

数字化验收工具 V1.0

用户使用说明书

数字化验收工具开发小组

2025 年 11 月

目录

1. 系统介绍.....	1
1.1 系统功能简介	1
1.2 运行环境	1
2. 安装与卸载.....	1
2.1 安装操作	1
2.2 卸载操作	2
3. 系统操作指南.....	2
3.1 验收前准备	2
3.1.1 测试数据内容	2
3.1.2 测试数据层级	2
3.2 登录操作	4
3.3 主界面介绍	5
3.4 模块操作	6
3.4.1 查重复文件模块	6
3.4.2 查遗漏、存储路径和命名规范模块	7
3.4.3 查质量模块	14
3.4.4 元数据模块	14
3.4.5 查数据挂载模块	15
3.4.6 查工作文件模块	15
3.4.7 存储载体模块	15
4. 常见问题.....	21
4.1 忘记检查结果	21
4.2 查漏、存储路径和命名规范模块报错	21
4.2.1 “生成逻辑地址文件”时报错一	21
4.3.2 “生成逻辑地址文件”时报错二	23
4.3.3 “生成逻辑地址文件”时报错三	23
4.3.4 “生成物理地址文件”时报错	24
5. 附录.....	25

1.系统介绍

1.1 系统功能简介

基本介绍：本软件旨在构建淮阴区数字化档案检查验收系统，通过自动化技术实现档案数字化成果的部分流程质量管控，涵盖文件查重、查漏、存储路径验证、命名规范检查等核心功能，确保数字化档案符合国家标准及行业规范，提升验收效率与准确性。

1.2 运行环境

本平台是在用 IntelliJ IDEA 2023 作为开发工具开发的。
提供了以下运行环境：Windows 11 家庭版 64 位。

2.安装与卸载

2.1 安装操作

本软件为绿色软件，安装流程如下：

1. 解压获取的压缩包（文件名为“数字化验收工具 DocCheckTool V1.0.zip”）；
2. 点击解压后文件夹内的.exe 程序，即可直接启动使用。

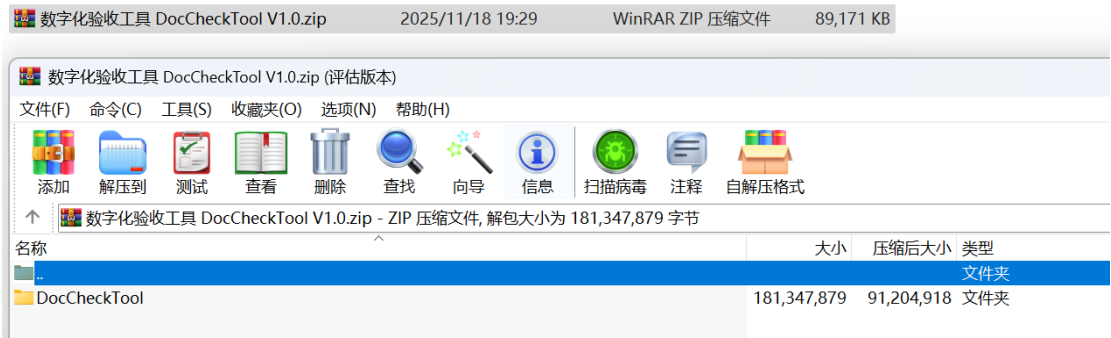


图 2-1 获取的压缩包

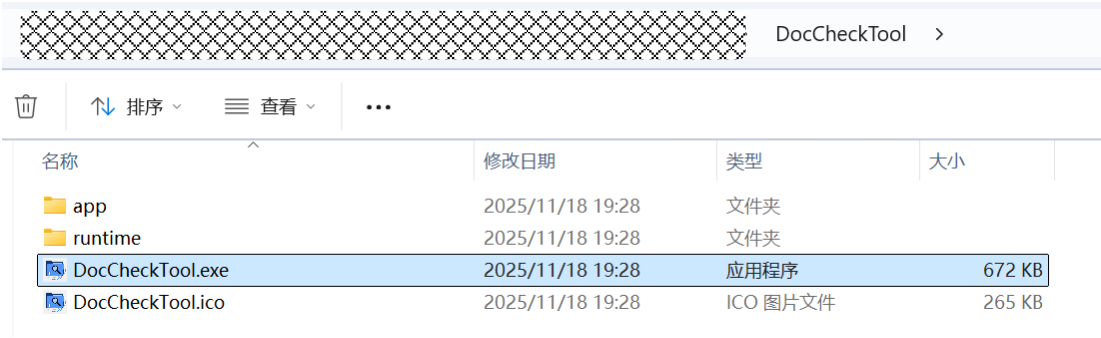


图 2-2 解压后的软件结构

2.2 卸载操作

无需复杂卸载流程，直接删除原始压缩包和解压后的文件夹，即可完成完全卸载。

3.系统操作指南

3.1 验收前准备

3.1.1 测试数据内容

待检查的数字化档案（以“测试数据”为例）需包含三个文件夹，具体要求如下：

- JPG 文件夹：存放页面级数字化成果，支持 jpg 、 jpeg 、 png 、 bmp 、 gif 、 tiff 或 jp2 数据格式；
- PDF 文件夹：存放文件级数字化成果，支持 pdf 或 ofd 格式；
- 目录文件夹：存放著录信息文件，支持 xls、 xlsx 、 xml 或 dbf 格式。

3.1.2 测试数据层级

以档号“ 9027-WS2015-D30-0001 ”的档案为例，各级文件存储路径规范如下：

(1) JPG 页面级文件

该档号的第一页页面级数字化成果（也就是名为：
9027-WS2015-D30-0001-0001.jpg 的文件）的存储路径为：

...\测试数据
 \JPG
 \9027
 \9027-WS2015
 \9027-WS2015-D30
 \9027-WS2015-D30-0001
 \9027-WS2015-D30-0001-0001.jpg

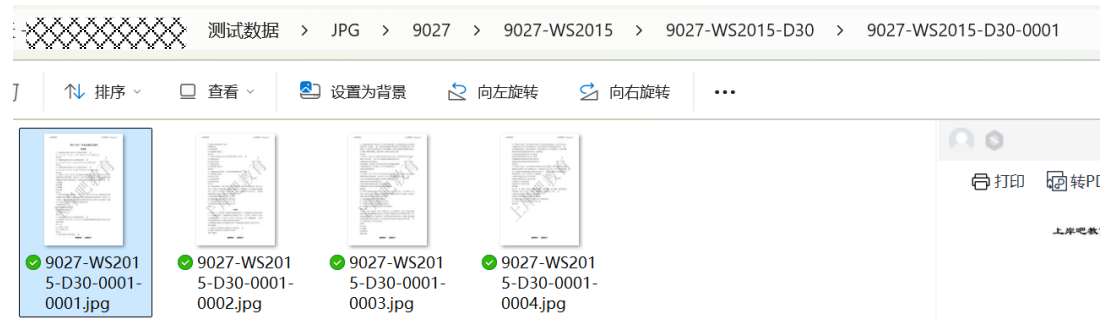


图 3-1 JPG 页面级文件存储路径示例

(2) PDF 文件级文件

该档号的文件级数字化成果（也就是名为：9027-WS2015-D30-0001.pdf
的文件）的存储路径为：

...\测试数据
 \JPG
 \9027
 \9027-WS2015
 \9027-WS2015-D30
 \9027-WS2015-D30-0001
 \9027-WS2015-D30-0001.pdf



