

# 纸质档案数字化验收标准体系研究

## 项目研究报告

云南省档案局（馆）

科研项目课题组

2012.11

国家档案局官网  
WWW.SAAC.GOV.CN

# 目 录

引言.....	1
一、国内外同类项目研究现状.....	2
1.1 国外研究现状.....	2
1.2 国内研究现状.....	3
二、项目工作基本情况.....	4
2.1 项目名称.....	4
2.2 项目承担单位.....	4
2.3 项目课题组成员.....	4
2.4 项目研究的目的意义及立项背景.....	4
2.5 项目研究的主要内容.....	5
2.5.1 数字化加工安全保密验收.....	5
2.5.2 档案图像扫描验收.....	5
2.5.3 图像处理验收.....	6
2.5.4 图像存储验收.....	6
2.5.5 目录建库验收.....	6
2.5.6 数据挂接验收.....	6
2.5.7 档案文件级目录及题名验收.....	6
2.5.8 档案整理验收.....	6
2.5.9 数据备份验收.....	6
三、项目组织实施情况.....	7
3.1 项目组织.....	7

3.1.1 项目研究.....	7
3.1.2 项目基础调研.....	8
3.1.3 项目完成情况.....	10
3.2. 项目研究人员、经费投入及使用情况.....	11
3.2.1 项目研究人员情况.....	11
3.2.2 经费投入及使用情况.....	11
<b>四、项目研究成果.....</b>	<b>11</b>
4.1 数字化加工安全保密验收.....	11
4.2 档案扫描验收.....	13
4.3 图像处理验收.....	17
4.4 图像存储验收.....	20
4.5 目录建库验收.....	22
4.6 数据挂接验收.....	21
4.7 档案文件级目录及题名验收.....	29
4.8 档案整理验收.....	37
4.9 数据备份验收.....	38
<b>五、项目研究成果推广应用.....</b>	<b>43</b>
<b>六、项目研究报告及相关验收规定、办法、教材及表单... 44</b>	<b>44</b>
6.1 档案数字化验收规定、办法及教材.....	44
6.2 档案数字化验收申请、报告及表单.....	45

## 四、项目研究成果

《纸质档案数字化验收标准体系研究》科研项目现已获得以下几方面的研究成果：

### 4.1 数字化加工安全保密验收

安全保密包含两方面的内容：实体安全保密和信息安全保密。通过软硬件及人员管理，实现全方位的安全管理。包括网络平台搭建、监控设备的安装（16个摄像头）等。

本单位的数字化管理中心对加工公司实行流程管理，通过对加工人员进行岗前培训，增强人员安全意识、保密意识、责任意识等，确保档案在数字化加工过程的安全完整。

首先，验收小组检查加工公司是否与省档案局签订“信息保密协议”、公司是否办理《中华人民共和国国家秘密载体复制许可证》、工作人员是否持有云南省保密局颁发的《涉密人员上岗培训证》。

第二、档案原件是否严格按照档案管理要求进行管理。在未拆卷以前，如发现有缺失文件现象，必须填写《馆藏档案卷内文件材料登记备查表》，经签字认可后再进行拆卷处理。

第三、下班之前，所有档案是否进箱、柜，档案箱、柜钥匙要由加工公司专人保管，并且要由专人检查水电、窗户是否已关闭完好后，方可离去。

第四、在加工过程中，如发现文件属密级以上，是否进行登记，文件所涉内容。同时，不能随便在文件上添写任何字及符号，只能用2b铅笔编写页码。

第五、严禁将手机、相机、摄像机、私有移动存储介质等设备带入加工现场。由于工作需要，每个公司只允许两名管理人员可以在加工现场使用手机，其余员工只能在加工现场外或下班之后才能使用手机，如有违反上述规定，一律严惩。

第六、验收小组查看现场时，注意下列事项：数字化加工过程中所使用的扫描仪、声频（视频）采集卡等设备，必须符合有关保密技术标准和保密要求，主机不得有U盘插口，数字化加工所使用的计算机及其存储介质，必须严格遵守计算机有关保密管理规定。使用的移动硬盘、U盘、光盘等存储介质，按涉密载体进行管理，不得在非涉密计算机及其网络上使用，并且每天公司拷贝数据所用硬盘等，必须由专人在管理中心进行登记后，方可领取使用，同时履行归还登记手续。加工公司在加工合同履行结束后撤离现场时，所使用的电脑硬盘、移动硬盘等数据载体需移交业主单位。

