

中医药古籍数字化工作手册

（第一版）

国家中医药管理局科技司

2023年2月

目 录

引 言.....	1
1 范围.....	3
2 规范性引用文件.....	3
3 术语和定义.....	3
4 古籍数字化工作流程.....	5
5 古籍数字化加工准备.....	5
6 古籍元数据著录.....	8
7 古籍图像数字化.....	13
8 数据命名.....	16
9 数据验收.....	19
10 数据发布利用.....	19

引 言

中医药古籍承载着中华民族特有的精神价值、思想智慧和生命知识，蕴含着丰富而宝贵的原创思维、独特理论和实践经验，是养生保健、防病治病理论与方法的宝藏，更是中医药科技创新和学术进步的源泉。

古籍数字化作为保护与传承中华珍贵典籍的重要手段之一，能够真实、清晰地反映古籍原貌，促进传统典籍文化传播，有效解决古籍保护和利用的矛盾。为进一步加强中医药古籍保护与服务，2023 年，我局在中央财政转移支付资金项目中实施“省级中医药古籍数字图书馆建设项目”。

为做好全国中医药古籍数字化工作，在参考国家古籍保护中心编制《古籍数字化工作手册》基础上，形成了《中医药古籍数字化工作手册》（第一版）（以下简称“手册”）。本手册适用于本项工作相关的各单位。

本手册用于指导汉文和少数民族普通形制中医药古籍的数字化。

本手册为试用本，随着古籍数字化工作的全面开展，我们会面临新的古籍数字化问题，同时对古籍文献的加工、存储、查询、利用等方面的新需求也将不断地提出。希望各古籍收藏单位详细记录并及时反映手册使用中的问题，我们将根据试用情况对手册进一步修订完善。

1 范围

本手册规定了普通形制古籍和缩微胶片以及数字图像主文件和衍生文件的数字化的主要技术和管理要求。

本手册适用于古籍数字化加工,包括古籍数字化加工准备、古籍元数据著录、古籍图像数字化、数据命名规则、数据提交、数据验收和发布管理等。

2 规范性引用文件

GB/T 21712 古籍修复技术规范与质量要求

GB/T6159.1-2003 缩微摄影技术词汇

GB/T 31076.1 汉文古籍特藏藏品定级第1部分:古籍

GB/T 31219.2 图书馆馆藏资源数字化加工规范第2部分:文本资源

WH/T 66 古籍元数据规范

WH/T 91 汉文古籍集外字描述规范

GB/T22113-2008 印刷技术 印前数据交换 用于图像技术的标签图像文件格式(TIFF/IT)

GB/T23286.1-2009 文献管理 长期保存的电子文档文件格式 第1部分:PDF1.4(PDF/A-1)的使用

ISO15444-1:2004 信息技术 JPEG 2000 图像编码系统

ISO32000-1:2008 文献管理 便携式文档格式 第1部分:PDF 1.7

DB34/T4333-2022 古籍数字化工作指南

3 术语和定义

3.1 古籍

古代典籍的简称。主要指1912年以前书写、印制的典籍。

3.2 数字化

运用信息处理技术将文献转化为计算机可识别和处理的数字信息的过程。

3.3 书叶

按文稿顺序排列的书写、印制的单张纸叶。

3.4 版框

书叶正面图文四边的围栏。

3.5 版心

书页左右对折的正中、在折叶时取作中缝标准的条状行格。雕版印刷的书籍版心通常印有书名、卷次、叶码，有的还印有一版文字总数、刊刻机构以及刻工姓氏等。

3.6 天头

图文或板框上方余幅。

3.7 地脚

图文或板框下方余幅。

3.8 数字图像

由数字组成的图像。用二维矩阵描述，表示实物图像的整数阵列。

3.9 缩微胶片

条或卷形状的缩微品。

3.10 像素

构成数字图像的最小单位。通常用若干不同色彩像素排列而组成的矩阵来表示数字光栅图。

3.11 分辨率

图像中存储的信息量。可以有多种衡量法。

注：图像分辨率和图像尺寸一起决定文件的大小和输出质量。该值越大，图像文件所占用的磁盘空间也越大，打印或修改图像等操作所需时间也就越多。

图像分辨率以比例关系影响着文件的大小，即文件大小与其图像分辨率的平方成正比关系，图像分辨率也影响到图像在屏幕上显示的大小。

3.12 点/英寸

扫描仪（打印机）在水平方向和垂直方向上的每英寸都能扫描（打印）的点数（dpi）。

3.13 缩（小比）率

缩微影像尺寸与相应原件尺寸的比例关系。

注：缩小比率总是小于1。

3.14 放大倍率

放大影像尺寸与原缩微影像尺寸的比例关系。

3.15 无损压缩

去掉或减少数据中的冗余，但这些冗余信息可以重新插入到数据中。无损冗

余压缩是可逆的过程，也称无失真压缩。

TIFF格式，灰度和彩色的色彩模式下，无损压缩为LZW方式。

3.16 有损压缩

允许一定程度的失真，可用于图像、声音、视频等数据的压缩，压缩比可达到几十倍甚至上百倍。有损压缩是不可逆的过程，损失的信息不能再恢复。

3.17 图像拼接技术

将多幅来自同一场景的具有一定重叠区域的小尺寸图像，通过软件系统拼合成一幅大型的无缝的图像。

3.18 档案典藏级

用字母A表示，是古籍数字化图像的主文件。用于数字图像的长期保存，可作格式转换和复制的母本。以不压缩或无损压缩方式存储。

3.19 复制加工级

用字母P表示，通过主文件转换产生的的衍生文件，用于加工复制各种精度、大小的屏幕浏览图像的母本文件。供专家、合作伙伴及专门组织成员网上有条件权限的访问。有较高的精度和较大的尺寸。

