

国家图书馆普通古籍书目数据库建设工作综述

董馥荣

(国家图书馆分馆普通古籍组)

国家图书馆分馆普通古籍书库拥有 160 余万册的典籍文献，收藏范围主要是 1911 年及其以前古典装帧形式的普通古籍，同时也包括 1911 年以后新印的线装书。建立普通古籍古籍书目数据库，尽快实现古籍书目数据的规范化、网络化，是国家图书馆近年业务工作的重点项目。从 1999 年至今，经过近百名工作人员的不懈努力，共整理图书 10 多万部，147 万余册，编制书目数据 24 万余条，预计在 2003 年初将完成全部普通古籍回溯书目数据的制作工作。普通古籍书目数据库包括专著、分析、汇集等多种书目数据类型，是一个综合性的大型古籍书目数据库。

一、建立普通古籍书目数据库的背景要求

国家图书馆入藏普通古籍始于 1909 年筹建京师图书馆时期，经过数十年的不断积累，到八十年代末，普通古籍总量已达 160 余万册，形成了品种繁多、内容丰富、为学术界所重视的古籍府库。普通古籍分类所使用的是 1929 年刘国钧先生编制的《北京图书馆中文普通线装书分类表》。按此表划分，普通古籍共分为目录学、经籍、史乘、地志、传记、古器物学、社会科学、哲学、宗教、文字学、文学、艺术、自然科学、应用科学、总记等十五大类。古籍自身的特殊性决定了古籍编目有较大的难度，又受到各种条件因素的制约，长期以来，普通古籍中只有 40% 能够提供给读者阅览，60% 的书籍处于简编或未编状态，很大程度上影响了读者对古籍的利用，也不利于本馆了解和掌握普通古籍的馆藏情况。为尽早将这部分普通古籍公诸于众，普通古籍组从 1983 年开始进行普通古籍大合流编目工作。至 1997 年，先后完成了目录学门、古器物学门、自然科学门、文字学门、地志门等五大门类的合流编目工作，并正式出版了目录学门、古器物学门、自然科学门、文字学门的普通古籍书本式目录。共合流图书 20 余万册，1 万 5 千多种，占总量的 12% 左右。若按此合流编目方式进行下去，编完所有的普通古籍还需要几十年的时间。这与当今现代化的建设进程是很不相称的。

随着计算机技术的高速发展，图书馆自动化程度日益提高，图书情报书目数据库技术也在突飞猛进。我馆从八十年代末就开始采用机读目录格式进行普通中文图书的编目工作，随后建立了一批书目数据库。普通古籍组在此基础上，从 1995 年下半年开始着手筹建普通古籍书目数据库的调研工作。适逢 1997 年分馆馆舍进行维修改造，到 1999 年才正式开始普通古籍书目数据库的建库工作。

二、建设普通古籍书目数据库的计划构想

由于简编和未编的古籍文献数量庞大，按传统的编目加工方法难以在短期内将所有书籍加工完毕，所以在建库之初就设计了分两步走的方案。第一步：用 3 至 5 年的时间完成基础书目数据的编制，以揭示馆藏为主要目的。著录内容比较简约，主要包括题名、责任者、版本、载体形态、分类、索书号等要素。第二步：在第一步工作的基础上，进一步对书目数据

进行细化整改，实现对古籍文献的深层次揭示，涉及到分类方面的，即将十五大类分类与四部分类进行对照，以及与规范数据库挂接等等。

三、建设普通古籍书目数据库的基础工作

建设普通古籍书目数据库的基础准备工作主要包括：对编目软件进行测试、试用并提出修改的意见、制订新的古籍编目工作流程、准备书目数据制作的工作标准、对古籍编目人员进行培训等四个方面。

（一）在编目软件方面，我们使用的是文津图书馆系统软件。从试用到最终系统软件古籍编目模式的确定，经过了相当长时间的磨合。目前文津软件基本能够满足我们古籍编目的使用要求，可以实现分类、题名、责任者、出版年、出版者、索书号、其他题名等多种途径检索。

（二）工作流程是我们遇到的新课题。根据实际情况，在数据制作中我们采取了两种方式：对已经完成合流编目的五大类目录，因为编目基础工作做得比较好，我们采取将卡片目录直接转换成 MARC 数据的方法；对未经合流编目的书籍，在现有古籍编目人员还不能熟练地进行文字录入的情况下，我们采取了填写工作单的方式编制书目数据，请专门的录入人员进行录入，再由编目人员对录入数据进行校对。这样做可以充分地扬长避短，既提高了古籍书目数据编目、加工的效率，也保证了数据的质量。通过近 4 年的工作实践，确实取得了良好的效果。

（三）古籍书目数据的工作标准。1999 年我们制作数据的标准是我馆编制的《古籍机读目录格式字段表》。该表规定了古籍机读目录所使用的字段、子字段及其记载各项古籍书目信息的格式。2001 年 10 月由中国国家图书馆编制的《汉语文古籍机读目录格式使用手册》正式出版后，数据制作标准为该手册。著录规则按照国家标准《古籍著录规则》执行。著录用文字根据中国图书馆学会学术研究委员会古籍整理专业委员会 1998 年对《古籍机读目录格式字段表》的讨论意见，采用了规范的简体字，当时主要考虑到字库原因，以及与馆内外其他文献类型书目数据库统一合库的问题。

（四）对古籍编目人员进行培训。从传统的手工编目到现代化的书目数据制作、加工，这是一个跨度非常大的转变，对古籍编目人员来说也是一次挑战。机读目录格式与传统的卡片目录格式之间存在着很大的差异。机读目录格式为了使书目记录规范、标准，便于计算机识别，对著录内容的划分和标识是非常严格的，而古籍本身具有多样性和不确定性，加大了编目著录的难度，如何准确地按照机读目录格式要求记录著录内容，是我们在编目中经常遇到的问题。字段的选取和指示符的运用，更是我们早期编目中的难点。针对这个问题，我们对编目人员进行了多次培训，一方面学习《中国机读目录格式使用手册》，另一方面还制订了《〈普通古籍书目数据工作单〉制作要求》、《普通古籍书目数据录入校对要求》等一系列的学习材料和工作细则，首先使编目员能在很短的时间内学会根据要求填写工作单，再通过进一步的培训，使编目员逐渐摆脱了用卡片套用机读格式的思维方式，真正理解了机读目录格式的含义，做到自觉地根据古籍机读目录格式的要求填写书目数据工作单。

四、普通古籍书目数据格式的逐步完善

由于此前没有制作古籍书目数据的经验,所以普通古籍书目数据库建设工作是在不断摸索,不断改进的状态下进行的。我们的古籍书目数据格式也在具体的实践工作中逐步完善。4年中我们根据机读目录格式的规定,先后对数据格式进行了多次的调整。最初数据格式与普通中文图书数据没有什么大的区别,只是在内容和字段的选择上有所不同。2001年,根据《中国机读目录格式使用手册》(修订版),我们增加了适用于古籍的字段,并且对古籍与民国以来线装书分别进行编目。前文已经提到,普通古籍书目数据库中包括1911年以后的新印线装书,开始做书目数据时在字段格式上没有加以区别,在增加了古籍专用字段后,我们在定长编码字段上对古籍和新印线装书加以区分,古籍启用了140、141字段,新印线装书仍沿用105字段,这样做不仅突出了古籍的特点,也有利于今后古籍和新印线装书书目数据的交流和共享。

另一方面,表现在连接字段的应用上。比如丛书子目数据,开始我们只用了225丛书项,出处内容在300字段描述,这样做书目数据在卡片显示时与手工做分析片很相似,但并不符合机读目录的格式要求,我们对此进行了改进,用461字段做连接,并增加了205、210和215字段,在300字段取消了出处内容。开始用461字段时也没有区分层次关系,所有子目均上连到总集(丛书),子目的从属关系表现得不清晰,现在我们使用了462字段,丛书内部的各种层次关系都能得到体现,充分发挥了机读目录格式的优势。与此相似的还有合刻、合订书的书目数据,开始做的时候只在300字段说明,书目数据之间没有用到连接字段,现在我们用了423、481、482字段,使有关书目数据之间的关系可以一目了然。

五、有待进一步研究解决的问题

虽然我们的普通古籍书目数据库工作取得了比较大的进展,但不可否认还存在着很多不尽如人意的地方。表现在机读目录格式的运用方面,虽经多次调整,但很难达到形式与内容的完美统一,有些内容仍然找不到合适的字段,这是应该继续探讨的问题。另一方面表现在书目数据本身需要进一步完善,如连接字段虽然已经使用,但并没有实现真正意义上的连接,还有系统外字的问题,这些将随着计算机书目数据库技术的不断发展而逐步解决。另外,由于各个阶段使用的数据格式不尽相同,造成各部分数据之间的差异,以及由于缺乏规范数据库所造成的数据检索点不统一等问题,都是我们今后进一步整改数据的工作重点。

普通古籍书目数据库的建立,仅仅是我馆古籍数字化初始阶段的工作成果,并不是我们的最终目的。它的建立,为今后古籍数据库的建设打下了良好的基础,同时也培养了一批熟悉机读格式的古籍编目人才,这也是普通古籍书目数据库工作的重要收获。

总之,普通古籍书目数据库的建设工作,是一个不断地学习和改进过程。在这个探索的过程中,我们走过不少弯路,也取得了宝贵的经验。国家图书馆普通古籍书目数据库的建立,不仅对本馆基础业务工作有着重要的意义,相信对于全国的古籍文献数字化工作也将起到有益的作用。