## 4.2 档案扫描验收

档案扫描是档案数字化加工的中心环节，加工公司必须根据所扫描档案幅面的大小选择相应规格的扫描仪或专业扫描仪进行扫描。档案扫描验收，主要检查扫描色彩模式是否按照合同的分辨率要求和模式情况进行，是否反映档案的原貌等。刚开始数字化加工工作的时候，我们也

有过一些争议，主要就是在档案是提供一般性利用，还是可以提供特殊利用，比如原文出版等，因为两者所要求的存储空间差别很大。根据这几年来来的数字化扫描加工工作的进一步发展，存储介质的飞跃发展，我们也逐渐统一了思路，规定了档案扫描的一些基本标准。比如，扫描分辨率：文件均采用彩色色彩模式，分辨率 $\geq 300\text{dpi}$ ；照片档案： $< 5$ 寸、 $600 \times N$ ， $5$ 寸、 $600\text{dpi}$ ， $> 5$ 寸、 $600 / N$ 。（注：5寸照片采用 $600\text{dpi}$ 分辨率；原照片小于5寸照片 $N$ 倍的，则分辨率为 $600 \times N$ ；原照片大于5寸照片 $N$ 倍的，分辨率为 $600/N$ ）。但在数字化扫描及验收的过程中，发现了很多意想不到的特殊情况。例如：

**4.2.1** 在扫描中发现档案纸张存在油性纸、复写纸、铅笔字、字迹模糊、纸张较差及较薄等情况。具体做法是：在保持档案原貌的前提下进行扫描，并提高分辨率或采用垫白纸再进行扫描，尽量反映档案的原始原貌。

**4.2.2** 在扫描中发现部分档案原件纸张的底色较深，所标页码扫描后看不清楚。具体做法是：用一张白色小纸条写上该页页码，放在该页右上角，然后再进行扫描。

**4.2.3** 在扫描中发现有重份档案、重份照片、重份表格等情况。具体做法是：只扫描一份。

**4.2.4** 在扫描中存在多张零散照片。具体做法是：将照片平铺在白纸上，并在白纸右上角标上页码后再进行扫描。

**4.2.5** 档案中存在部分超大、超长、图纸及表格，无法进行扫描。具体做法是：用白纸打上“此页为特大图纸”字样，扫描此页来代替特

大图纸并在备注栏注明。如：



4.2.6 在扫描中发现档案上有多张粘贴的小纸条或部分档案纸张粘贴原文且破损严重等情况。具体做法是：为保持档案原貌，用平版扫描仪进行扫描，尽量将档案可视部分扫描出来，如果有遮挡部分，重新进行标页后再另扫一页。

4.2.7 档案中有部分大画幅（大于 A3 幅面）无法一次性扫描。具体做法是：分幅扫描后进行完整拼接或用相机拍摄。如：

国家档案局官网  
WWW.SAAC.GOV.CN



4.2.8 档案数字化过程中存在照片底片、锦旗、奖状等实物档案、音像档案。具体做法是：把实物档案放于白纸上并在右上角标注相应的页码进行拍照或只对音像封面进行扫描。如：





4.2.9 在扫描民国档案时，很多是桐子页，由于纸张较破较薄，正反面都有内容，如果正常扫描，就会把背面的内容印到正面。具体做法是：扫描时在桐子页中间夹放一张白纸。

4.2.10 在扫描民国档案时，档案原件存在纸张较差、破损严重，虫蛀、霉变、老化、缺字、字迹模糊等情况。具体做法是：尽量照档案原样进行扫描，字迹模糊的在扫描时垫上“原件字迹不清晰”进行扫描，有的老化已碎的，用专用信封将已碎的档案装袋不进行扫描。

4.2.11 在扫描民国档案中发现部分档案有严重粘贴或无法展平等情况。具体做法是：扫描成多幅图像，再进行图像拼接。

另外，我们要求各公司在扫描过程中，凡是发现或遇到档案存在的各种情况，都必须填写数字化加工备查表。如：

国家档案网  
www.saac.gov.cn

数字化加工备查表

全宗号	XX全宗	目录号		案卷号	
数字化加工公司	云南数字化	法人	张三	加工时间	2012年1月
拆卷人	田晓	扫描人	张呈文	装订人	毛艳平
数字化加工情况说明： 第1、159页原件破损/第1-5页原件严重破损/整本破损/整本霉烂/整本虫蛀/此卷有实物档案/第25、169页严重霉烂未扫描、装订/邮票12张					
填写要求：凡卷内纸质目录信息与卷内档案文件信息不一致、有重大改动和错漏字的以及其他需要说明的。					

