

数字化过程中的古籍保护问题

——“明别集丛刊”项目的工作体会

□ 栾伟平

摘要 在“明别集丛刊”项目工作中,认识到数字化过程中的古籍保护问题,主要是:甄选好底本,选择合适的扫描仪器,采用合适的扫描模式,以及采取合适的图像处理方式,等等。

关键词 古籍数字化 古籍保护 “明别集丛刊”项目

“明别集丛刊”是北京大学图书馆和黄山书社合作的大型古籍影印项目,是迄今为止规模最大的明人别集丛书,选取明代近 2000 位作者的诗文集,占现存全部明代作者及其别集的三分之二强,基本囊括了明代重要作者的重要别集。该项目选取了 1500 多种原书,200 多种胶卷,共扫描 50 多万筒子页(包括胶卷还原),其选目与扫描工作已经完成。笔者在北京大学图书馆古籍部就职,参与了选目工作,并承担该项目的复核和后期处理工作,对于数字化过程中的古籍保护问题,有一些粗浅体会,今整理如下。

1 数字化前的底本甄选工作

选择好的底本,这是古籍数字化的重要前提。据笔者参与此项工作的体会,在选择底本时,需注意以下几个问题:

1.1 底本应有较大版本价值

底本应选取刻印年代较早、流传稀少、具有代表性、是此后诸本的祖本,等等。这个问题主要属于版本目录学和文献学的范畴,这里不展开讨论了。

1.2 版本鉴定应保证准确性

注意核查,给出所选底本的正确版本信息。在确定某一古籍的版本时,最好参考《中国古籍善本书目》,与其中所载条目相对照,看是否相符。出于历史原因,我馆古籍目录存在着一些错漏或不准确的地方。如果与《中国古籍善本书目》不相符,需细查原因——是《中国古籍善本书目》有误,还是我馆的目录不准确?还是各有正误之处?一般说来,《中国古籍善本书目》的版本著录正确性较高,但也有缺

陷。如:《中国古籍善本书目》10947 条“荣木堂集三十二卷 明陶汝鼎撰 清顺治康熙递刻本”,本馆原书为三十七卷,自然要遵从原书。再如:《中国古籍善本书目》9859 条“范文忠公初集十二卷 明范景文撰 年谱一卷 清王孙锡撰 清康熙四十年范毓秀、范绳祖等刻本”。关于刊刻年代,我馆目录作“清康熙十二年”,据原书序跋,该书刻于康熙十二年,康熙四十年序为后加,故不从《中国古籍善本书目》。“明别集丛刊”中所采用的约 700 种善本书,笔者曾一一查阅其在《中国古籍善本书目》中的著录。每当遇到与《中国古籍善本书目》不符合之处、且从原书中无法判断正误的情况,笔者尽量去我馆之外的藏馆查阅原书,如果没有条件看到原书,则寻找书影,来对照是否同版。书影的获得,基本有三种方式。第一,本馆所藏影印本,如“四库存目丛书”、“四库禁毁丛书”、“四库续修丛书”、“丛书集成”,等等。第二,某些图录类图书,如《中国版刻图录》;国家珍贵古籍名录图录,已经出版的有《第一批国家珍贵古籍名录图录》、《第二批国家珍贵古籍名录图录》;各大图书馆的古籍善本图录,如《中国国家图书馆古籍珍品图录》、《南京图书馆馆藏图录》等;古籍展览图录,如《册府撷英——国家珍贵古籍特展图录(2009)》、《楮墨芸香:国家珍贵古籍特展图录(2010)》等。第三,相关网上搜索。三者结合起来,对版本确定有很大帮助。

1.3 同一版本有不同复本者

在选目过程中,如遇到同一版本有不同复本者,应综合考虑古籍内容、印刷状况、装订情况、纸张条

件,等等,尽量选择内容完整、无缺页或者缺页较少、刷印清晰、断版、补版较少的书。不能选中缝较紧的书,如果中缝太紧,扫描时,靠近装订线的文字容易变形,甚至有可能根本扫描不上,需要拆书,才能达到理想的扫描效果。但如果两种书都中缝紧,而其中的一种有包角的话,则尽量选择没有包角的书,因为包角还原起来较困难。总之,想尽办法避免拆书,尤其是尽量不拆有包角的书,以保持古籍的原貌。

