

浅述古籍出版项目数字化标准和操作规范

郭静

古籍数字化是我国悠远璀璨的传统文化与文明传播和传承的有效途径。古籍数字化的实施主要是利用缩微摄影、拍照或扫描技术采集图像，建立影像数据库或进行影印出版，还可以在此基础上通过文字自动识别（OCR）或人工录入等进行元数据著录，生成文本数据库或进行文字出版。影印出版能相对更好地反映古籍原貌、更接近于原作的“存真”处理方式。这种方式保存了文献典籍的真实形态，未经转写，故而避免一些文字和内容上的误读误解，可靠性较高，满足了相关学术研究的需要。

一 数字化背景下的古籍出版项目主体

古籍的数字出版不仅是简单地复制古籍，也属于古籍整理和学术研究的范畴。从整体来看，出版项目的选题、书单确定、版本确定到最终的制作、发行，均由出版社完成。目前很多专业的古籍出版社都已经具备较为完备的古籍出版项目流程和营销体系，一些高校研究单位或个人科研项目也进行了相关数字化研发。但由于最重要的善本古籍原书资源保存于图书馆，数字化过程中遇到的具体问题需由出版社、图书馆双方主体合作解决（见图 1）。如不同时期、不同版本的古籍在装帧、版式、墨色、纸张、字体上，甚至内容上都有不同，因而在具体数字化出版过程中，应特别注意版本的考订和选择，并及时与出版方编辑沟通。

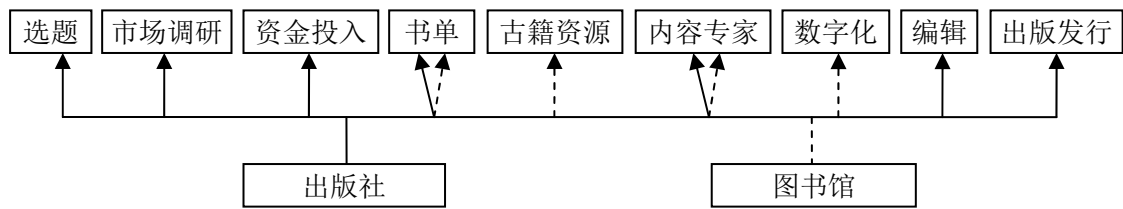


图 1 古籍出版项目主体合作流程

二 古籍出版项目数字化标准

古籍是一种特殊的文化载体，对其进行数字化也有相应特殊的处理要求，而并非简单的复制。不同于现代书籍，古籍由于存在线装、蝴蝶装、经折装等不同装帧，在页面划分上也存在筒子叶、半叶和双半叶等不同的技术方式。数字化时根据古籍的装帧、版面情况，选定如何扫描书叶。古籍纸张由于年代久远，有的纸质已变得酥脆，有的书叶粘连，这些破损情况都需特别注意，不适宜扫描的应即刻中止。

目前在缩微摄影、拍照和扫描三种图像采集方式中，以扫描最为常用。扫描是使用不同幅面专业书刊扫描仪对古籍文献逐叶扫描，存储为计算机图像文件。特点是图像质量高且稳定，受环境因素影响较小，适合影印出版。不足之处是生产效率受图像分辨率及操作员工熟练度影响。这里主要讨论古籍扫描的数字化标准。

1. 页面：扫描仪幅面尺寸内的藏品，多以一个筒子叶或双半叶采集；扫描仪幅面尺寸外的册页类藏品，至少采集半叶；扫描仪幅面尺寸外的其他形制藏品，采取从左至右—从上至下—从右至左逐块顺次的方式扫描，再由后期拼图处理。分画幅扫描时，各扫描区域边缘应有 3 厘米（含）以上的重复扫描区。

2. 图像质量：按 1:1 比例扫描。图文扫描要清楚完整，无透字现象，真实反映原书霉斑、水渍等情况。扫描时需加放色标，使成像有对比度，更好地还原古籍原始面貌。扫描一般书籍，只需在封面、题跋、卷端、封底等页面放置色标；扫描拓片、法帖、地图等图像资料时，需要每拍都放置色标。图像端正，书叶平整。出版项目现在常见图像质量具体要求如下表：

位深	RGB 模式 24 位
分辨率	300—600ppi
文件格式	Tiff（LZW）/ JPG
色调再现	ICC 配置文件

表 1 古籍出版项目图像采集质量要求

一些要求不高的出版项目书影也可以存储为分辨率 300ppi 及以上的 JPEG 格式。如果藏品是纸张较薄的散叶、手稿，需在书籍承托板的黑色绒布上垫上白板或宣纸衬纸，保证扫描色彩不失真。另外，设定自动裁剪时注意尽量使书叶充满整个画面，不需留较多空白，不可裁掉天头地脚。

三 古籍出版项目数字化操作规范

经出版社确定数字化用书清单并签署出版协议后，出版项目的具体操作便交由数字化工作室完成。大体上出版项目的数字化流程如下：

1. 前期准备

（1）核对书目：确认藏品书号、版本、册数、馆藏地等信息。

（2）提书：将核对无误的书目传递给库房工作人员。库房工作人员按照项目要求筛选、提取古籍。数字化工作室的藏品管理员需认真细心与库房做好交接工作，清点提取书目的种数、册数，做好书库的藏品出入库登记。