以上做法，是我们在数字化扫描过程中以及在进行多次验收的基础上，总结出来的常见特殊情况及处理方法，我们也对各公司进行了规范和统一。在数字化档案的验收过程中，我们也按照以上的标准，对抽查的每一份档案进行逐页的检查验收，让每一份、每一页档案都能做到有来源、有记录，最大限度的保证档案的齐全完整。

### 4.3 图像处理验收

档案扫描图像处理结束后加工公司必须对扫描图像的偏斜度、清晰度、失真度等进行质检，对存在偏斜、污渍、不完整等问题的扫描图像进行纠偏、去污、拼接等技术处理，确保图像完整无误。加工公司质检完成后须认真填写“档案数字化流程工作单”，对图像处理结果进行相应说明。

# 档案数字化加工流程单

本卷	全宗号	目录号	案卷号	总页数	页	备注
前 处 理	实际整理:		条目录	张纸	签名:	
	存在问题	A 纸张破损 B 跳号 C 内容缺失 D 内容模糊 E 原件污渍 F 原件粘贴 G 原件缺失 H 超大幅 J 缺字 K 内容重复 L 漏编页码 M 重条 N 跳条 O 缺条 P 题名不正确 Q 多起止页号 W 页码重编 R 重条 S 虫蛀				
	整理结果					
前 处 理 检 查	实际检查:		页		签名:	
	存在问题	A 漏编页码 B 页码重编				
	整理结果					
著 录	实际扫描:		条		签名:	
	存在问题					
	整理结果					
质 量 检 查 (条 目)	实际处理:		条		签名:	
	存在问题					
	整理结果					
文 件 扫 描	实际处理:		幅		签名:	
	存在问题	A 大于A3 B 大于四张A3 C 铅笔字 D 灰度 E 真彩 F 原件内容模糊 G 原件污渍 H 原件粘贴				
	整理结果					
图 像 处 理	实际核查:		幅		签名:	
	存在问题	A 纸张破损 B 字迹不清 C 原件模糊 D 印章不清 E 原件内容缺失 F 拼接大图 G 杂点多 H 重页 J 背面 K 重扫				
	整理结果					
质 量 检 查 (图 像)	实际核查:		幅		签名:	
	存在问题	A 字迹不清 B 背面 C 字被遮 E 字被裁 E 空白页 F 文件名不对 G 图像处理不干净 H 纠偏 J 去污 K 裁边				
	整理结果					
装订还原	<input type="checkbox"/> 贴边处理		<input type="checkbox"/> 无遗漏		签名:	
装订检查	<input type="checkbox"/> 页码颠倒		<input type="checkbox"/> 订漏		签名:	

初验和验收，重点从图像数据完成质量、是否纠偏、是否去污三个方面进行审核检查。在图像处理过程中，要求保持档案信息真实原貌，不得擅自修改档案信息。

**图像数据质量：**对图像偏斜度、清晰度、失真度、完整性及漏扫、排列顺序与原件的一致性进行检查。

**纠偏：**对出现偏斜的图像必须进行纠偏处理，以达到视觉上基本不感觉偏斜为准，允许的倾斜度不得大于1度；对方向不正确的图像必须进行旋转还原，以符合阅读习惯。以达到视觉基本不感觉偏斜为准，并符合阅读习惯。

**去污：**对图像页面中出现的会影响图像质量的杂质，如黑点、黑线、黑框、黑边等必须进行去污处理。处理过程中必须遵循在不影响可懂度的前提下展现档案原貌的原则。

档案扫描图像处理结束后加工公司必须对扫描图像的偏斜度、清晰度，失真度等进行检查。

对不符合图像质量要求的，重新进行图像处理。对图像文件不完整或无法清晰识别的，重新扫描。对漏扫文件必须及时补扫并正确插入图像。扫描图像的排列顺序与档案原件不一致时，必须及时进行调整。图像质量力求图像完整、清晰，无影响图像美观的黑边和污点，图像不偏斜或倒置，符合正常阅读习惯。对初验和验收过程中图像存在的质量问题，数字化管理中心和全省专家验收委员会认真填写“数字化档案质量验收记录表”图像处理情况，督促加工公司对有问题的图像数据进行相应的整改。

数字化档案质量验收记录表

全宗号	1011	目录号	001	案卷号	0000 1	文件目录数 (条)		画幅数量 (幅)	156
						合格文件目 录数(条)		合格画幅数 (幅)	155
文件总数 (件)		合格文件 数(件)		合格率 (%)		整体是否合 格(文件合格 率≥97%为合 格)		整体是否合 格标识(合格 为1, 不合格 为0)	
存在问题的内容记录									
文件号	档案整 理	图像 扫描	图像处 理	图像 存储	数据挂 接	数据 备份	拟定 题名	数据 检索	备注
001			第16页 漏扫						