3.20 发布服务级

用字母D表示，通过衍生文件的复制加工级转换生成。因资源用途和使用对象、使用场景不同应遵从规定的技术要求操作处理。

4 古籍数字化工作流程

古籍数字化工作步骤依序为：古籍数字化加工准备、古籍元数据著录、古籍图像数字化、数据命名、数据验收、数据发布利用。

5 古籍数字化加工准备

5.1 古籍藏品出库清点

古籍藏品出库后，进行初步检查。数字化加工人员按照古籍库房指定的时间、地点取书，办理出库手续后，在专门人员的监护下直接送到工作场地暂存。存放地点、温湿度等应符合古籍保护的要求。

数字化加工人员应认真阅读随书提交的出库工作单，对照工作单核对题名、部次、册数，对工作单中标注的缺叶、错叶、重叶以及其他提示要逐一核查。

5.2 书况检查与登记

古籍数字化书况检查是通过对古籍原本的逐叶翻检，了解古籍每叶的保存状

况，为古籍保护与修复提供依据，为数字化的后续环节提供指导。书况检查时，详细记录古籍状况，同时，结合古籍的版本、版式、装帧形式、中缝宽度、透字等情况，确定古籍数字化方式和使用设备。

在书况检查过程中完全按书页原貌统计总叶数，记录古籍的缺叶、漏叶、错叶、重叶、空白叶等情况，注明是否需要补配、去重、变更叶码顺序等，登记扫描或拍摄的叶数，形成较为完整的书况检查记录。书况检查记录将作为数字化加工、质检、标引等后续环节的参考和依据。

5.3 缩微胶片状况检查与登记

数字化之前，应对缩微胶片进行检查，判断胶片是否适合数字化。如不符合下列情况，可先对缩微胶片进行处理后再进行数字化加工，以确保数字图像的质量。

5.3.1 缩微胶片物理形态无卷曲、变形、脆裂、粘连、乳剂层脱落等情况。

5.3.2 缩微胶片密度、解像力等技术指标应满足以下规定：

- (1) 原底片背景密度值为 0.9~1.2。
- (2) 第二代负片背景密度值为 0.9~1.2。
- (3) 各代正片背景密度值应不高于 0.25。
- (4) 在同一画幅内，不同位置影像的背景密度差不应超过 0.40。
- (5) 各代缩微品的最低解像力应达到下表所给的数值：

缩（小比）率	二号解像力测试图		
	第一代	第二代	使用拷贝片
1 : 7	11.0	10.0	9.0
1 : 8	10.0	9.0	8.0
1 : 9	9.0	8.0	7.1
1 : 10	9.0	8.0	7.1
1 : 11	9.0	8.0	7.1
1 : 12	8.0	7.1	6.3
1 : 13	8.0	7.1	6.3
1 : 14	7.1	6.3	5.6
1 : 15	7.1	6.3	5.6

1 : 18	6.3	5.6	5.0
--------	-----	-----	-----

5.3.3 缩微胶片无可可见性微斑、变色、生霉等情况。

5.3.4 无影响缩微胶片图像可读性的情况。

5.4 破损古籍修复与处理方式

5.4.1 古籍修复

古籍的修复,依据中华人民共和国国家标准《古籍修复技术规范与质量要求》(GB/T 21712-2008)执行。

普通古籍的修复,由具有相关资质的单位和人员进行,确保修复质量。

入选《国家珍贵古籍名录》古籍的修复,按照《文化部关于进一步加强古籍保护工作的通知》,原则上只能由国家级古籍修复中心组织开展。古籍数字化进行书况检查时如发现需要修复的此类古籍,应作好记录,并与负责本地区古籍修复工作的国家级古籍修复中心取得联系,履行相关手续。

古籍修复完成后再进行数字化。

5.4.2 古籍缩微胶片破损处理

缩微胶片扫描之前应进行检查,若发现胶片存在问题不适合进行数字化处理,则应查看是否保存有上一代胶片,若存在则可直接使用上一代胶片扫描或使用上一代胶片重新制作胶片后再进行扫描。若本单位未保存上一代胶片,则需检查胶片来源,若是全国图书馆文献缩微复制中心制作的胶片,则需尽快联系该中心进行修复或重新制作胶片。

5.5 数字化加工设备及配件

5.5.1 扫描设备规格

扫描设备建议选择零边距或非接触式扫描仪。设备 A3 幅面的光学分辨率 600dpi 以上,CCD 感光元件不低于 5000 像素点,色彩位数 24bit。扫描设备应是无紫外线的同步冷光源。

5.5.2 数码拍照设备规格

数码照相机的有效像素不低于 2100 万像素。建议选择有效像素 3300 万像素以上的数码设备。

5.5.3 缩微数字转换设备规格

扫描设备建议选择专业级缩微胶片扫描仪。设备光学分辨率 400dpi 以上,画幅最短边 CCD 感光元件不低于 7000 像素点,色彩位数 8bit 灰度。扫描设备具有

胶片保护装备，不划伤、损伤胶片。

5.5.4 托稿台

托稿台为古籍数码拍照的平台。托稿台材质应选用经过脱酸处理的，书籍托架应稳定牢固。托稿台可以借助专门的带有空气抽提装置，也可以自动评测被拍摄物重量，通过液压调节拍摄物与上下压平装置的空间和力度，控制拍摄物位置和平整度。

5.5.5 背景布置

拍摄或扫描古籍，应利用背景纸来营造比较干净的氛围。背景纸，为单一颜色，不带各种线条和图案。拍摄背景不宜选用暖色调的丝绒、毛毯作为背景，大片的颜色反光容易使古籍成像偏色。

中灰色纸板，是拍摄古籍的最佳背景。在曝光控制中，中灰色，即18%灰板是一种精确的测光板，表面是漫反射，不会出现光斑。通过18%灰板测到的光线与落到被拍摄物体上的光线是完全相同的。

黑色纸板，也是古籍扫描中的通常选择的背景。由于背景和古籍原件反差较大，更有利于突出文献样貌。

6 古籍元数据著录

6.1 著录内容

6.1.1 文献整理登记

文献整理是数字化工作的首要环节。为数字化加工、质检、标引等后续工作提供参考和依据。工作人员检查原书状况，清点文献册数，根据《文献整理登记表》内容要求，客观、准确记录文献基本信息。