图 3-2 PDF 文件级文件存储路径示例

(3) 目录文件

以“总和.xlsx”为例，存储路径：

...\测试数据
 \目录
 \总和.xlsx

注意：表格内容需符合“生成逻辑地址文件”的字段要求，避免报错。



图 3-3 目录文件存储路径示例

3.2 登录操作

- 1.启动解压后的.exe 应用程序，进入登录界面；
- 2.输入指定的用户名和密码；
- 3.单击“登录”按钮，即可进入系统主界面。



图 3-4 登录界面

3.3 主界面介绍

系统主界面从上到下分为三个核心区域，功能如下：

- 菜单栏：包含“文件”“帮助”等核心操作入口；
- 项目信息栏：需输入项目名称、验收时间、全宗号、档案门类、归档年度等信息，便于后续文件命名和数据追溯；
- 模块操作栏：包含 3 个核心检查模块 4 个指示操作模块，支持下拉选择或滚轮切换。



图 3-5 主界面

3.4 模块操作

请按照 1-7 的顺序完成所有检查模块操作，确保验收流程完整。

3.4.1 查重复文件模块

功能说明

批量检测页面级或文件级数字化成果中的重复文件，通过计算文件哈希值实现精准比对。



图 3-6 查重复文件模块操作界面

操作步骤

1. 载入数据：点击“选择文件夹”按钮，或直接在文本区域输入文件地址，分批载入页面级（JPG）和文件级（PDF）数据；
2. 开始检查：点击“开始检查”按钮，系统自动计算文件哈希值并比对，“结果反馈”区域将显示扫描文件总数、目标文件数、非目标文件数、重复文件组数及重复文件数量；
3. 核实确认：根据系统反馈的重复文件组，逐一核实是否为真实重复文件；
4. 记录结果：将确认后的检查结果填入《查重登记表》。

3.4.2 查遗漏、存储路径和命名规范模块

功能说明

分页面级和文件级两类，通过比对逻辑地址与物理地址，检测文件遗漏、存储路径错误及命名不规范问题。

操作要求

需先检查页面级数字化成果（JPG），再检查文件级数字化成果（PDF），两类检测流程一致，需重复操作两次。



图 3-7 查遗漏、存储路径和命名规范模块操作界面

页面级检测操作步骤

1. 选择模式：点击下拉列表，选择“页面级”模式（第一次操作默认是页面级）；



图 3-8 选择“页面级”模式

2. 载入目录文件：点击“选择文件”，先选择目录文件格式（xls、xlsx、xml 或 dbf），再选中目标目录文件；

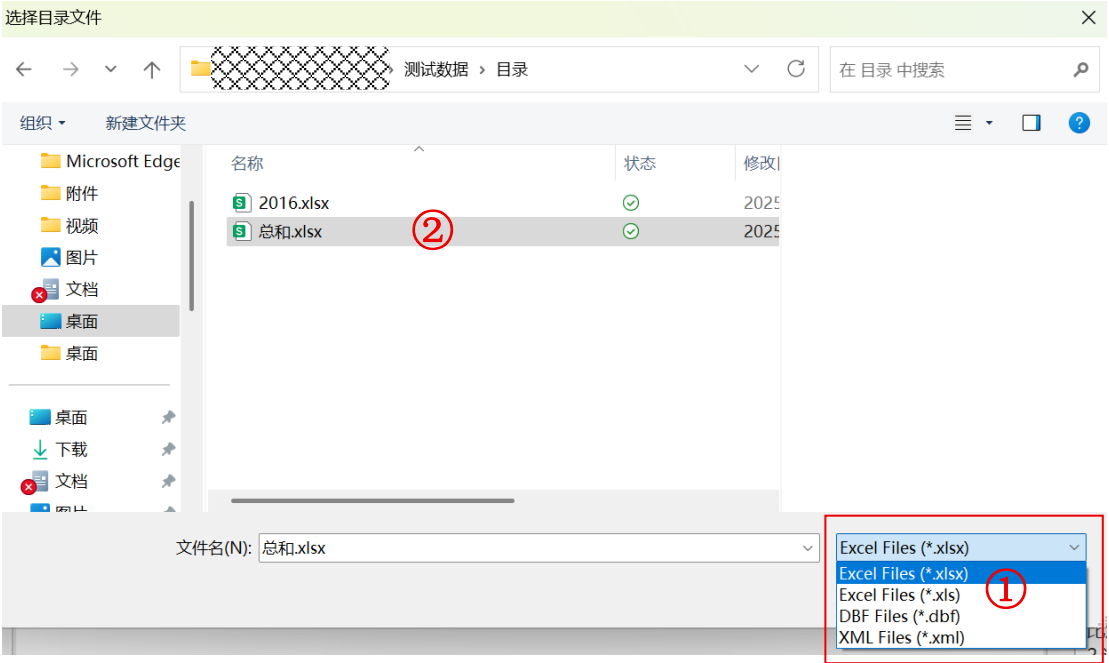


图 3-9 载入目标目录

3. 生成逻辑地址文件：点击“生成逻辑地址文件”，选择存储位置（建议放在目录文件夹内），修改文件名为“xxxx（项目名称）-页面级-逻辑地址文件.csv”；

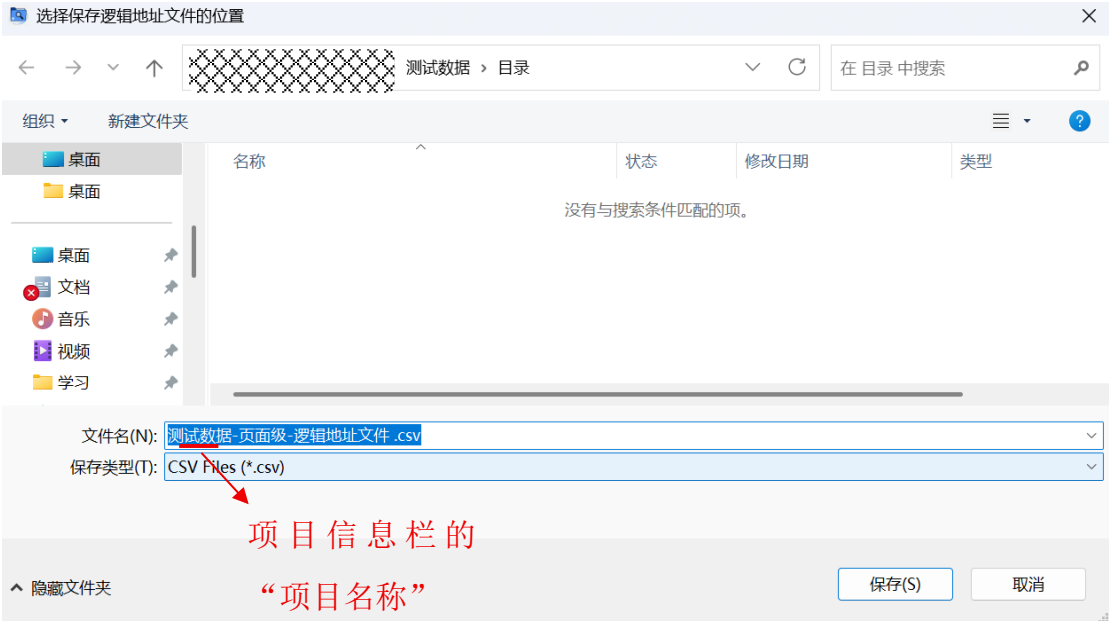
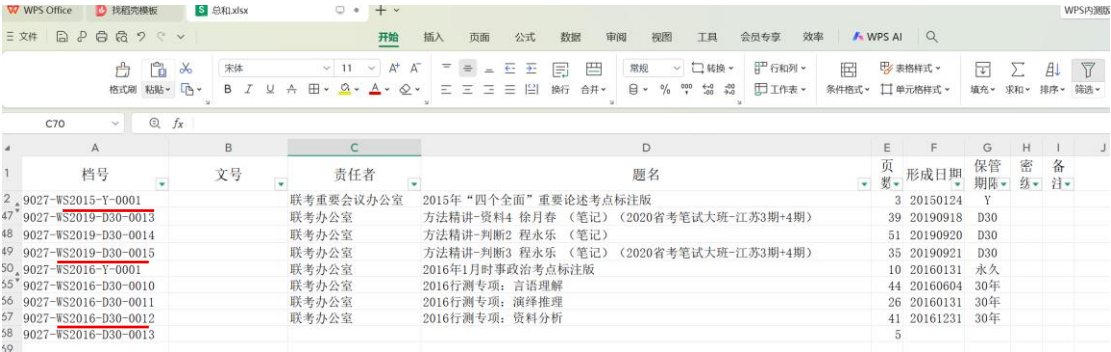


图 3-10 生成页面级逻辑地址文件

4. 载入页面级文件夹：选择目录对应的 JPG 文件夹（需包含目标档号的页面级文件）；（如这个“总和”目录里的内容是 WS2015，WS2016 和 WS2019

的著录信息，那么要导入的文件夹就应该只包括这三部分的 JPG 文件夹）。



档号	文号	责任者	题名	页数	形成日期	保管期限	密码	备注
9027-WS2015-Y-0001		联考重要会议办公室	2015年“四个全面”重要论述考点标注版	3	20150124	Y		
9027-WS2019-D30-0013		联考办公室	方法精讲-资料4 徐月春 《笔记》(2020省考试大班-江苏3期+4期)	39	20190918	D30		
9027-WS2019-D30-0014		联考办公室	方法精讲-判断2 程永乐 《笔记》	51	20190920	D30		
9027-WS2019-D30-0015		联考办公室	方法精讲-判断3 程永乐 《笔记》(2020省考试大班-江苏3期+4期)	35	20190921	D30		
9027-WS2016-Y-0001		联考办公室	2016年1月时事政治考点标注版	10	20160131	永久		
9027-WS2016-D30-0010		联考办公室	2016行测专项: 言语理解	44	20160604	30年		
9027-WS2016-D30-0011		联考办公室	2016行测专项: 演绎推理	26	20160131	30年		
9027-WS2016-D30-0012		联考办公室	2016行测专项: 资料分析	41	20161231	30年		
9027-WS2016-D30-0013				5				

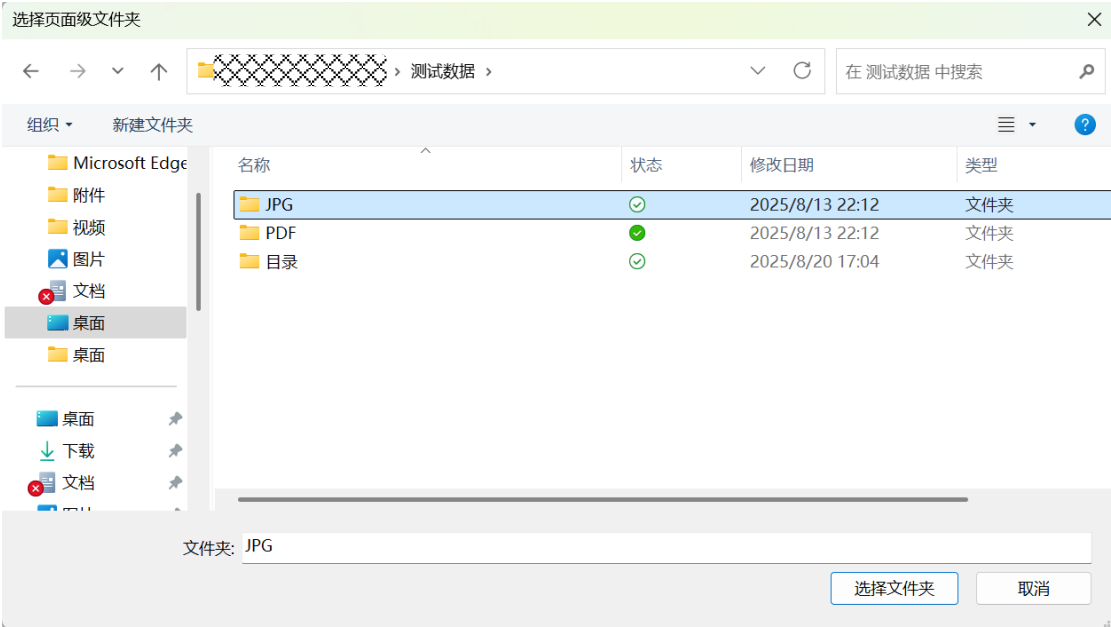
图 3-11 “总和.xlsx” 的内容



名称	状态	修改日期	类型	大小
9027-WS2015	成功	2025/8/13 22:12	文件夹	
9027-WS2016	成功	2025/8/13 22:12	文件夹	
9027-WS2019	成功	2025/8/13 22:12	文件夹	

图 3-12 包含“总和.xlsx"所有内容的页面级文件的文件夹

（包含 2015，2016 和 2019 的文件夹是 9027，由于 JPG 文件夹下面只有 9027 一个文件夹，所以导入 JPG 也是代表只包含这三个数字化成果的文件夹）



名称	状态	修改日期	类型
JPG	成功	2025/8/13 22:12	文件夹
PDF	成功	2025/8/13 22:12	文件夹
目录	成功	2025/8/20 17:04	文件夹