验收人员签字： 年 月 日

#### 4.4 图像存储验收

加工公司在档案图像处理初验合格后，须对图像文件进行及时存储。存储格式分别采用 TIFF 格式和 PDF 格式两种形式。在保证扫描图像清晰可读的前提下，以尽量减少存储容量为原则，严禁以 Word 的插图形式保存为 Word 文件。图像数据存储：同一批次档案图像数据要求分别采用 TIFF 格式和 PDF 格式两种形式进行数据存储，母盘采用 TIFF

格式进行图像数据备份，数据挂接备份则采用 PDF 格式进行存储。档案原文图像数据审核验收主要检查加工公司在档案图像处理完成后是否按图像格式进行数据存储，凡存储格式不符合以上规定而采用 JPG 等格式保存的数据，一律重新处理图像格式。图像存储的验收主要是检查数字化加工公司所提交的光盘或者移动硬盘等其它存储介质，其图像存储的格式是否以 TIFF (LZW) 格式和 PDF 格式存储；存储介质中的每一份文件，是否都有一个与之对应的唯一档号，并以该档号为扫描后的该份图像文件命名。

首先检查数据格式是否标准，所选定的数据格式应能直接或间接通过 XML 文档进行数据交换；其次检查档案的著录，采用人工校对或自动校对的方式，对目录数据库的建库质量进行检查，核对著录项目是否完整、内容是否规范和准确。

## **4.5 目录建库验收**

目录建库工作是此研究项目体系中一项很重要的组成部分，它包括数据格式选择、档案目录数据库的建立和档号唯一性选择，在档案目录数据库中每一份文件赋予一个与之相对应的唯一的档号，该档号又作为这份文件扫描后的图像文件命名的唯一性，在目录数据库建库工作中显得十分重要。

### **4.5.1 数据格式选择**

我们在纸质档案数字化验收标准体系研究项目中，对目录建库主要是选择通用数据格式 (DBF、TXT、MDB)，各数字化加工公司的数据移交

时统一转换为·DBF格式，向云南省档案局（馆）的局域网数据服务器进行导入，其所选定的数据格式都能直接或间接通过XML文档进行数据交换。

#### 4.5.2 档案目录数据库的建立

按照国家档案局《档案著录规则》（DA/T 18）和《云南省档案馆档案文件级目录著录要求》等规定项目建立规范化的档案目录数据库，其结构如下：

**现行档案文件级目录数据库结构**

字段	字段名	类型	宽度
1、	全宗号	C（字符型）	4
2、	目录号	C（字符型）	3
3、	案卷号	N（数值型）	4
4、	卷内顺序号	N（数值型）	4
5、	档号	C（字符型）	40
6、	案卷档号	C（字符型）	40
7、	责任者	C（字符型）	100
8、	题名	C（字符型）	250
9、	人名索引	C（字符型）	250
10、	文件编号	C（字符型）	80
11、	文件日期	T（日期型）	8
12、	日期备注	C（字符型）	12
13、	起页号	N（数值型）	4
14、	止页号	N（数值型）	4
15、	页数	N（数值型）	4
16、	保管期限	C（字符型）	4

国家档案局官网  
www.saac.gov.cn

17、	录入时间	T (日期型)	8
18、	录入员	C (字符型)	10
19、	录入部门	C (字符型)	16
20、	扫描部门	C (字符型)	16
21、	密级	C (字符型)	4
22、	备注	C (字符型)	40

### 历史档案文件级目录数据库结构

字段	字段名	类 型	宽度
1、	全宗号	C (字符型)	4
2、	目录号	C (字符型)	3
3、	案卷号	N (数值型)	5
4、	卷内顺序号	N (数值型)	3
5、	档号	C (字符型)	40
6、	案卷档号	C (字符型)	40
7、	责任者	C (字符型)	100
8、	题名	C (字符型)	250
9、	文件日期	T (日期型)	8
10、	日期备注	C (字符型)	12
11、	起页号	N (数值型)	4
12、	止页号	N (数值型)	4
13、	页数	N (数值型)	4
14、	保管期限	C (字符型)	4
15、	录入时间	T (日期型)	8
16、	录入员	C (字符型)	10
17、	录入部门	C (字符型)	16
18、	扫描部门	C (字符型)	16