古籍数字化文献整理登记表，包括《文献整理登记表（全书）》及《文献整理登记表（各卷）》。

(1) 《文献整理登记表（全书）》填写古籍整体检查情况，以一部书为单位，一部书一行。包含下列6个字段：

序号	名称
1	ID
2	加工記錄標識號
3	名錄號

4	題名卷數
5	相關說明
6	冊數

(2) 《文献整理登记表(各卷)》填写古籍各卷检查情况,先以册分,一册内再以卷为单位,一卷一行。包含下列 21 个字段:

序号	名称
1	ID
2	内部序號
3	加工記錄標識號
4	冊次
5	卷次
6	葉數
7	封面
8	前護
9	後護
10	封底
11	開本
12	板框
13	透字
14	夾框
15	夾字
16	皺折
17	缺葉
18	殘葉
19	簽條
20	夾紙
21	备注

6.1.2 描述元数据

(1) 《书目数据表》是古籍的书目信息，原则上每部古籍填写一行。包含下列 23 个字段：

序号	名称
1	ID
2	加工記錄標識號
3	名錄號
4	普查編號
5	書目記錄標識號
6	索書號
7	分類
8	題名卷數
9	其他題名
10	主要責任者
11	其他責任者
12	存（缺）卷
13	版本（含補配）
14	冊數
15	板框
16	版式
17	合訂狀態
18	裝幀形式
19	叢編
20	附注
21	批跋
22	館藏單位
23	製作單位

(2) 《卷目数据表》是每部古籍的卷目信息。

每部古籍第一条记录填写该部古籍是否分卷，决定是否制作具体的卷目数据。

若该部古籍为一卷或不分卷，不需要制作具体的卷目数据，则仅填第一条记录：“卷名”项中填“無卷目”，“属性”项中填“2”；

若该部古籍为分卷，需要制作具体的卷目数据，则第一条记录：“卷名”项中填“有卷目”，“属性”项中填“1”。

《卷目数据表》包含下列 10 个字段：

序号	名称
1	ID
2	内部序號
3	加工記錄標識號
4	叢書標識號
5	子目題名
6	層級號
7	冊號
8	卷名
9	葉碼
10	屬性

6.1.3 管理元数据

(1) 《外字表》是古籍的外字信息，依据《汉语文古籍机读目录格式使用手册》（国家图书馆编，北京图书馆出版社，2001年10月第1版）中393字段系统外字附注的外字描述方法进行描述。

以一部书为单位，每个外字填写一行。包含下列 6 个字段：

序号	名称
1	ID
2	内部序號

3	加工記錄標識號
4	外字
5	描述
6	位置

(2) 《管理信息表》是古籍在数字化加工过程中的管理信息。每部古籍一行。

包含下列 12 个字段：

序号	名称
1	ID
2	加工記錄標識號
3	技術信息
4	格式
5	分辨率
6	色彩
7	色彩位深
8	壓縮
9	加工方式
10	事件類型
11	操作
12	操作日期

6.1.4 元数据制作说明

珍贵古籍数字化工作的元数据制作，分别按文献整理登记、描述元数据、管理元数据等三部分工作，制作 6 张库表，包括文献整理登记表（全书）、文献整理登记表（各卷），书目数据表、卷目数据表，外字表、管理信息表。

详细制作说明，参见附录一

6.2 著录要求

制作数据应符合下列要求。

(1) 使用 Microsoft Office Access 2003 进行数据制作，以保证数据库文

件的规范性；

(2) 所有数据库表格均用通行繁体字填写。使用字符集为 Unicode 5.0 UTF-8；

(3) 字体字号为宋体 11 号；

(4) 著录信息应严格按照文献实际内容进行描述，标引词与标引对象文件应正确链接。著录文字处理错误率不超过 0.3%

7 古籍图像数字化

7.1 古籍页面划分

古籍页面（以线装为例）划分类型主要有筒子页（叶）、半叶、双半叶等。古籍数字化根据古籍的装帧、版面等实际情况，按上述划分类型选定后再数字化采集图像。

页面划分方式	筒子页（叶）	半叶	双半叶
示例			

7.2 数字图像采集方式

7.2.1 数字扫描（拍照）

每日扫描（拍照）前进行数字化设备的基本色彩校正，测量数字输入设备和数字输出设备的色彩属性。

7.2.2 缩微胶片数字转换

使用专业级缩微胶片数字转换设备。当缩微胶片整体密度较为均匀、图像质量较好、图像大小基本相同时，宜采用高速扫描仪对其进行快速整卷/盘扫描，否则宜采用低速扫描仪逐画幅扫描。

7.2.3 数字图像制作规格

资源级别	采集方式	技术规格	
档案典藏级	古籍原件	位深	24 位

(A)	扫描（拍照）	扫描分辨率	文献小于 10X12.5cm, 用 600DPI 以上 文献大于 10X12.5cm , 小于 A3, 用 400 DPI 以上
		数字拍照	2100~3300 万像素
		文件保存格式	TIFF (LZW) RAW (sRAW)
		色调再现	ICC 配置文件
	缩微胶片 数字转换	位深	8 位
		光学分辨率	400 DPI 以上
		文件格式	TIFF (LZW)
复制加工级 (P)	古籍原件 扫描（拍照）	位深	24 位
		分辨率	档案典藏级分辨率
		像素	档案典藏级拍照像素
		文件保存格式	JP2000 无损压缩
		色调再现	ICC 配置文件
	缩微胶片 数字转换	位深	8 位
		光学分辨率	档案典藏级分辨率
		文件格式	JP2000 无损压缩

