启发和借鉴。

上述网上少数民族图书目录有一个共同的特点就是使用拉丁转写或文本扫描。

三、国际中国少数民族古籍目录数据库项目设计。

建议成立“国际中国少数民族古籍目录数据库协作中心”。任务为：

国际：为有藏书但没有专业人员的藏书单位提供专业编目服务；与有藏书也有专业技术人员的藏书单位或著有相关专题目录的学者专家协商建立网上国际联合目录。

国内：同样使用拉丁转写进行建库，待具备民族文字平台时自动添加民族文字。争取与有关方面开展协作，可以先从条件比较成熟的《中国蒙古文古籍总目》做起。

具体的技术问题有：

1、规定模式。国际联合目录可以有两种模式：

模式一：按照统一的格式、统一的要求重新录入数据。

模式二：直接利用已经出版的目录，通过 OCR 文字识别技术实现数字化。

2、协调文字的使用。国际联合目录需要协调著录文字中民族文字与英语、汉语及其它藏书国语言的使用。在上述各国所出版的目录中，除了记录题名、著者、引文的民族文字（拉丁转写）外，叙述文字是各不相同的，是使用各国自己的文字。在国际联合目录中，文字问题应该怎样统一呢？

在第一种模式中，有两种选择：（1）叙述文字同时采用三种语言：民族语、英语、汉语。

叙述文字应该尽可能地使用现行的民族文字（拉丁转写），如蒙文、藏文等，然后再将其译成英文、汉文。而对那些现在已经不再使用的古文字，比如回鹘文，除了题名、著者、引文用民族文字（拉丁转写）外，叙述文字用英、汉两种文字。英文是世界通用语言，使用汉语是因为中国少数民族文献与汉语有着广泛的联系。（2）叙述文字用英语。除了题名、著者、引文用民族文字（拉丁转写）外，只用英文做叙述文字，汉文可做题名翻译。

在第二种模式中则尽可能地保留了各国的文字。可以进行加工，建立统一的索引，也可以利用原书的索引进行检索。

3、协调 MARC 格式的使用。目前国际图联有 UNIMARC，中国有 CNMARC，美国有 USMARC。我感觉这些 MARC 主要是格式上有所区别，在著录内容上没有太大的区别，而国际联合目录需要使用统一的 MARC 格式。从中国的角度，从国内文献数量较多的角度考虑，可以采用 CNMARC；但从目前国际上的趋势来看还是使用 USMARC 的比较多。近来，还出现另外一种趋势，即采用 Dublin Core 替代 MARC。

4、确立网络系统。途径有二：（1）与现有的国际联机编目系统接轨，如 RLIN、OCLC CJK，（2）开发新的专门的系统。国际少数民族古籍联合目录对系统的要求，一是国际性，二是兼容多种文字。目前，该目录计划使用民族文字的拉丁转写、英语、汉语，若是采用第二种模式则还包括德语、法语、日语、俄语等，而从长远来看是要添加少数民族文字的，如藏文、蒙古文、满文等。

5、统一拉丁转写方案。目前，对于一种少数民族文字国际国内都存在有多种拉丁转写方案，需要确立一个统一的转写方案。

6、确定著录要求。根据《国际标准书目著录（ISBD）》的要求，考虑各种文字的特点并借鉴已出版目录的编辑经验确定统一的著录要求。

看起来这是一个规模庞大的、复杂的、专业技术要求很高的计划，但也是一个可以从个别做起，然后汇个别为整体的可实施计划。

该项目经费来源途径：向有关方面申请专项资金，争取社会各界的赞助，主办国提供基本资金，各合作单位提供协作资金，将成果商品化等。国内国外对此项目有兴趣、并提供基本基金者都可以成为该项目的主办单位。

该项目的实施可以为中国少数民族古籍文献的数字化和网络化打开局面，促进少数民族文字平台的开发，对这些珍稀文化遗产的资源共享、抢救整理、开发利用具有深刻、长远的历史意义和国际意义。

希望该项目建议与设计能够得到国际、国内的积极反应和大力支持。

除了上述全球性的合作外，还可以根据需要开展局部的、地区性的合作。本人愿意为有需求的国家提供力所能及的服务。