19、	密级	C (字符型)	4
20、	备注	C (字符型)	40

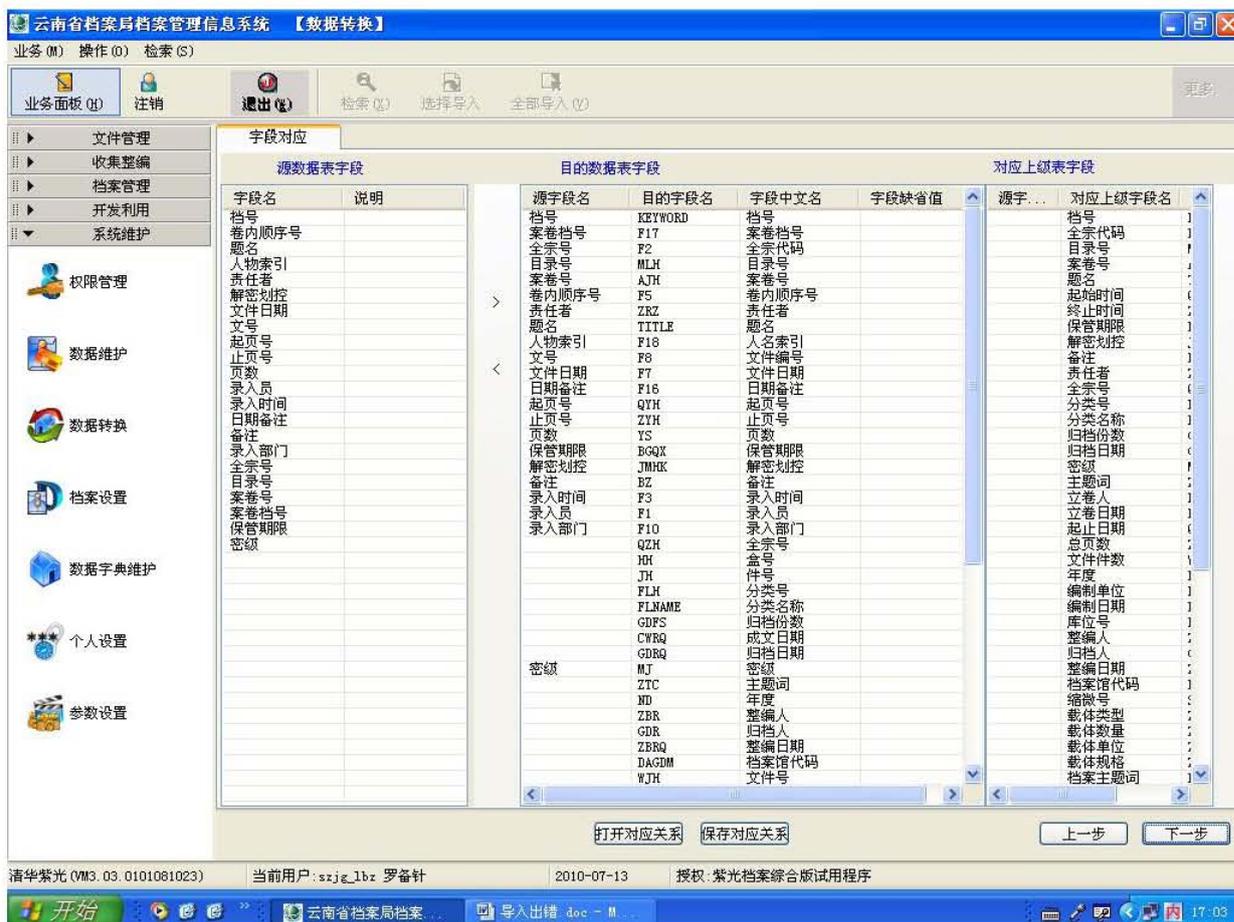
### 归档文件（新规则）目录数据库结构

序号	字段名称	字段类型	字段长度
1、	档案馆代码	字符型	6
2、	编制单位	字符型	30
3、	全宗号	字符型	4
4、	年度	字符型	4
5、	保管期限	字符型	4
6、	机构（问题）	字符型	20
7、	件号	数字型	5
8、	档号	字符型	60
9、	责任者	字符型	100
10、	题名	字符型	250
11、	人名索引	字符型	250
12、	文件编号	字符型	100
13、	时间	日期型	8
14、	日期备注	字符型	12
15、	页数	数值型	4
16、	密级	字符型	4
17、	备注	字符型	30
18、	录入时间	日期型	8
19、	录入员	字符型	20
20、	录入部门	字符型	20
21、	扫描部门	字符型	16