【注】当前只提交档案典藏级图像数据

7.3 数字图像采集要求

7.3.1 古籍原件扫描（拍照）

数字化环境注意防护光源，避免透光或反射光的影响。

扫描后的图像清晰，数据文件叶码连续，没有重叶、缺叶，错叶、折叶等情况（原书缺叶、错叶除外）。补扫缺叶图像要与同册图像文件的大小一致，颜色接近。

(1) IT8 标准色彩导表（以下简称色卡）的扫描和拍照。色卡是色彩管理的重要手段，色卡应用可以准确还原图像颜色。每部古籍数字化前，先扫描或拍

照色卡。使用色卡时，操作人员应该戴手套以保护色卡的清洁度。

(2) 每部古籍第一册封面的拍摄。将（带尺寸）标尺放置于古籍原件左侧，距离古籍 0.1~1 厘米，一同拍摄。

(3) 按 1:1 比例扫描，叶面外围要求留白，宽度不超过 1 厘米；书页间距不超过 0.1 厘米。

(4) 以原书的上边沿为基准，以中缝为中心线，保持原文献的天头、地脚的尺寸不变，左右两边的尺寸基本不变。图像倾斜角度不大于 0.2 度。

(5) 原件表面有其他粘贴物件时，先将原件与粘贴物（即粘贴物覆盖于文献）一起扫描，然后将粘贴物掀开，再次扫描原件。

(6) 原件透背叶字迹，有虫蛀、漏洞时，需垫上古籍适用的衬纸后扫描。

(7) 分画幅扫描时，各扫描区域边缘必须有 3 厘米（含）以上的重复扫描区。

(8) 数字图像文件用 photoshop 软件检查图像不失真（图像放大至实际尺寸 100%），检查清晰度。

(9) 数字图像文件与古籍原件颜色不一致，须先进行设备色彩校正，再重新进行扫描或拍照工作。

7.3.2 缩微胶片数字化转换

(1) 胶片的选择。建议选择（工作母片）第二代负像缩微品进行数字化。

(2) 缩微胶片的保护

1) 扫描过程中，注意对缩微胶片的保管和保护，注意防火、防水、防光。

2) 扫描过程中，工作人员应戴洁净的棉质薄手套，轻拿缩微胶片的边缘。

(3) 倍率选择。按照缩微胶片拍摄时的缩小比率来选择对应的放大倍率进行还原，理论上要求还原后的图像尺寸与原件尺寸相同。

(4) 对比度选择。在扫描过程中，应根据缩微胶片图像的密度、解像力进行调整和设定。

(5) 亮度选择。根据缩微胶片图像的密度、解像力进行调整和设定。若整卷扫描，建议每卷正式扫描前进行预扫描，依照预扫描结果，调整数字化加工设备的扫描亮度参数。

7.4 数字图像处理要求

古籍数字图像处理是在未改变原扫描或拍照图像的色彩、分辨率、格式、压缩等情况下进行。数字图像文件处理内容及要求如下：

- (1) 纠偏处理。对出现偏斜的图像进行纠偏处理，对方向不正确的图像进行旋转还原，以符合阅读习惯。
- (2) 图像剪裁。拍摄原件外边缘裁至 0.5~0.75 厘米。
- (3) 不能进行锐化或者图像增强处理，不能更改图像的颜色，尽量减少对图像文件的后期处理。

7.5 数字图像检查要求

数字化单位检查图像数据的质量，建议扫描（拍照）完成后立即进行质检。数据检查的内容和要求如下：

- (1) 图像文件（各种格式）放大到 1:1 状态，逐叶检查。检查文件是否有透光、透字、有彩点、彩线、太淡、太浓、黑边、污点、歪斜、模糊（马赛克等）或图像内容不完整等现象。若不符合图像质量要求应进行图像校正或重新扫描（拍照）。
- (2) 发现文件漏扫时，应及时补扫并正确插入图像。
- (3) 检查扫描（拍照）图像的规格要求和技术参数。
- (4) 检查图像叶码是否连续，不得跳叶。
- (5) 文献以册/件为单位检查标引（登记、描述、管理）数据是否完整、准确。
- (6) 所有文件保存位置正确，可以有效打开和显示。
- (7) 按照命名规则，检查目录、文件、数据库、文档、介质等名称是否正确。
- (8) 检查各类说明、统计、验收等文档是否齐全。

8 数据命名

8.1 元数据库表

8.1.1 古籍数字化文献整理登记表

文献整理登记表（全书）

文献整理登记表（各卷）

8.1.2 描述元数据

书目数据表

卷目数据表

8.1.3 管理元数据登记表

外字表

管理信息表

8.2 加工记录标识号

加工记录标识号为8位阿拉伯数字，由单位代码（4位）+单位内部古籍数字化流水号（4位）组成。

8.3 目录命名

数据保存目录由三级结构组成：

第一级目录名称为加工记录标识号；

第二级目录名称为古籍册号。每部古籍有多册实体，每一册命名为4位数字，从0001开始，依次按流水号命名；

第三级，保存图像数据文件。命名方式见8.4

8.4 文件命名

8.4.1 色卡

色卡拍摄的两种方法，

方法一，一册古籍用单台设备完成，只拍摄一拍色卡，图像文件名为：seka+后缀，如seka.tif；

方法二，一册古籍用两台设备，即两个半叶同时拍摄，需要分别拍摄色卡。设备一拍摄色卡文件名为seka1+后缀，如seka1.tif；设备二拍摄色卡文件名为seka2+后缀，如seka2.tif。

那么seka1所对应一组古籍图像应为该册古籍的奇数叶，seka2对应的一组古籍图像应为该册古籍的偶数叶。

8.4.2 原件内容

古籍原件，以册为单位进行数字化。从古籍封面、前护、正文、后护、封底等依次加工。文件名为由4位数字组成，从0001开始，以流水号方式命名。

8.4.3 特殊处理

古籍原件的透字、残叶、签条、粘贴物等情况复杂，为客观保留件状况，古

籍数字化时，采取一叶多拍处理，如原件有粘贴物，先将原件与粘贴物（即粘贴物覆盖于文献）一起扫描，然后将粘贴物掀开，再次扫描原件。这样，就获得 2 个图像。个别情况也可能扫描多次，留下多个图像。