纸张方面,尽量选用白棉纸或皮纸等较厚而坚韧的纸张,尽量不选薄竹纸,特别是薄而脆的黄色竹纸古籍。从保护古籍来看,在扫描过程中,要翻页,要按压,竹纸的书更容易受损。从扫描效果看,同样的内容,在黑白扫描的情况下,白棉纸或皮纸等的扫描效果更清晰;另外,古籍纸张太薄的话,容易透出反面的字,在黑白扫描的情况下,这些透字会变成图像上的底灰,处理起来较困难。至于透字问题,有用下列方法处理的:即扫描时加衬纸,夹在筒子页内,把两个半页隔开,从而避免透字。不过,古籍纸张多较松脆,衬入的纸张易把书口划破。所以,这次“明别集丛刊”扫描时,我们没有采用衬纸。透字问题则希望扫描公司来处理。

选书时,如果仅选用该书的部分内容、需要插入纸签标明起止的话,尽量用软而薄的手工竹纸,千万不能用厚而硬的机械制纸,以免损坏古籍原本。另外,有些宣纸虽然是手工制作,但比较厚而硬,也尽量不要选用。纸签上的说明文字尽量用铅笔,以免污染古籍原本。

1.4 卷页单一定要做好

选择好版本后,一定要做卷页单,也就是页次记录单,逐页记录序跋、卷次、附录等,计算总页码。卷页单是扫描的依据,要详细记录缺页、半页等情况。如果古籍的某些页有损坏,也要记录下来,这不但是为了检查图像扫描的准确性,也是为了在扫描结束后,与扫描公司交接古籍原本时,作为点收依据。当然,也是为了以后的修复需要。卷页单是非常重要的,“明别集丛刊”项目采用了部分胶卷,我馆请扫描公司还原胶卷时,未提供卷页单,导致还原出了错:有时好几种书在一个胶卷上,可能只有一种书为我们所用,而扫描公司的操作员分辨不出来,或者还原得太多,或者只还原了一部分,或者根本就没有还原。结果,只好在提供卷页单后,重新还原。另外,

需提醒扫描公司的操作人员,即使卷页单有错误(比如计算总页码错误,或者记录个别卷次的页码错误等),也不许擅改,一定得告知参与选目的相关工作人员后,才能改动,以避免造成扫描图像错误。

2 图像化过程中需注意的问题

2.1 尽量保证不损伤古籍

为了保护古籍,和扫描公司交流好非常重要。因扫描公司人员流动大,操作员的素质与水平参差不齐,需要在扫描前提醒他们一些扫描原则。古籍扫描与普通书扫描有很大不同:古籍一定要轻拿,轻放,轻翻;而且,一定不能像扫描普通书那样,用带有螺旋纹的指套,因指套上的螺旋纹易刮伤古籍纸张。在进行大型数字化项目的时候,扫描公司一般会派一个组长。这位组长一定得认真负责,如果事事处处只顾及扫描速度,而不管古籍是否受损、扫描效果是否良好,这样的组长不能用。如对扫描组长不满意,我们可以督促扫描公司换人。如果操作员不爱护古籍,我们也要请他离开。最好和扫描公司就古籍保护问题,签订一份合同,详细规定古籍扫描有哪些规矩,违反了的话,或者损伤古籍的话,需要承担什么责任。另外,要告知扫描公司,没有取得我们同意,不能私自拆书。又,某些古籍的封面最初印刷时,覆有薄的透明纸保护,须告知扫描公司,尽量不要撕开覆纸,可用彩色扫描,如果彩色扫描也不清楚的,宁可不扫描封面,也要保持古籍原貌。

2.2 确保扫描效果良好

这也是立足于保护古籍的需要。如果扫描效果不好,又无法进行后期处理的话,很可能需要重新扫描,而多次扫描原书,对古籍损害很大。古籍扫描效果是否良好,与扫描仪器、扫描参数设置、扫描的颜色模式等等,都有关系。

“明别集丛刊”项目采用了3种图像化设备:柯美(MINOLTA)PS7000扫描仪、数码相机、虹光(Avision)FB6080E扫描仪。FB6080E是平板式扫描仪,其图像效果在3种设备中最好,但因为与古籍直接接触,可能对纸张松脆的古籍有损,而且扫描速度较慢。其他两种复制设备速度较快,扫描时只需页面朝上,平放于托书台,但也需要操作员手动翻阅古籍,并按压,对古籍还是可能有微损。而且,有的古籍较厚,用后两种复制设备时,扫描图像可能会变

形。最好采用既能与古籍零接触,又能保证图像效果的扫描仪器。

除了选择好的扫描仪外,还要调整好相关扫描参数,如分辨率、焦距、亮度、对比度、图像文件格式等。如果设置不当,在扫描过程中,容易造成图像细节丢失,即使后期通过图像软件进行处理,也无法获得理想效果。“明别集丛刊”项目采用的分辨率是400dpi,文件格式是tiff。

颜色模式的选择。在古籍图像化中,通常有彩色、灰度以及黑白3种扫描类型。“明别集丛刊”项目基本采用黑白扫描。笔者发现,黑白扫描有较大局限性,在某些情况下并不适合。首先,还原胶卷时,不适合用黑白模式。“明别集丛刊”先是用黑白模式还原胶卷,但发现效果很不理想,很多时候,底灰重得挡住字了,或者底灰遮住了版心。后来,只好换用灰度模式,效果好多了。其次,黑白扫描模式无法完整再现古籍原件上的红色字迹或图像。批校本、评点本、彩色套印本等等,最好采用彩色扫描。比如:“明别集丛刊”选用的《丹崖集》,该书是清研古楼抄本,李盛铎批校。李用朱笔校改了上百处,用黑白扫描模式,校改痕迹非常模糊。李盛铎为此本所做的朱笔跋语,“癸丑霜降后五日,假友人藏钞本校,改正如右”,用黑白模式扫描不出来。由于黑白扫描失真得厉害,该书换用了彩色扫描。再如:日本延宝五年刻本《永觉和尚禅余外集》,书上有大量的红色圈点,书眉及行间有大量朱批,用黑白模式扫不清晰,只好换用彩色模式。另外,四库底本一般在首页钤有“翰林院印”的红色满汉文大方印,用黑白模式扫描不清。有的四库底本有红色眉批,还贴有飞签,部分飞签上钤有四库分校官的朱印,也在黑白模式下,无法体现。此外,版画不能用黑白扫描,即使是黑白版画也不适合。因为黑白扫描层次不丰富,无法体现出版画的细腻笔调。黄山书社准备用单黑印刷“明别集丛刊”,所以,还得把彩色图像转化成黑白图像。因此,在用彩色模式扫描个别书的时候,要注意颜色深浅,颜色过深,转化成黑白图像后,有些字会被底灰遮住;颜色过浅,转换成黑白图像后,又不太清晰。

笔者认为,彩色扫描是最保真的方式。但彩色扫描速度慢,而且比黑白扫描总额贵。因此,采用哪

种扫描模式,要和出版社多沟通,问询出版社需要达到什么扫描效果,并灵活掌握。最好先试扫描一部分,找出最好的扫描方式后,再进行大规模扫描。

3 图像化之后的处理问题

扫描完成后,需要做图像后处理,在以下两个环节容易出问题:

3.1 拼版环节

扫描完成后,需要对两个半页进行拼接,形成一个筒子页。如果操作员不认真,容易出现以下错误:左右拼版颠倒,以致版心到了边上;缺半版;左右两个半版有部分叠加,从而挡住版心;左右拼版歪了,以致两版之间出现夹角。需要对扫描公司特别说明的是:有的筒子页半页有字,另半页是空白,这半个空白页也必须拼上,不能删除。否则,容易把这半页和下一个筒子页的前半页拼在一起,从而导致整本书的拼版错误。

3.2 裁切环节

扫描公司在处理图像数据时,会对图像边缘进行裁切,以去除多余部分。不特别说明的话,扫描公司一般默认为以版框为裁切依据,也就是说,去掉天头、地脚以及左右边框外的内容。但是,有很多古籍是评点本,评点者在版框外写评语;还有的收藏者,在版框上或者在版框外钤印。以上内容如果被裁切掉了的话,古籍图像就失真了。应告知扫描公司,以古籍原书页面的边缘为裁切的依据。

以上是笔者参与“明别集丛刊”项目中,关于古籍数字化与古籍保护的一些初步想法。挂一漏万,抛砖引玉,希望得到大家的指点。

参考文献

- 1 中国古籍善本书目编辑委员会编.中国古籍善本书目(集部).上海:上海古籍出版社,1998
- 2 北京图书馆编.中国版刻图录.北京:文物出版社,1960

作者单位:北京大学图书馆,北京,100871

收稿日期:2012年11月17日

(转第121页)