#### 4.5.3 档号唯一性选择

档案目录数据库中的每一份文件都有一个与之相对应的唯一档号，必须以该档号作为这份文件扫描后的图像文件命名。

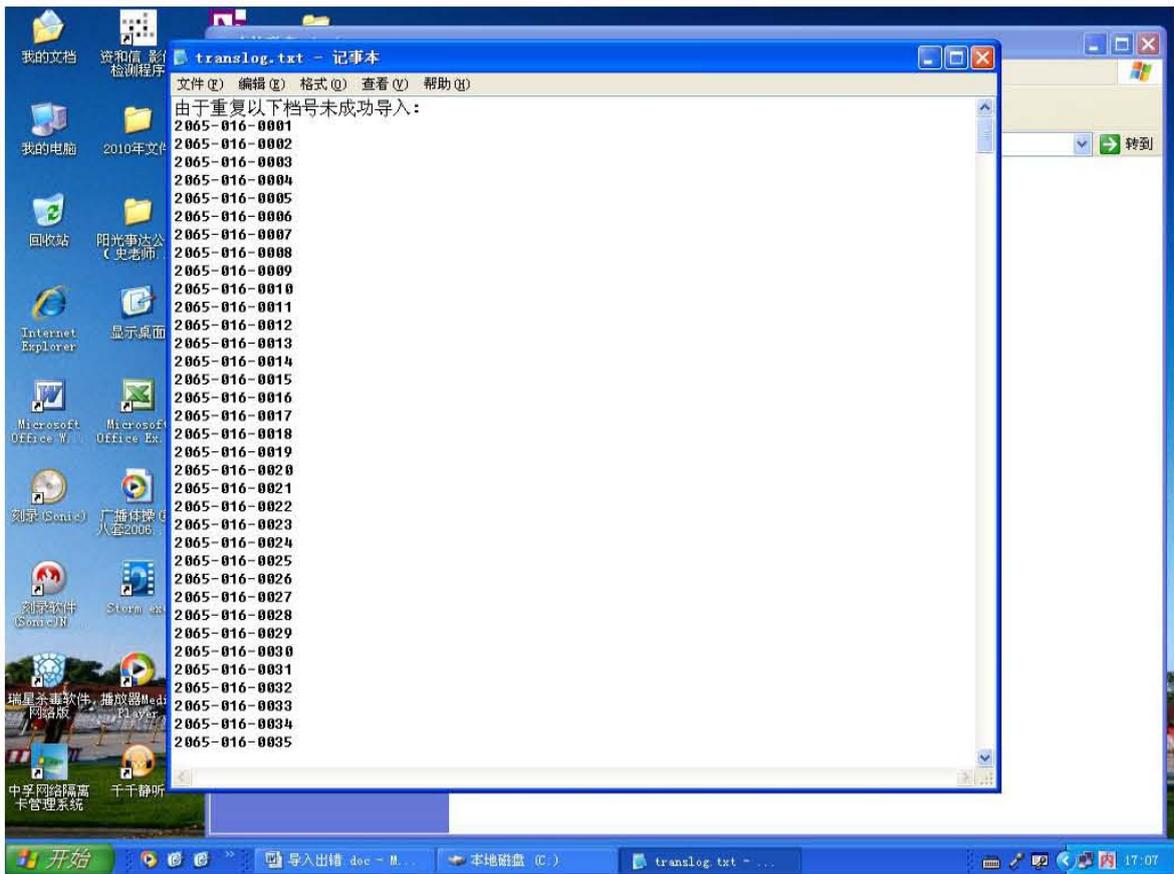
### 【目录数据导入的字段对应关系】



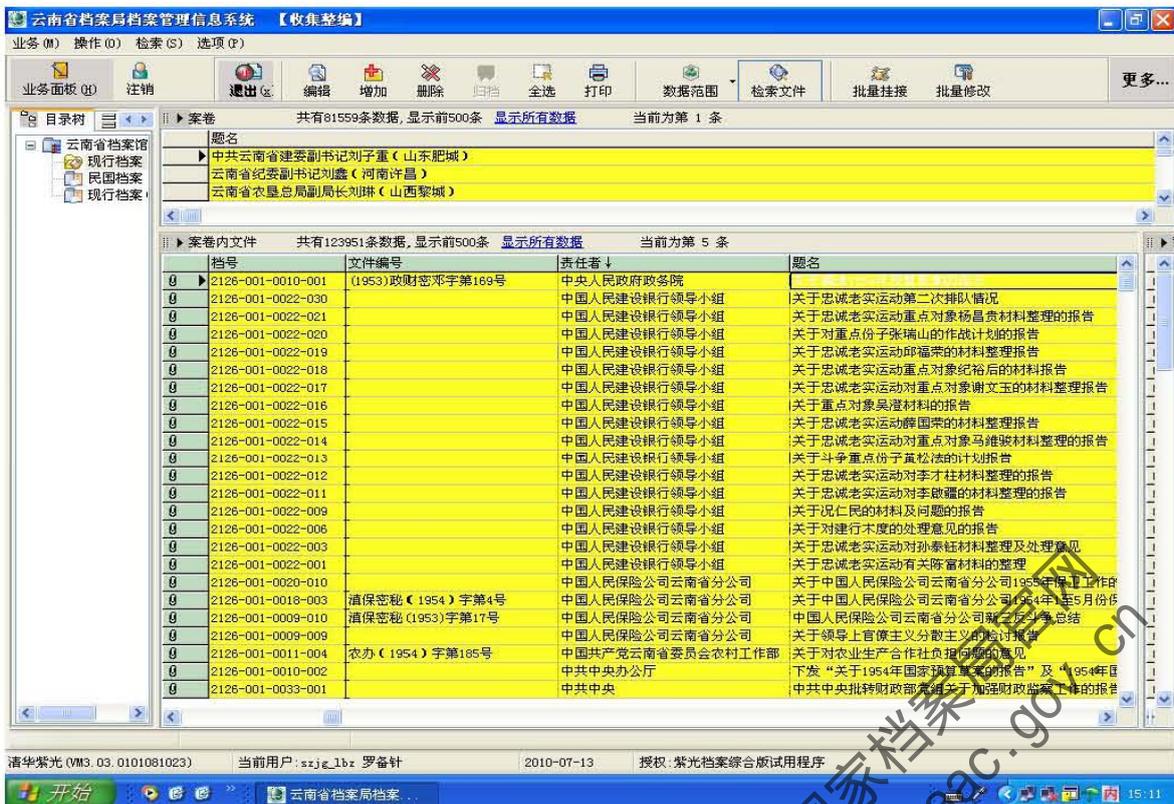
如果档号重复，系统提示退出不予以导入。如下图的数据就是档号未成功导入目录的提示：