特殊处理的图像文件名为：4 位数字 + 小写字母

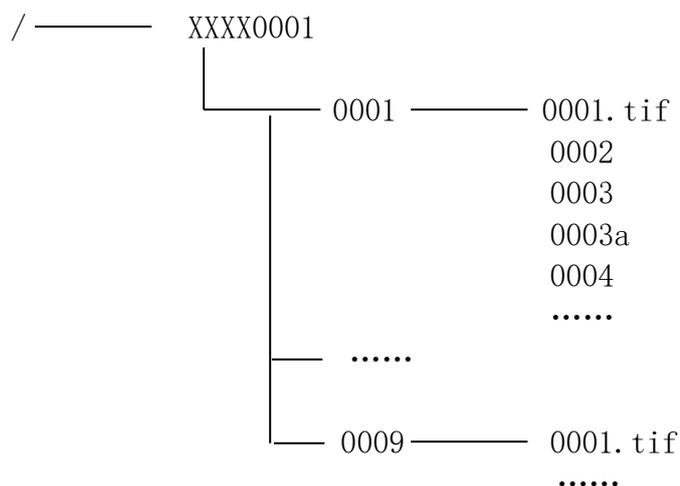
其中，4 位数字为原件内容的顺序流水号；小写字母，从 a 开始，顺序命名。如，原件第 5 叶有粘贴签条，先与粘贴物一起扫描，然后将粘贴物掀开，再次扫描。两个图像文件命名：0050.tif

0050a.tif

8.5 示例

【示例 1】以十三經註疏三百五十三卷（元刻明修本）为例：

加工记录标识号为 XXXX0001，该书第一卷的卷册号为 0001，第一叶的叶号为 0001，文件保存目录为 XXXX0001/0001，图像文件名为 0001.tif。



【示例 2】以行素草堂金石叢書十六種（清光緒三年至十七年吳縣朱記榮槐廬刻本）为例：

行素草堂金石叢書十六種

.....

金石三例續編三種（清光緒十一年刻本）

.....

誌銘廣例二卷（清梁玉繩撰）

金石三例續編三種之誌銘廣例二卷，在本套书的第 10 册和 11 册，则第一卷第一叶保存目录及文件名为 XXXX0002/0010/0001.tif，其中第一部分 XXXX0002 为加工记录标识号，第二部分 0010 为第一卷所在册的顺序号，第三部分 0001

为第一卷封面。

9 数据验收

9.1 验收方式

古籍数字资源制作完成后应由古籍数字化管理人员对成品质量进行验收。

包括数字化古籍文献整理登记表、描述元数据登记表、管理元数据登记表、图像数据以及说明文件。针对数据品质、数据结构合理性和内容完整性，采取抽样检查方式，抽检样本数依据提交数据的质量确定验收比例，由验收人员随机抽选。验收人员撰写数据验收报告。

9.2 验收标准

(1) 送检数据与《古籍数字资源提交单》的内容一致，各种格式数据和数据说明文件一一对应，不可夹杂无关文件。

(2) 各种标引、说明文件的文字、符号、版式、位置和名称准确，其综合错误率不超过 0.3%。

(3) 图像数据扫描方式、扫描规格、文件格式、文件命名、图像处理、压缩方式等符合要求，其综合错误率不超过 1%。

(4) 成品数据备份数量、保存介质命名、数据存放方式、数据内容符合规范要求，且各类型保存介质内无坏死文件、不准携带病毒，错误率为 0。

(5) 达到验收标准的数据视为合格，在规定错误率范围内检查出的数据问题由加工单位进行修正；超出错误率、未达到验收标准的数据由加工单位对全部提交数据进行整改、重新数字化等返工处理。

10 数据发布利用

10.1 数据保存

宜采用三种长期保存途径：

- (1) 硬盘存储。
- (2) 异地灾备保存。
- (3) 适时将硬盘中的数字图像转换为缩微胶片存储。

10.2 数据库建设

应兼具实用性、规范性与开放性，数据库中的数据、软硬件系统宜采用通用的格式与标准化协议，便于交换和共享，数据库及数字资源应安装至本地服务器。

数据库系统应具有输入、输出、存储、修改、更新等基础功能。建立起基于

元数据仓储模式下的查询、检索平台，逐步实现词义匹配、全文检索以及多种知识信息关联。应提供图像基础细节及缩放、翻页、目录、册次选择等基本服务功能；设置访问统计及用户反馈模块，确保古籍数字化项目应用的有效性。应包括系统用户管理、系统日志管理、数据库备份与恢复等安全管理功能。

遵循“共建共享”宗旨，提供数据接口供其他系统调用。遵循易用性原则，保障用户便捷获取系统服务，实现直接浏览器访问，鼓励支持移动端展示。注重信息安全访问和产权保护。

附录一 元数据制作说明

珍贵古籍数字化工作的元数据制作，分别按文献整理登记、描述元数据、管理元数据等三部分工作，制作 6 张库表，包括文献整理登记表（全书）、文献整理登记表（各卷），书目数据表、卷目数据表，外字表、管理信息表。

一、文献整理登记表（全书）

《文献整理登记表》属于古籍数字化工作前整理部分，是书况检查的记录文件，为数字化加工、质检、标引等后续工作提供参考和依据。填写时应客观、准确。鉴于全书与各卷存在一对多的关系，文献整理登记工作，分为两个表制作：《文献整理登记表（全书）》、《文献整理登记表（各卷）》。

《文献整理登记表（全书）》填写古籍整体检查情况，以一部书为单位，一部书一行。

《文献整理登记表（全书）》包含下列 6 个字段：

1、**ID**：表格中每行数据的顺序号，由 Access 自动编号。

2、**加工記錄標識號**：8 位阿拉伯数字，由单位代码（4 位）+单位内部古籍数字化流水号（4 位）组成。单位代码由国家古籍保护中心统一分配。单位内部古籍数字化流水号由单位自行分配，从 1 开始，不足 4 位以 0 补齐。

【注】加工记录标识号是 6 个表格的关联字段，填写时务必准确，以免出现链接错误。

3、**名錄號**：入选《国家珍贵古籍名录》的古籍的名录编号。5 位阿拉伯数字，不足 5 位以 0 补齐，如 01001。

4、**題名卷數**：依《全国古籍普查登记手册》相关规定著录。应与《书目数据表》中的一致。

5、**相關說明**：填写前面各项未能涵盖的内容。

6、**冊數**：阿拉伯数字+计量单位（如“冊”、“捲”、“軸”等）。应与《书目数据表》中的一致。

二、文献整理登记表（各卷）

《文献整理登记表（各卷）》填写古籍各卷检查情况，先以册分，一册内再以卷为单位，一卷一行。若一卷跨两册，则填写两行，分别填写该卷在这两册的相应情况。若古籍为一卷或不分卷，则直接填写册次情况，一册一行。每册封面、前护的叶数在该册第一条记录中填写，后护、封底的叶数在该册最后一条记录中