（3）确认书况：开始加工前需确认藏品保存情况，如有较为严重的破损、虫蛀、夹框、掉叶等情况，应及时进行详细记录，并与古籍修复人员确认是否适合加工，是否需要先经拆线、局部修复等处理。书况检查记录可为数字化加工以及后续修复工作提供参考和依据。

（4）衬纸：善本古籍大多纸张较薄，扫描时有可能出现透字现象，影响影印效果。加工前需检查纸张是否透字，如透字或虫蛀较为严重，应垫入大小软硬度适中的白色衬纸再进行加工。如果书籍本身较厚，衬纸时需注意不能将书叶直接卷折压在下面，需用镇纸压住翘起的书叶以保持平衡，以免造成折痕损伤书籍。另外，对于书背钉线过紧的书籍，衬纸过厚会导致夹框。此时应视具体情况，先衬 1/2 或 1/3 册左右书叶，加工完撤掉衬纸，再继续衬剩余的书叶。

（5）测量书籍大小：加工前用精度较高的透明直尺测量藏品的开本大小、板框、天头和地脚的长度，以满足后续出版的需求，以及完善书目信息登记内容。此步骤也可在加工完成后进行。

2. 图像采集

（1）启动设备：启动扫描仪预热。启动 PC 机，打开扫描系统控制和成像处理软件。扫描前将玻璃面板擦拭干净，避免灰尘和杂质影响扫描图像效果。

(2) 放置藏品：书籍要平整得摆放，书页如果起伏不平会导致扫描出来的图像变形。将书脊固定在左右两个承托板之间，然后用竹启子等光滑的工具将书页轻轻平整。

(3) 文献存储：设定扫描图像储存路径，建立文件夹。文件夹结构一般为：项目名称>藏品号+藏品名称>册序（1、2、3）。在扫描系统控制软件中设定好 PPI 数值和锐度等参数。

图像采集过程中还有些需要特别注意的事项。如书页表面有其他粘贴物件时（浮签等），应先将书页与粘贴物一起（即粘贴物覆盖于书页之上）扫描一拍，粘贴物纸张薄的话需衬纸于书页上。然后将粘贴物掀开，使之不遮挡书页上的文字图画，再次扫描一拍。应特别注意的是浮签等一定要归回原处，保留藏品原始信息。

另外，扫描完毕以及暂时不需扫描的藏品应放在数字化工作室的保险柜中暂存。数字化工作室无藏品管理员看护时 also 需将古籍暂存于其中。

3. 图像质检

对扫描完的图像进行质检是数字化工作中非常重要的一环，前承图像采集，下接数据管理。质检主要是检查扫描图像是否完整、正确。一般使用 ACDSee 数字图像处理软件进行浏览和修改，较大的图像可使用 Photoshop。具体操作为：质检人员逐叶翻检古籍原件，依据图像质量要求与采集好的图像一一核对。如发现问题，夹好纸条、交还扫描人员进行重扫或补扫，再次质检合格后将重扫或补扫后的图像代替原图像。

常见的质检问题主要有：（1）书页漏扫；（2）文字模糊、锯齿、发虚；（3）色彩异常，有彩点、彩线；（4）书页透字；（5）有杂毛、碎屑等异物；（6）图像扭曲、变形；（7）书页重复；（8）浮签遮字没有掀起；（9）图像文件损坏、图像内容不完整；（10）夹框、夹字严重；（11）页角翻折；（12）切边过多。

检查图像无问题以及补扫、重扫完成后，按照命名规则对图像进行批量重命名。以一个文件夹（一册）为单位，命名规则为：书号+书名+第 X 册+流水号（001 至 999）。确认无误后，由藏品管理员清点并交还藏品，与库房工作人员做好交接。

4. 数据管理及提交

(1) 数据管理: 将图像数据由 PC 机转移至存储介质中。由于 Tiff 格式图像容量较大, 目前多用大容量移动硬盘来存储, 或再用光盘备份。确定好存储地后, 及时将藏品自身信息、扫描信息和存储信息登记至数字化工作单上, 做好数据管理工作。

(2) 数据提交: 向出版社提交数据时应按照对方的具体要求进行图像处理, 如裁切黑边、调整分辨率、转存为其他存储格式等, 将数据刻录光盘或拷贝至 U 盘、移动硬盘交付, 做好交付记录。

古籍出版项目的数字化是一项需要各方人员配合, 并要对流程中各步骤认真负责的精细工作。对数字化工作室的人员来说, 不仅需要熟悉其中各项质量标准 and 操作规范、提高自身专业素养, 工作中还要时刻以保护、爱惜藏品为重, 珍惜手中历经千年不可再生的文化资源, 从而创造出合格优秀的古籍数字化产品。

参考文献

耿铭《古籍影印的生命在于“存真”》,《攀登》2009 年第 3 期, 101—108 页。

王立清《关于多元古籍数字化主体的探讨》,《图书馆学研究》, 2011 年第 4 期, 53—58 页。