### 【未导入数据详细记录提示】

国家档案局官网  
WWW.SAAC.GOV.CN



## 【检查核对目录内容】

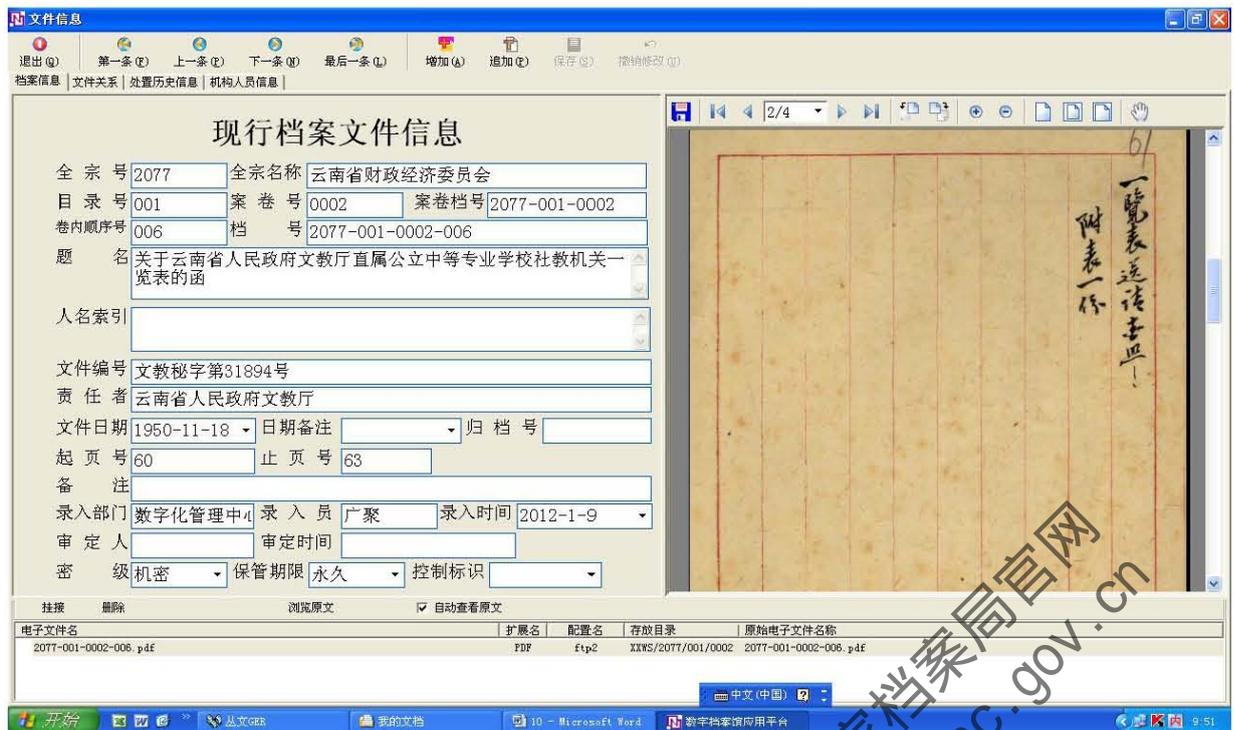


## 4.6 数据挂接验收

挂接是档案数字化转换过程中形成的目录数据库与图像数据库，通过质检环节确认为“合格”后，各数字化加工公司将其加载到云南省档案局（馆）的数据服务器上。通过紫光档案管理系统软件，以纸质档案目录数据库为依据，将每一份纸质档案文件扫描所得的一个或多个图像存储为一份图像文件。通过每一份图像文件的文件名与档案目录数据库中该份文件的档号的一致性和唯一性，建立起一一对应的关联关系，便可批量、快速地进行挂接。

数字化加工公司将每一份纸质档案文件扫描所得的一个或多个图像采用多元 TIFF 格式存储成单个图像文件夹，同时与云南省档案馆档案目录数据库对接，将图像数据挂入数据库中。