填写。

《文献整理登记表（各卷）》包含下列 21 个字段：

- 1、**ID**：表格中每行数据的顺序号，由 Access 自动编号。
- 2、**内部序號**：古籍内部顺序号。每个加工记录标识号下都从 1 开始编号。
- 3、**加工記錄標識號**：应与《文献整理登记表（全书）》中的一致。
- 4、**冊次**：阿拉伯数字，如 1、2、3。每卷所在册次都要填写，不省略。
- 5、**卷次**：卷次标识，如卷一、卷二、卷上、卷下、天干地支、元亨利贞、六艺、成语、诗韵等，参照各卷卷端著录。

若该部古籍由多种著作组成，“卷次”可依卷端所题填写，如易圖、周易上經第一、周易下經第二、周易象上傳第一、周易象下傳第二等。

正文卷一之前的内容，除封面、前护及单独成卷的卷首、目录外，统称“卷前”，如内封、序言、未单独成卷的目录等。

正文末卷之后的内容，除后护、封底及单独成卷的卷末、附录外，统称“卷後”，如原书后跋、藏书家的题跋等。

若目录单独成卷，统一填“目录”，不填“目录上”、“目录下”。

若古籍为一卷或不分卷，则只填册次，不填卷次。

6、**葉數**：阿拉伯数字。除封面、前护、后护、封底单独统计叶数外，各卷叶数在此填写。

7、**封面**：统计各册封面的叶数，填写在该册第一条记录中。

8、**前護**：统计各册前护的叶数，填写在该册第一条记录中。

9、**後護**：统计各册后护的叶数，填写在该册最后一条记录中。

10、**封底**：统计各册封底的叶数，填写在该册最后一条记录中。

11、**開本**：阿拉伯数字，精确到小数点后一位，计量单位为厘米，先纵（高），后横（宽）。如 29.9×19.2。

【注】著录内容，只写入“数字”和大写乘号“X”。勿写“厘米”

12、**板框**：阿拉伯数字，精确到小数点后一位，计量单位为厘米，先纵（高），后横（宽）。如 19.4×14.3。

【注】著录内容，只写入“数字”和大写乘号“X”。勿写“厘米”

13、**透字**：由书叶半叶可见另半叶字迹即为透字。此处填写各卷透字的书叶位置，如整卷、第 2 葉、第 5-8 葉。

14、**夾框**：板框在装订线内即为夹框。此处填写各卷夹框的书叶位置，如整卷、第 2 葉、第 5-8 葉。

15、**夾字**：图文在装订线内即为夹字。此处填写各卷夹字的书叶位置，如整卷、第 2 葉、第 5-8 葉。

16、**皺折**：书叶不平整即为皱折。此处填写各卷皱折的书叶位置，如整卷、第 2 葉、第 5-8 葉。

17、**缺葉**：整叶缺失即为缺叶。此处填写各卷缺叶的位置，如第 2 葉、第 5-8 葉。

18、**殘葉**：书叶残缺即为残叶。此处填写各卷残叶的位置，如整卷、第 2 葉、第 5-8 葉。

19、**簽條**：粘贴在书叶上的纸签即为签条。此处填写各卷签条所在书叶位置，如第 2 葉、第 5-8 葉。

20、**夾紙**：夹在书叶中的纸条即为夹纸。此处填写各卷夹纸所在书叶位置，如第 5-6 葉之间。

21、**备注**：填写前面各项未能涵盖的内容。

三、书目数据表

《书目数据表》是古籍的书目信息，按以下相关规定进行填写。

原则上每部古籍填写一行。

责任者不同的几种著作在抄写或刻印时被组合在一起，没有总题名的，做一条记录，填写一行。

不同版本的合函、合订古籍，按版本不同情况，做多条记录，填写多行，即有几种版本就做几条记录，同时分配不同的加工记录标识号，每个加工记录标识号对应的册号都从 1 开始重新计算。

丛书、丛编做一条记录，填写一行，子目不单做一条记录。

丛书零种做一条记录，填写一行。

《书目数据表》包含下列 23 个字段：

- 1、**ID**：表格中每行数据的顺序号，由 Access 自动编号。
- 2、**加工記錄標識號**：应与《文献整理登记表（全书）》中的一致。
- 3、**名錄號**：应与《文献整理登记表（全书）》中的一致。

4、**普查編號**：该部古籍在《古籍普查登记表格》中的“普查编号”。普查编号由省级行政区划代码（6位阿拉伯数字）+单位代码（4位阿拉伯数字）+流水号（7位阿拉伯数字）组成，如110000-0101-0000001。若该部古籍，尚未分配普查编号，此处则留空。

5、**書目記錄標識號**：该部古籍在该单位编目系统或OPAC中的系统ID号，比如MARC中的001字段内容。

6、**索書號**：各单位用于古籍排架、提取、借阅的索书号。

7、**分類**：依据《全国古籍普查登记手册》中的《汉文古籍分类表》进行分类，至少分到二级类，每级类之间以/间隔，如經部/叢編、史部/紀傳類/正史之屬。

8、**題名卷數**：依《全国古籍普查登记手册》相关规定著录。应与《文献整理登记表（全书）》中的一致。

9、**其他題名**：除“题名卷数”中题名外的不同题名。依《全国古籍普查登记手册》相关规定著录。

10、**主要責任者**：一种古籍具有几种不同责任方式的责任者时，列居第一种责任方式的责任者。格式为（朝代/国别）+责任者名称+责任方式。两个责任说明、两个责任者之间均以两个半角空格间隔。当某一责任方式具有多个责任者时，责任方式仅著录于最后一个责任者之后，如（宋）程頤 （宋）朱熹撰。