图 3-13 载入包含 “总和.xlsx"内容的页面级文件的文件夹

5. 生成物理地址文件：点击“生成物理地址文件”，选择存储位置（建议放在目录文件夹内），修改文件名为“xxxx（项目名称）-页面级-物理地址文件.csv”；

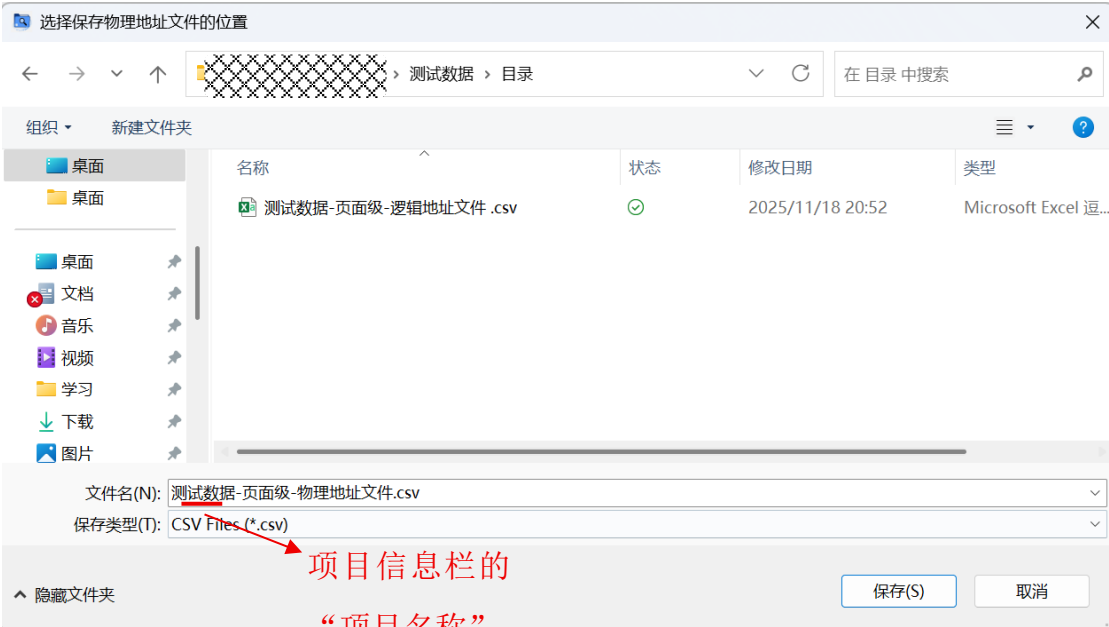


图 3-14 保存页面级物理地址文件

6. 开始比对：点击“开始比对”按钮，系统自动比对逻辑地址与物理地址，反馈差异结果；



图 3-15 “页面级”模式比对结果展示

7. 核实记录：根据反馈结果核实问题原因，将结果填入《查遗漏、查存储路径和命名规范登记表》。

文件级检测操作步骤

1. 选择模式：点击下拉列表，切换至“文件级”模式；



图 3-16 选择“文件级”模式

2. 载入目录文件：重复页面级检测的第 2 步（可选择与页面级一致的目录文件）；

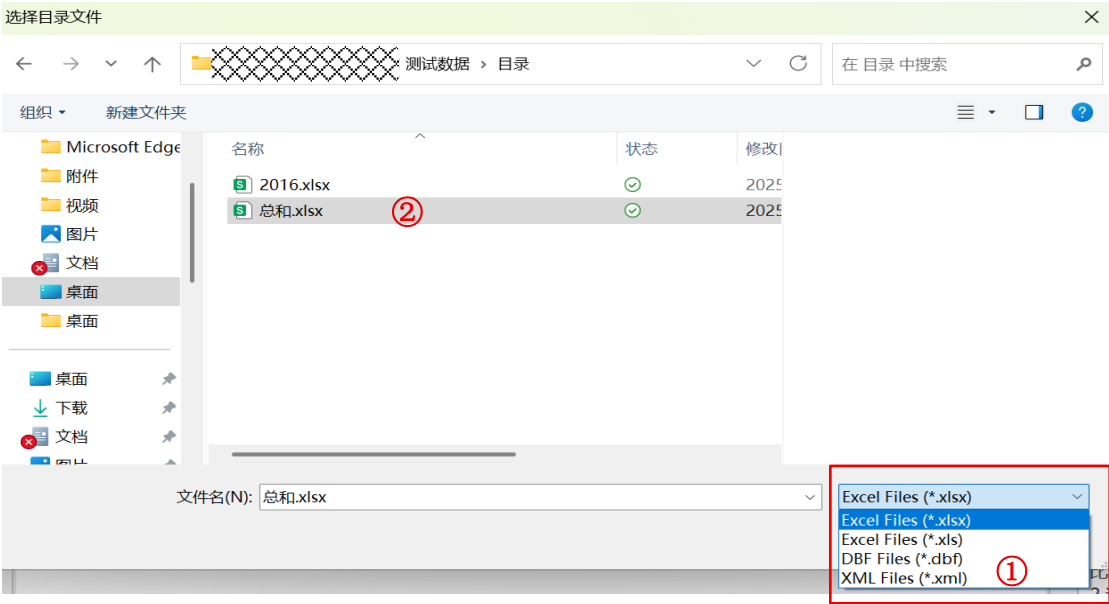


图 3-17 载入目标目录

3. 生成逻辑地址文件：点击“生成逻辑地址文件”，修改文件名为“xxxx（项目名称）-文件级-逻辑地址文件.csv”，存储位置建议与页面级一致；

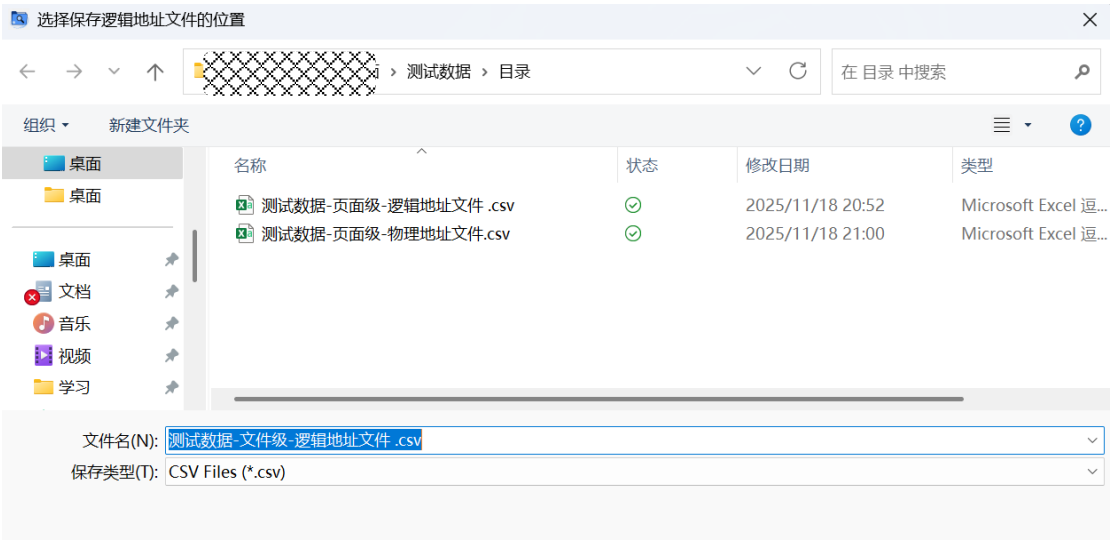


图 3-18 生成文件级逻辑地址文件

4. 载入文件级文件夹：选择目录对应的 PDF 文件夹（需包含目标档号的文件级文件）；（如这个“总和”目录包含的是 WS2015，WS2016 和 WS2017 的著录信息,那么要导入的文件夹就应该只包括这三部分的 PDF 文件夹）。



图 3-19 包含“总和.xlsx”所有内容的文件级文件的文件夹

（包含 2015，2016 和 2019 的文件夹是 9027，由于 PDF 文件夹下面只有 9027 一个文件夹，所以导入 PDF 文件夹也是代表只包含这三个数字化成果的文件夹）



图 3-20 载入包含“总和.xlsx”所有内容的文件级文件的文件夹

5. 生成物理地址文件：点击“生成物理地址文件”，修改文件名为“XXXX（项目名称）-文件级-物理地址文件.csv”，存储位置与页面级一致；

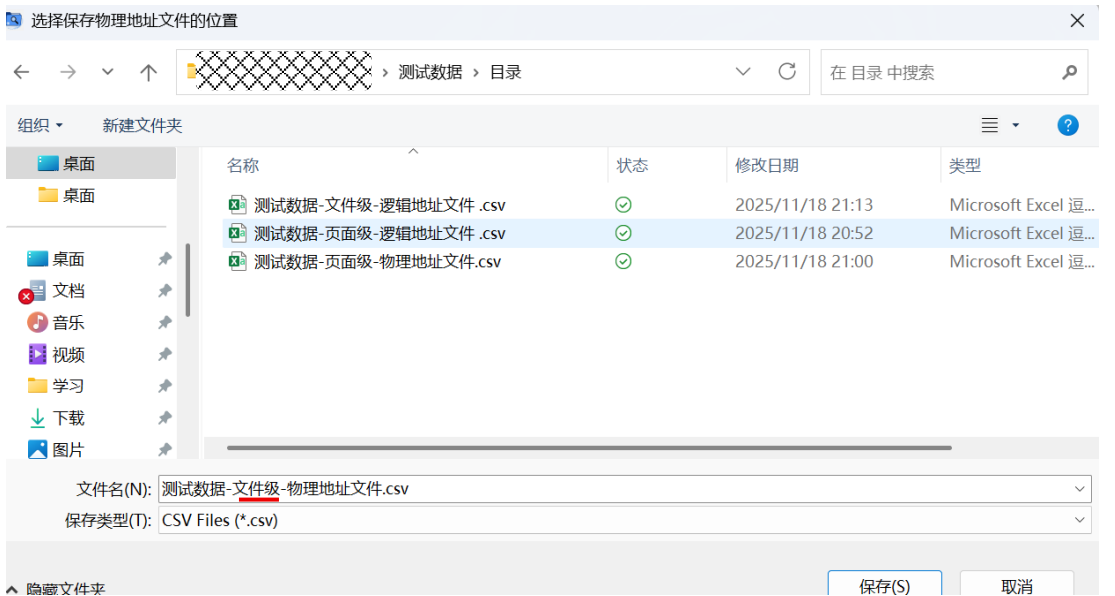


图 3-21 生成文件级物理地址文件

6. 开始比对：点击“开始比对”按钮，系统反馈比对结果；

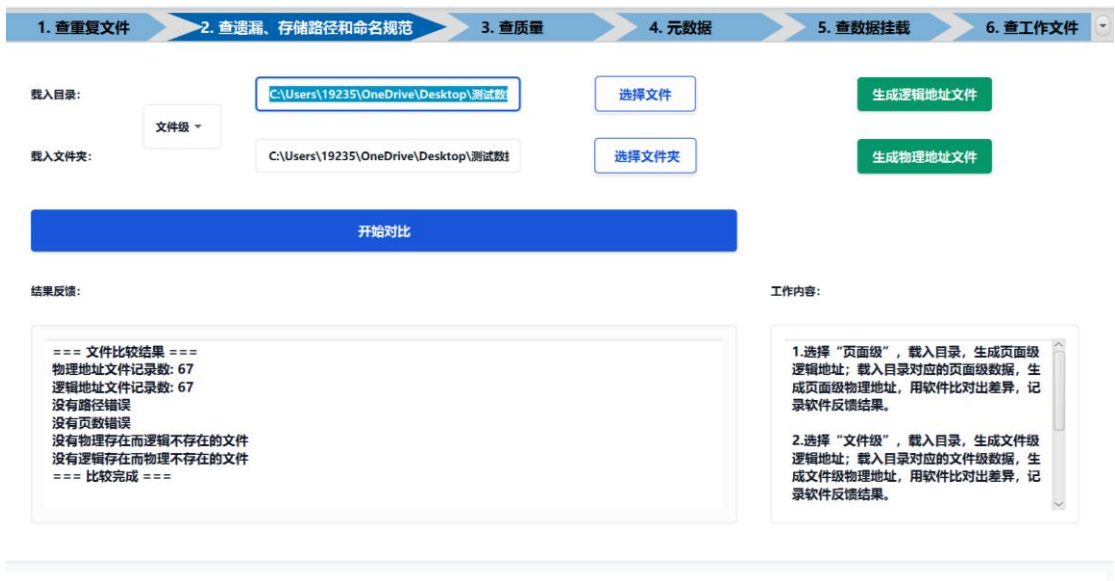


图 3-22 “文件级”比对结果展示

7. 核实记录：重复页面级检测的第 7 步，完成文件级问题记录。

结果反馈说明

- 出现：“物理文件在逻辑文件中未找到”，说明可能是文件存放地址错误、命名不规范或目录中没有该档号记录或目录中该档号页数少于实际页数；

- 出现：“逻辑文件在物理文件中未找到”，说明可能是漏扫，命名不规范、目录中档号错误或目录中的文件页数超过实际页数。

3.4.3 查质量模块

功能说明

检测数字化成果的质量是否符合规范，系统仅提供操作提示，需人工完成检查。

操作要求

按照系统界面“工作内容”提示，对照相关质量标准，人工核查数字化成果的清晰度、完整性等质量指标，将结果填入《质量检查登记表》。

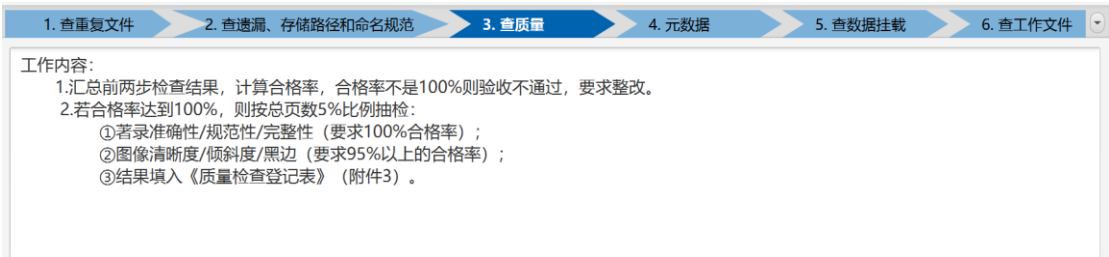


图 3-23 查质量模块操作界面

3.4.4 元数据模块

功能说明

核查数字化项目信息、技术环境及技术参数的完整性与规范性。

操作要求

对照《元数据检查登记表》，人工检查并登记相关信息，确保元数据无遗漏、无错误。

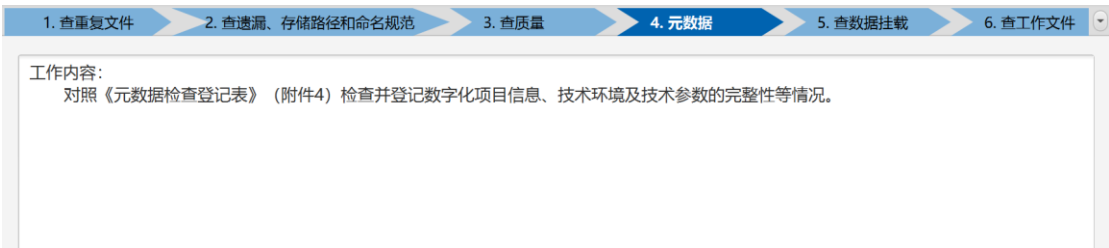


图 3-24 元数据模块操作界面

3.4.5 查数据挂载模块

功能说明

验证数字化成果与相关系统的挂载兼容性，系统提供操作提示，需人工完成检查。

操作要求

按照系统界面“工作内容”提示，人工核查数据挂载的有效性和准确性，将验证结果填入《挂接检查登记表》。

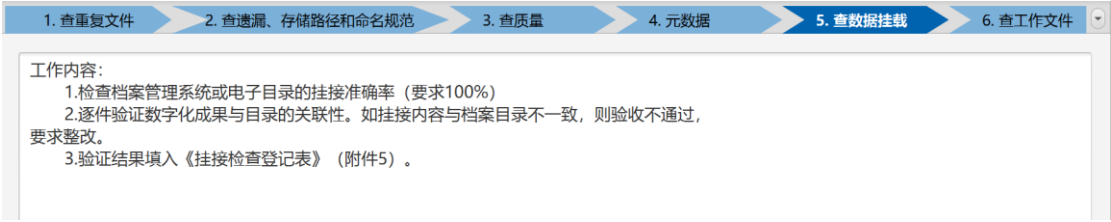


图 3-25 查数据挂载模块操作界面

3.4.6 查工作文件模块

功能说明

检查数字化工作台帐的规范性及与成果的一致性。

操作要求

对照《工作记录检查登记表》，人工核查工作台帐内容，确保与数字化成果一一对应，并完成登记。

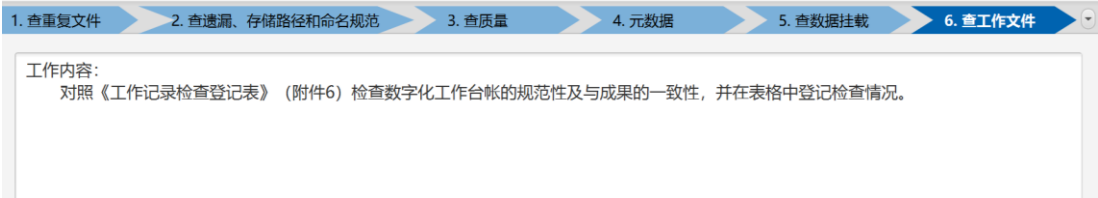


图 3-26 查工作文件模块操作界面

3.4.7 存储载体模块

功能说明

核查存储载体的合规性，生成数字化成果哈希值表并验证压缩包完整性。



图 3-27 存储载体模块操作界面

操作步骤

- 1. 存储载体核查：对照《存储载体检查登记表》，记录存储载体的类型、数量、内容及可读性情况；
- 2. 载入数字化成果：分批点击“选择文件”按钮，导入 JPG、PDF 等数字化成果文件夹（不含目录文件夹），若需更换批次，点击“清除”按钮；

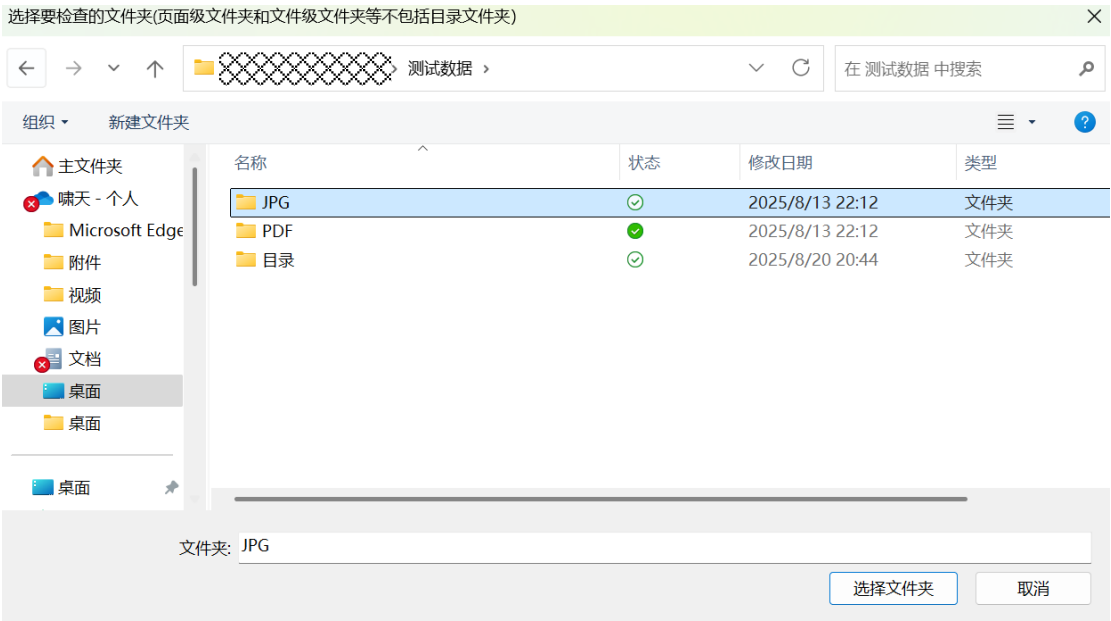


图 3-28 载入所有已检查的页面级数字化成果

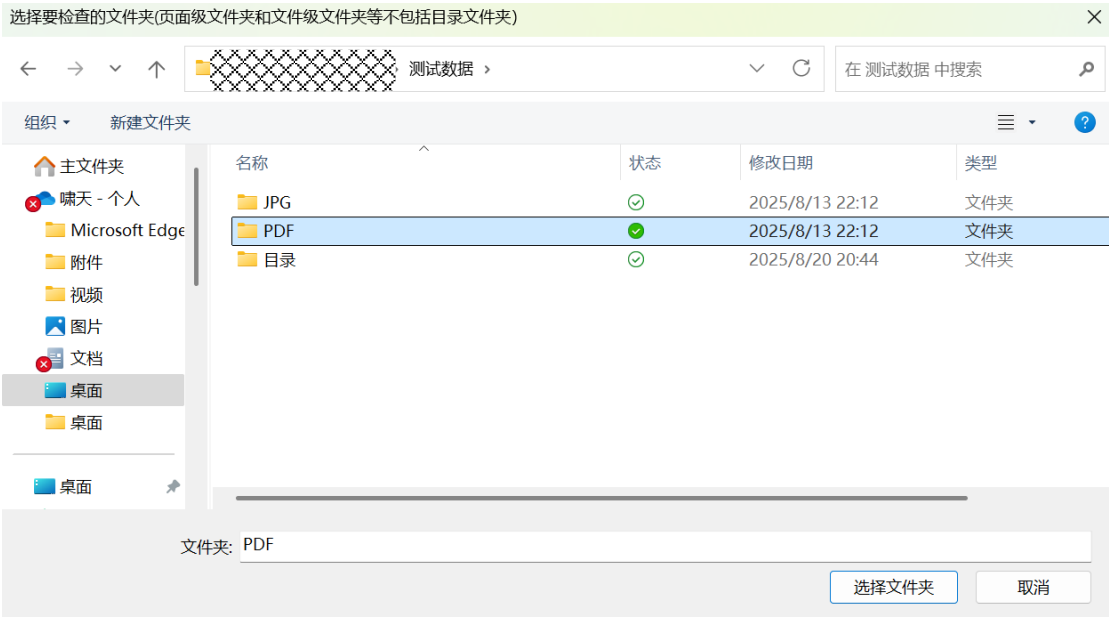


图 3-29 载入所有已检查的文件级数字化成果

注意：若是单个文件夹导入出错，建议选中路径（包含分号）后“delete”删除。



图 3-30 关闭不需要的文件夹

若是需要检查下一批数字化成果请点击“清除”按钮。

3. 生成哈希值列表文件：点击“生成哈希值列表文件”，选择存储位置（建议放在目录文件夹内）；

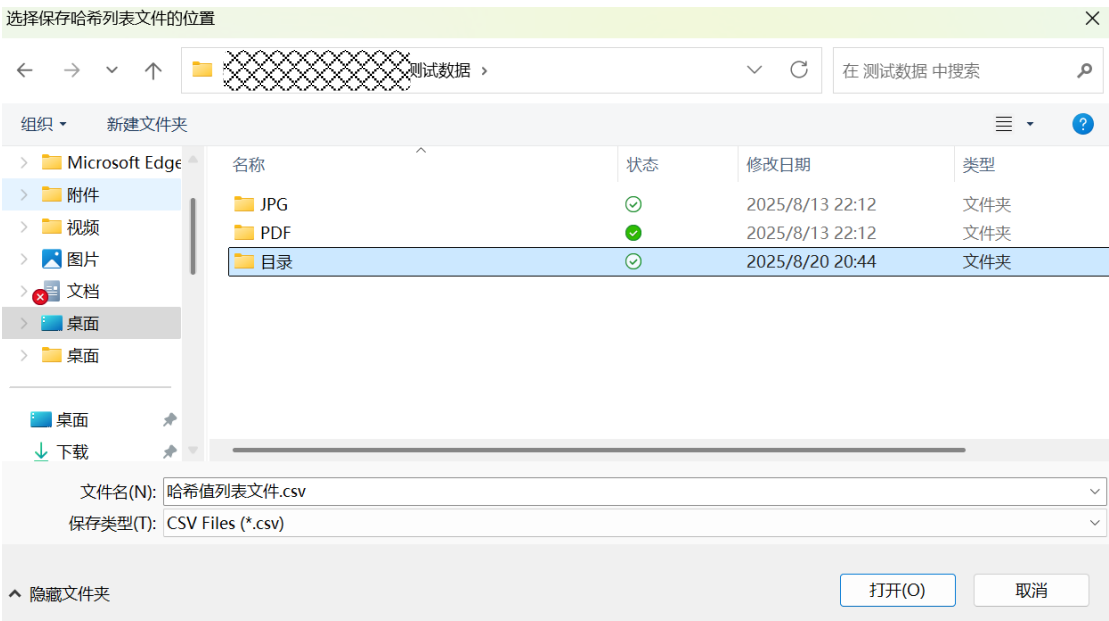


图 3-31 选择生成哈希值列表文件的位置



图 3-32 生成哈希值列表文件结果展示

4. 打包检测文件：将目录文件、哈希值列表文件、逻辑地址文件、物理地址文件等打包，生成“数字化验收检测包.rar”，存储在目录文件夹同一层级；



图 3-33 4. 要打包成 rar 的内容

先查看现在的“目录”文件夹下是否包含以下文件：1.数字化成果的目录 2.第二步生成的四个过程性文件 3.刚刚生成的“哈希值列表文件”。确保目录文件夹内容无误后打包成"数字化验收检测包.rar"，存放在“目录”同一层文件夹下。

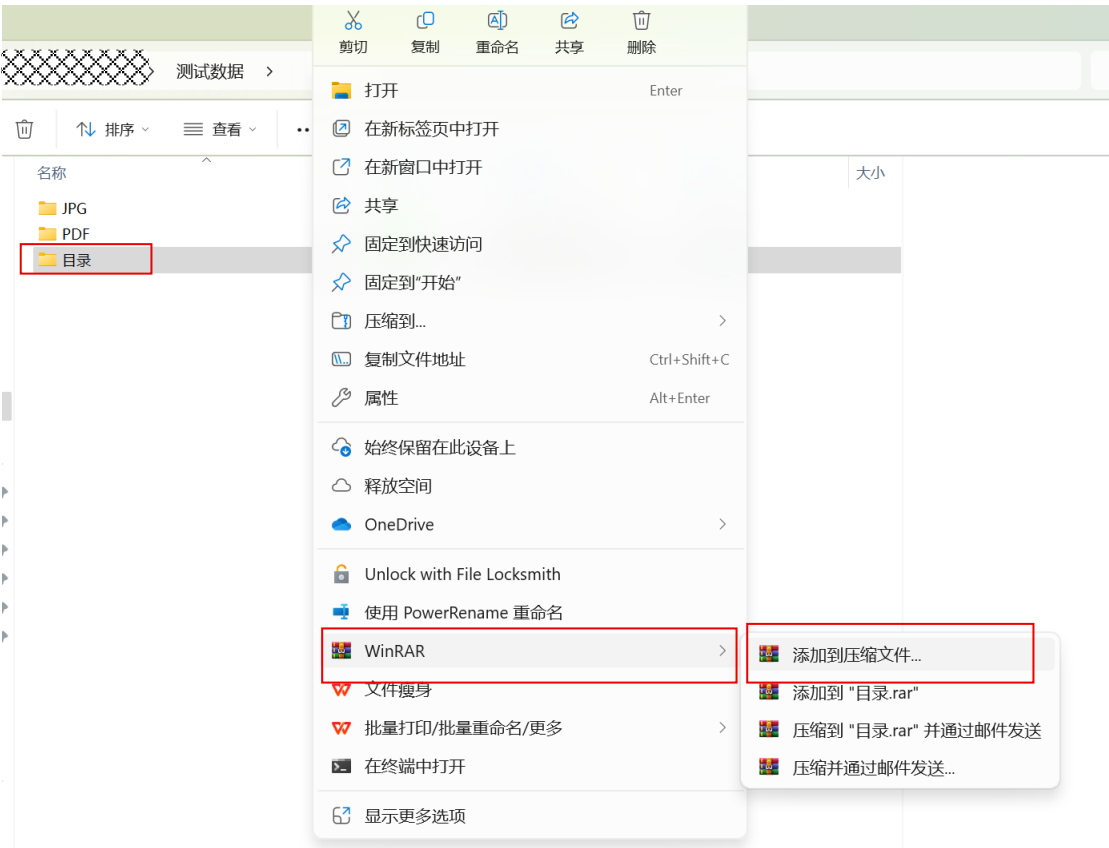


图 3-34 使用 winrar 打包检测文件



图 3-35 打包后的结构

5. 载入压缩包：点击“载入压缩包”的“选择文件”按钮，选中生成的“数字化验收检测包.rar”；



图 3-36 载入压缩包

6. 计算哈希值：点击“计算哈希值”按钮，将结果填入《存储载体检查登记表》。



图 3-37 6. 计算哈希值结果展示

至此，“测试数据”检查完毕！

4.常见问题

4.1 忘记检查结果

解决方案：查看软件解压包下的“logs”文件夹，其中“release.log”为当日检查结果日志，历史记录可查看“release-（对应日期）”格式的日志文件。

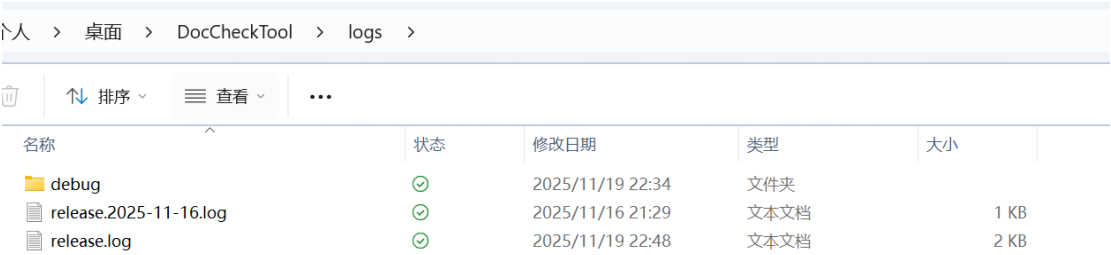


图 4-1 日志文件所在位置

解决方法：到“文件”-“设置”里面增加“步骤任务超时时间”。

4.2 查漏、存储路径和命名规范模块报错

（若要检查的“总和”目录内容如下所示）

测试数据目录									
档号	文号	责任者	题名	页数	形成日期	保管期限	密级	备注	
9027-WS2015-Y-0001		联考重要会议办公室	2015年“四个全面”重要论述考点标注版	3	20150124	Y			
9027-WS2015-D30-0001			存在空白档号和页数	5					
9027-WS2015-Y-0002	联时发[2015]1号	联考时事局	2015年1月时事政治考点标注版	5	20150208	Y			
9027-WS2015-Y-0003	联时发[2015]2号	联考时事局	2015年2月时事政治（重点标注版）	5	20150306	Y			
9027-WS2015-Y-0004	联时发[2015]3号	联考时事局	2015年3月时事政治考点标注版	5	20150409	Y			
9027-WS2015-Y-0005	联时发[2015]4号	联考时事局	2015年04月国内时事政治考点标注版	5	20150507	Y			
9027-WS2015-Y-0006	联时发[2015]5号	联考时事局	2015年05月国内外时事政治考点标注版	6	20150611	Y			
9027-WS2015-Y-0007	联时发[2015]6号	联考时事局	2015年6月时事政治考点标注版	7	20150712	Y			
9027-WS2015-Y-0008	联时发[2015]7号	联考时事局	2015年7月时事政治考点标注版	5	20150801	Y			
9027-WS2015-Y-0009	联时发[2015]8号	联考时事局	2015年8月时事政治考点标注版	8	20150910	Y			
9027-WS2015-Y-0010	联时发[2015]9号	联考时事局	2015年9月时事政治考点标注版	6	20151009	Y			
9027-WS2015-Y-0011	联时发[2015]10号	联考时事局	2015年10月时事政治考点标注版	7	20151112	Y			
9027-WS2015-Y-0012	联时发[2015]11号	联考时事局	2015年11月时事政治考点标注版	8	20151204	Y			
9027-WS2015-Y-0013	联时发[2015]12号	联考时事局	2015年12月时事政治考点标注版	10	20151231	Y			
9027-WS2015-Y-0014		联考重要会议办公室	2015年全国两会十大热点	2	20150525	Y			
9027-WS2015-Y-0015	联时发[2015]1号	联考时事局	2015年两会十大热点	10	20151110	Y			

图 4-2 目录内容

4.2.1 “生成逻辑地址文件”时报错一

“Excel 文件缺少列标题行”与“Excel 文件中未找到 ‘ 档号 ’ 列或 ‘ 页数列’ ”



图 4-3 缺少列标题行



图 4-4 列名缺失报错示例

解决方案:

1. 检查目录文件是否存在多余的总标题行，若有则删除；
2. 核实列名是否准确，需包含“档号”“页数”列（文字间无空格）；
3. 若为 xml 或 dbf 格式目录，参考附录 表 1 载入各类型目录允许的字段名 / 列名

4.3.2 “生成逻辑地址文件”时报错二

“第？行档号为空，停止处理” 或 “第？行页数为空，停止处理”

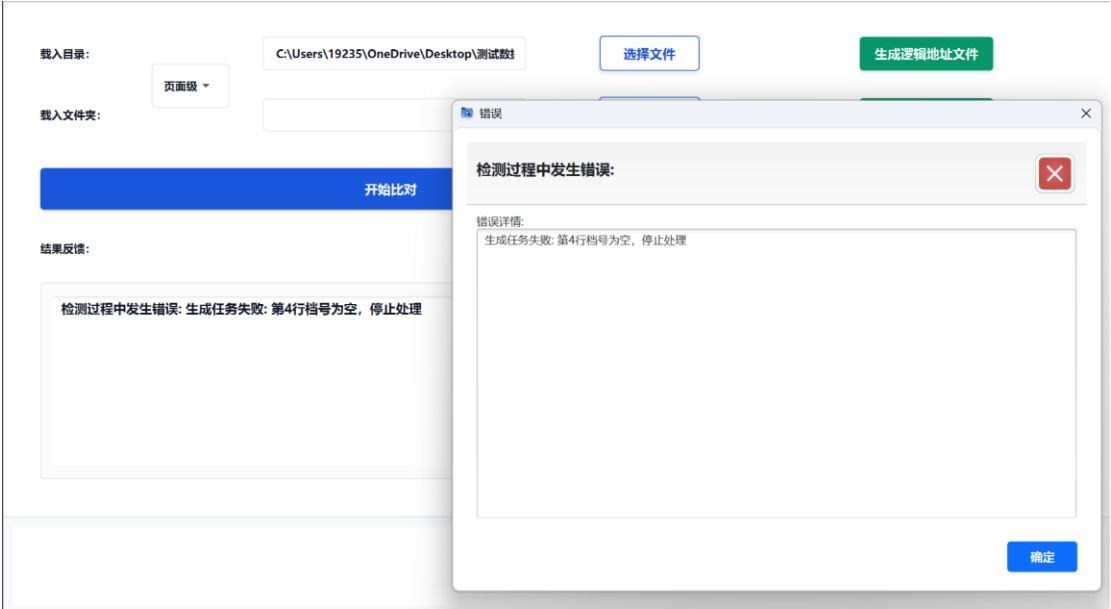


图 4-5 数据为空报错示例

解决方案：检查对应行的“档号”或“页数”列数据，确保无空值后重新操作。

4.3.3 “生成逻辑地址文件”时报错三

“存在重复档号”

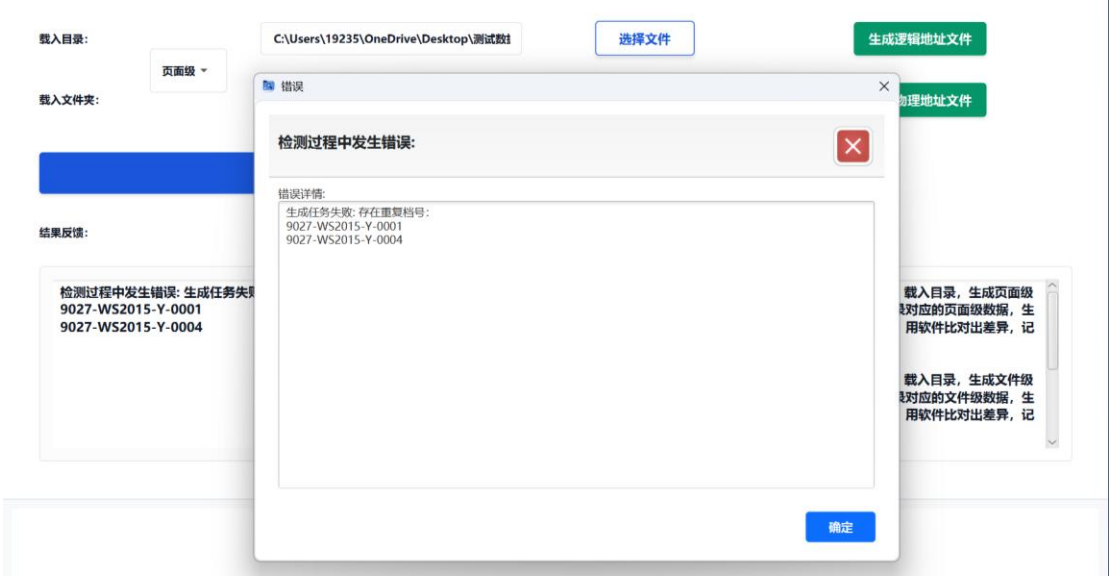


图 4-6 重复档号报错示例

解决方案：核查目录文件“档号”列数据，根据实体档案修改重复的档号，确保唯一性后重新操作。

4.3.4 “生成物理地址文件”时报错

“存在重复文件名”

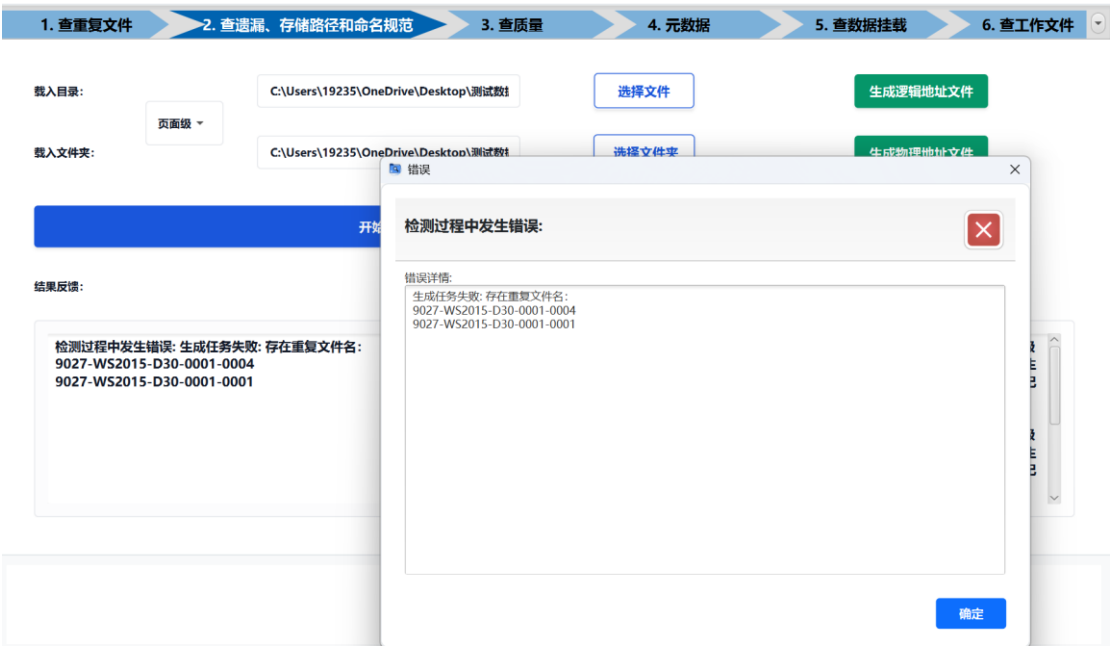


图 4-7 重复文件名报错示例

解决方案：检查对应的 JPG 或 PDF 文件夹，核实是否有相同文件名的文件，根据实际内容修改后重新操作。

5.附录

文件类型	表示“档号”	表示“页数”	Xml 中表示一行记录的根元素标签
xls	档号	页数/包含“页数”的列名	
xlsx			
dbf	档号	页数	
	DH	YS	
	fileNo	pages	
	dangan	pageCount	
xml	档号	页数	"row", "record", "data", "item", "档案"
	DH	YS	
	fileNo	pages	
	dangan	pageCount	

表 1 载入各类型目录允许的字段名 / 列名