### [ 图像挂接后的结果 ]



数据挂接验收，主要是检查图像数据与档案实体的页码总数是否一致，图像文件的名称与档案目录数据库中该份文件的档号、标题是否相同，图像文件总数与目录总数是否一致，挂接的数据是否有重复、遗漏等。

## [目录与图像文件检查方法]

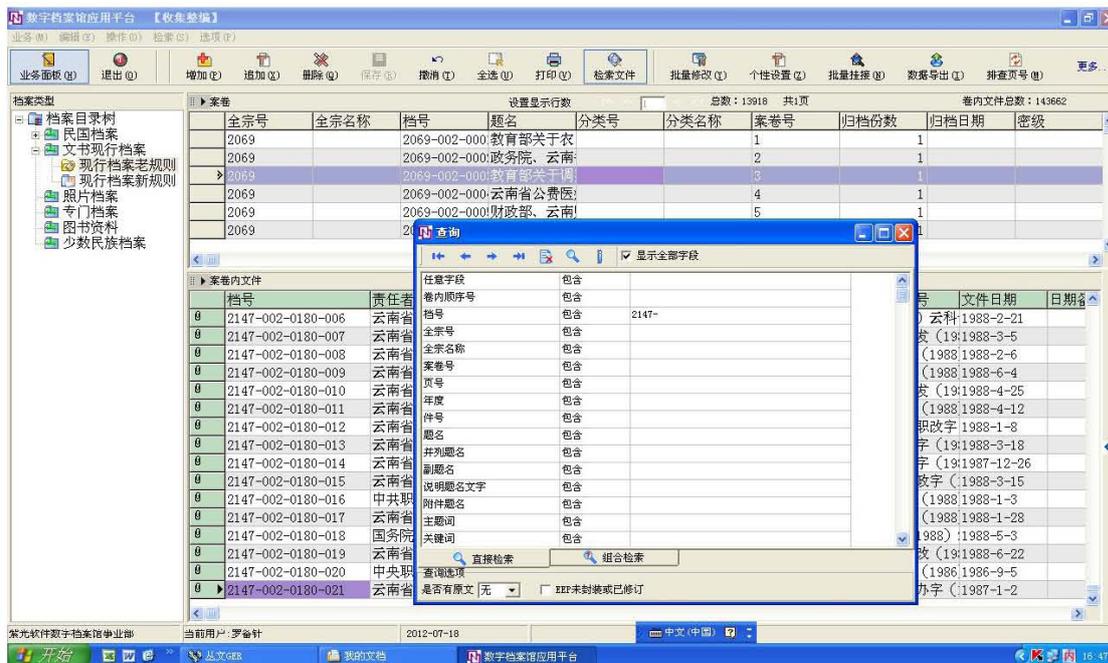


## [图像文件总页统计方法]



在挂接后，我们使用档案管理软件即可测试出其是否多挂、少挂或漏挂。如下图是查找漏挂图像的检测方法：

### [图像漏挂检测方法]



检查挂接情况，数据挂接后，以档案文件级目录数据库为依据，对挂接的图像文件进行检查，核对图像文件的命名格式是否符合要求，发现错误及时做出修正。