11、**其他責任者**：一种古籍具有几种不同责任方式的责任者时，除主要责任者外的其他责任者。格式为（朝代/国别）+责任者名称+责任方式。两个责任说明、两个责任者之间均以两个半角空格间隔。当某一责任方式具有多个责任者时，责任方式仅著录于最后一个责任者之后，如（明）徐孚遠 （明）陳子龍測議。

责任者不同的几种著作在抄写或刻印时被组合在一起，没有总题名的，第一种著作填写在“题名卷数”中，第一种著作的责任者根据责任情况分别填在“主要责任者”及“其他责任者”中，其他著作及其责任者填在“附注”中，著作之间以&间隔。

例 1：

題名卷數	周易傳義十卷上下篇義一卷
主要責任者	（宋）程頤 （宋）朱熹撰
附注	&易圖集錄一卷易五贊一卷筮儀一卷 （宋）朱熹撰

例 2:

題名卷數	春秋經傳集解三十卷
主要責任者	(晉) 杜預撰
其他責任者	(唐) 陸德明釋文
附注	&春秋名號歸一圖二卷 (蜀) 馮繼先撰&年表一卷

12、**存(缺)卷**: 存(缺)卷数+括注具体存(缺)卷次, 卷数、卷次用中文表示(也可用阿拉伯数字表示, 后期可利用程序进行转换), 不连续的卷次之间用顿号间隔。若古籍由多个著作组成, 在“存(缺)卷”中, 多个著作之间以逗号间隔, 其中某一著作完整, 则填“全”, 不填具体卷数。例 1 (正确表达): 存四卷(一至三、六)

例 2 (用阿拉伯数字亦可): 存 4 卷(1-3、6)

例 3: 存六十八卷(一、四至十二、十九至二十一、二十三至二十八、三十、三十九至六十七、七十三至九十、一百三十)

例 4: 獨漉堂詩集十五卷文集十五卷

存二十九卷(詩集全, 文集一至八、十至十五)

13、**版本(含補配)**: 依《全国古籍普查登记手册》相关规定著录。如宋乾道七年(1171)蔡夢弼東塾刻本(卷四十三配清光緒元年楊保彝影宋抄本)。

14、**冊數**: 阿拉伯数字+计量单位(如“冊”、“捲”、“軸”等)。应与《文献整理登记表(全书)》中的一致。

15、**板框**: 阿拉伯数字, 精确到小数点后一位, 计量单位为厘米, 先纵(高), 后横(宽)。如 19.4×14.3。

【注】著录内容, 只写入“数字”和大写乘号“X”。勿写“厘米”

16、**版式**: 依《全国古籍普查登记手册》相关规定著录。行数、字数、双行小字字数用中文表示(也可用阿拉伯数字表示, 后期可利用程序进行转换), 如九行二十字小字雙行同白口左右雙邊、十行十八字小字雙行二十三字黑口左右雙邊、十行十八字白口左右雙邊。

17、**合訂狀態**: 不同版本的合函、合订古籍, 按版本不同情况, 做多条记录, 填写多行。在这些记录的“合訂狀態”中, 分别填写“與某某書合訂”, 并括注对方的加工记录标识号。例: 資治通鑒綱目五十九卷首一卷 明弘治十一年(1498)書林慎獨齋刻本

續資治通鑒綱目二十七卷 明弘治十七年（1504）書林慎獨齋刻本

这两种合订，按其版本情况，应做两条记录：

加工記錄標識號	16010001
題名卷數	資治通鑒綱目五十九卷首一卷
版（帶補配）本	明弘治十一年（1498）書林慎獨齋刻本
合訂狀態	與“續資治通鑒綱目二十七卷”（16010002）合訂
加工記錄標識號	16010002
題名卷數	續資治通鑒綱目二十七卷
版（帶補配）本	明弘治十七年（1504）書林慎獨齋刻本
合訂狀態	與“資治通鑒綱目五十九卷首一卷”（16010001） 合訂

18、**裝幀形式**：填写该部古籍的装帧形式，如綫裝、卷軸裝、經折裝、蝴蝶裝等。

19、**叢編**：若古籍为丛书零种，则需填写其所属丛编题名。

20、**附注**：填写前面各项未能涵盖的内容。另外，责任者不同的几种著作在抄写或刻印时被组合在一起，没有总题名的，除第一种著作外的其他著作及其责任者填在“附注”中，著作之间以&间隔。

21、**批跋**：在古籍流传过程中，藏书家对该部古籍所作的批校题跋等信息。一个批跋信息由朝代/国别+批跋者名称+批跋方式组成。两个批跋信息之间、两个批跋者之间均以两个半角空格间隔。当某一批跋方式具有多个批跋者时，这些批跋者之间以朝代为序，同朝代者以生年为序，批跋方式仅著录于最后一个批跋者之后，如明朱承爵 清沈窳 清王芑孫 清潘遵祁 清盛昱題款。

22、**館藏單位**：古籍收藏单位。

23、**製作單位**：数字化加工单位。

四、卷目数据表

《卷目数据表》是每部古籍的卷目信息。

每部古籍第一条记录填写该部古籍是否分卷，决定是否制作具体的卷目数据。

若该部古籍分卷，需要制作具体的卷目数据，则在第一条记录中，“叢書標

識號”中填丛书标记（丛书为1，非丛书为2），“卷名”中填“有卷目”，“屬性”中填“1”，以下各行填写具体的卷目数据，“屬性”均填“0”。

若该部古籍为一卷或不分卷，不需要制作具体的卷目数据，则仅填第一条记录，“丛书标识号”中填丛书标记（丛书为1，非丛书为2），“卷名”中填“無卷目”，“属性”中填“2”。

《卷目数据表》包含下列10个字段：

- 1、ID：表格中每行数据的顺序号，由 Access 自动编号。
- 2、内部序號：古籍内部顺序号。每个加工记录标识号下都从1开始编号。
- 3、加工記錄標識號：应与《文献整理登记表（全书）》中的一致。

4、叢書標識號：由多位数字和“_”组成，基本格式为A_B_C。A为丛书标记，丛书、丛编填“1”，非丛书、丛编填“2”。每部古籍第一条记录中“丛书标识号”只填写丛书标记。B表示子目流水号，用非定长数字表示。当丛书、丛编的子目仍为丛书、丛编时，使用“_C”，可重复，C表示子目流水号，用非定长数字表示。

丛书标识号可以重复，不能为空。

例：行素草堂金石叢書十六種（清光緒三年至十七年吳縣朱記榮槐廬刻本）

其第10个子目为：金石三例續編三種（清光緒十一年刻本），丛书标识号为1_10；

其第10个子目的第3个子目为：誌銘廣例二卷（清梁玉繩撰），丛书标识号为1_10_3

5、子目題名：此处填写丛书的各级子目题名及其卷数。另外，一部古籍由多种著作组成，各种著作的题名及其卷数也填在此处。

6、層級號：表示各级卷目之间的层级及顺序关系。每一级的层级号为3位阿拉伯数字，不足3位以0补齐。上下级之间以-连接。

若为第一级卷目，则为001、002等。

若为第二级卷目，则为001-001、001-002、002-001、002-002等。

以下各级类目，依次类推。

层级号按实存卷的先后顺序分配，不按实存卷的卷次分配。如该部古籍存三卷（三、五至六），则卷三、卷五、卷六的层级号依次分配为001、002、003。

同一级的层级号之间不能跳号。如该部古籍分配的层级号为：001、002、

003-001、003-002、004，其中 003-001 之前缺少 003，应予增补。

7、**冊號**：该子目题名或卷名在整部书的册位置，为顺序号，由 4 位阿拉伯数字组成，不足 4 位以 0 补齐。应与图像文件保存的文件夹名称一致。

“冊號”，除每一部古籍的第一条数据（即“有卷目”、“無卷目”）外，该字段不能为空。

8、**卷名**：由各卷卷端之卷次标识及名称组成，卷次标识与名称之间以两个半角空格间隔，视各卷卷端的卷次标识前有无“卷”字决定是否带“卷”字。如“一 五帝本紀第一”、“二 夏本紀第二”（卷端为“史記一”、“史記二”，未带“卷”字）。

若卷次标识、卷名称仅具其一，则按实际情况著录。

若古籍分卷且有卷前或卷后等内容，则卷前、卷后也需填写。

若“子目题名”中填有内容，相应卷名则为空。

每部古籍第一条记录的“卷名”，根据古籍是否分卷，是否需要制作具体的卷目数据，填写“有卷目”或“無卷目”。

9、**葉碼**：填写子目题名或卷的首页在该册中的图像文件顺序号，应与图像文件名称中的数字对应。

例：卷一第一页的图像文件名为 0008.tif，此处即填 8。

子目题名，後漢書志三十卷，在第 33 册的第 3 个图像，此处即填 3。

10、**屬性**：若该部古籍分卷，需要制作具体的卷目数据，则在第一条记录中，“属性”填“1”，以下各行填写具体的卷目数据，“属性”均填“0”。若该部古籍为一卷或不分卷，不需要制作具体的卷目数据，则仅填第一条记录，“属性”填“2”。

五、外字表

《外字表》是古籍的外字信息，每个外字填写一行。

《外字表》包含下列 6 个字段：

- 1、**ID**：表格中每行数据的顺序号，由 Access 自动编号。
- 2、**內部序號**：古籍内部顺序号。每个加工记录标识号下都从 1 开始编号。
- 3、**加工記錄標識號**：应与《文献整理登记表（全书）》中的一致。
- 4、**外字**：描述外字的结构及组成部分。使用《汉语文古籍机读目录格式使

用手册》（国家图书馆编，北京图书馆出版社，2001年10月第1版）中393字
段系统外字附注的外字描述方法进行描述，如“淘-彳”、“澄（彳→扌）”等。

5、**描述**：补充说明该字字形及拼音等信息，如表示“淘”字去掉左边三点水、表示“澄”字左边三点水更换为提手旁。

6、**位置**：该外字所在整部书的册号和图像文件位置（册号为4位阿拉伯数字，不足4位以0补齐）、图像文件（与图像文件名一致）。册号、叶码之间以“\”间隔，如“0001\0018”（表示在第1册第18叶，对应图像文件路径为0001\0018.tif）。

六、管理信息表

《管理信息表》是古籍在数字化加工过程中的管理信息。每部古籍一行。

《管理信息表》包含下列12个字段：

- 1、**ID**：表格中每行数据的顺序号，由Access自动编号。
- 2、**加工記錄標識號**：应与《文献整理登记表（全书）》中的一致。
- 3、**技術信息**：与信息资源的创建、加工、使用相关的物理参数、技术手段与标准以及硬件环境，如赛数11000。
- 4、**格式**：信息资源的内容形式，包括资源内容与其元数据的类型，如TIFF、RAW。
- 5、**分辨率**：数字图像的扫描分辨率或拍照像素，如400dpi或3300万像素。
- 6、**色彩**：彩色、灰度、黑白等。
- 7、**色彩位深**：图像采样的色彩位深，如8位、24位。
- 8、**壓縮**：表示压缩方法和压缩率，可进一步细分为“压缩方法”和“压缩率”。如不壓縮、無損壓縮。本手册规定无损压缩方式为LZW。
- 9、**加工方式**：数字信息资源的加工方式。如原書掃描、原書拍照、縮微轉換、格式轉換。
- 10、**事件類型**：獨立加工或委託加工。
- 11、**操作**：本手册规定数字化后图像的处理方式，只有剪裁、纠偏。上述操作应在未改变原扫描或拍照图像的色彩、分辨率、格式、压缩等情况下进行。
- 12、**操作日期**：图像采集、图像操作处理的时间。书写形式为XXXX-XX-XX

附录二 古籍数字资源提交单

交接时间			
提交单位 代码及单 单位名称		交接人员	
接收单位		接收人员	
交接内容			
<p>1. 硬盘编号及存储量:</p> <p>2. 文献数量</p> <p>3. 数据类型与数量</p> <p>4. 说明文档</p>			

备注: