

5. 明清档案数字化质量管理体系建设

本体系依托一史馆大规模数字化外包项目，核心是项目管理和自动化技术应用。下面，我们就以一史馆数字化项目为例，根据数字化加工流程介绍质量管理体系建设。

5.1 明清档案数字化质量管理之项目管理

根据数字化质量管理体系框架，项目管理体系包括组织管理、人员管理、制度管理、设备管理 4 个方面工作。

5.1.1 组织管理

项目采取外包形式，一史馆为甲方，聘用专业数字化公司承担项目实施。一史馆负责提出专业要求，对项目实施过程进行监管，审核验收最终提交数据；外包公司主要负责项目具体实施。

一史馆方面，设立三级管理组织，由领导小组、项目组和项目监管组构成。数字化领导小组为项目最高层级管理组织，负责项目宏观协调和最终验收。数字化项目组负责外包项目的具体实施，设立安全和技术总监，分别负责项目实施过程安全和数字化加工技术要求管理。项目监管组下设现场监管小组和质检工作小组，分别负责数字化加工现场项目实施的监管和提交数据的初步审查验收。

外包公司方面，同样设立三级管理组织，由公司管理层、项目经理和业务单元三级管理机构组成。公司管理层负责与馆方数字化项目组对接，负责公司项目整体实施和资源保障。项目经理

与馆方项目监管组对接，负责项目具体实施，负责项目安全、质量和进度的整体控制。业务单元负责人受项目经理直接管理，负责本业务模块工作的具体指导和管理，负责项目的执行和操作。

组织关系如图：

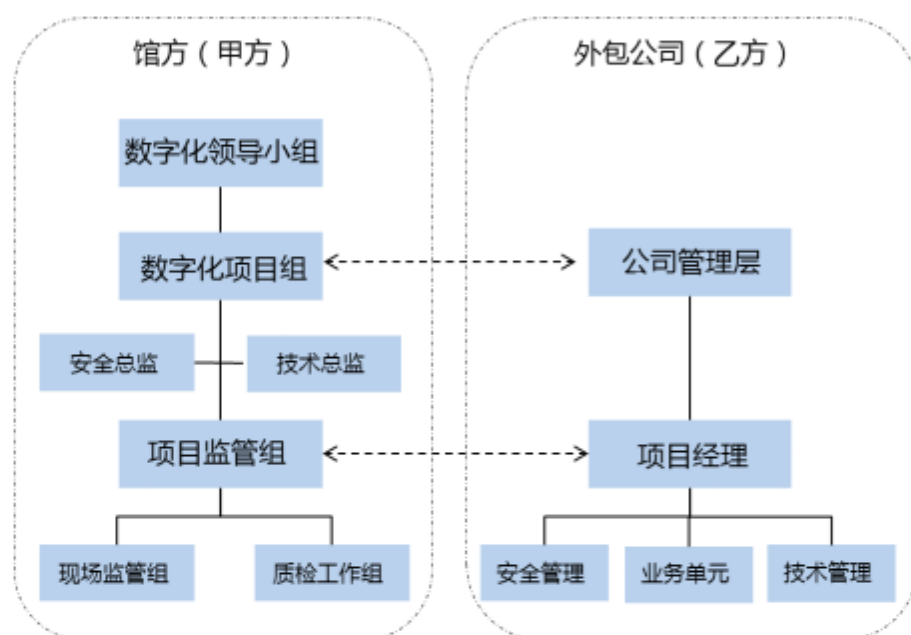


图 7 组织关系图

上述管理组织设置，由宏观协调到整体实施，再到现场具体监管，层级分明，职责明确，对接关系清晰。一方面有利于各方职责分工明确，确保整体工作有序，另一方面有利于信息上传下达，极大地提高了沟通效率，对于数字化加工全面质量管控的实现起到了有效支撑。

5.1.2 人员管理

数字化外包工作实施最终通过各业务单元工作人员来实现全面质量的管控，关键在于人员管理工作的好坏。一方面应该确保工作人员数量的稳定，另一方面要确保工作人员业务能力的

稳定。通过对公司人员数量、业务能力的控制，以及数字化加工过程数据流转责任追溯机制，从而实现数字化图像加工质量的稳定。

1. 建立及时纠错及责任回溯机制。在数字化加工各个环节制定详细的登记、交接和数据流转登记制度，明确数字化加工过程中各个环节责任人，实行流转登记核查，加强数字化加工过程执行监督，确保及时发现问题、及时追溯、及时改正。

2. 实施考核合格上岗制度。要求公司对员工进行业务培训和技能考核，公司考核合格后，项目监管组对公司培训考核合格的员工再次进行考核，考核通过后方可上岗，确保数字化加工人员业务能力达标。

3. 保持外包公司加工人员稳定。严格要求公司管理及加工人员数量，在项目期间，公司管理人员原则上不能更换，每天保证有三名管理人员到场；加工人员每月变动原则上不能超过总人数的10%；建立公司管理人员和加工人员出勤登记簿；建立公司人员每月人员流动统计表。

4. 建立公司人员备份机制。要求公司配备一定数量业务能力合格的机动人员，防止人员变动造成公司加工人员数量和业务能力下降，从而确保数字化加工各环节的有效运转和数字化加工质量的稳定。

5. 人性化管理。

根据全年工作计划，制定灵活而人性化的放假安排；敦促公司制定了积极的激励政策，对每月表现优秀的员工进行奖励；敦促公司制定员工出游计划，不定时的带员工进行户外活动放松身心，调节情绪；敦促公司出台奖励政策，每月评选明星员工，评选出表现优秀的员工，将其照片贴到明星墙上给予公示表扬，每年年底评年度选优秀员工并进行奖励。

5.1.3 制度管理

1. 管理制度

项目的有效实施离不开完善的制度保障。根据一史馆数字化加工项目特点，我们针对质量管理体系框架的各个组成部分，建立了包括人员管理、现场管理、设备管理、安全管理、数字化加工项目管理以及数字化加工质量管理等一整套制度体系。内容涵盖数字化加工安全规范、技术标准、操作要求、加工人员行为规范、现场监管、设备运维以及流程要求等，对数字化加工项目实施提供了依据。具体如：《数字化加工及管理基本准则》、《数字化加工档案原件安全管理实施细则》、《档案整理及数字化项目验收管理办法》、《数字化加工现场管理办法》、《数字化加工设备管理办法》、《数字化加工XX环节操作规范》以及公司人员行为规范、培训考核、奖惩制度等。

2. 质量规范

由于历史档案形制复杂、档案状况不一，很难用一个数字

化加工标准来衡量。为此，我们在建立完善制度保障体系基础上，采取分层设计的方法建立整个数字化加工质量规范体系。首先建立数字化加工质量基础标准，确定图像质量基本要求和操作规范，划定数字化加工图像质量红线；其次根据不同类型和保管状况的档案，在基本质量要求和操作规范的基础上制定详细的质量标准和实施细则，以平衡数字化加工过程中档案安全、质量和进度的关系；最后，统一项目质检工作程序及标准用语，将常见错误分类并制定相应处理机制，确保执行统一，最终实现数字化加工质量检查标准化、可控化。具体如：《纸质档案数字化加工技术规范》、《馆藏档案数字图像命名管理规定》、《纸质档案数字化加工验收标准》、《数字化加工元数据要求》、《纸质档案数字化加工实施细则》等。

5.1.4 设备及软件管理

课题组认为，实现生产的流水化和标准化是确保数字化加工质量的前提，为此，项目组为外包公司提供了统一型号的快速平板式扫描仪，设置了统一规格的画幅尺寸、扫描位置和图像参数。在设备使用上，所有设备进入加工现场前都经过检查和调试安装，设备的使用范围和功能也都进行物理和软件限制，以确保数据的安全和稳定。在设备运维方面，除每天对加工现场和设备进行除尘等日常清洁外，由专人定期检查玻璃板、背板等设备配件，通过色卡检查设备色彩状况，一旦出现玻璃划痕、背板变色等，由

专业人员进行更换。

在软件设计方面，针对一史馆历史档案的项目特点，我们整体上设计了扫描、优化、著录、校对、质检、自动挂接、查询检索、数据转换、数据上传、数据备份、数据打印等数字化加工相关功能。支持“目录数据库”和“图像数据库”的创建、录入和导入、导出，并自动批量挂接，支持实时工作量的统计输出，并能够自动生成工作进度统计图表。

5.2 数字化加工各环节质量管理及技术应用

5.2.1 数字化加工环节 workflow 技术应用

1. 数字化加工环节介绍

数字化加工项目实施阶段，共分为扫描前处理、数字图像扫描以及成品数据质检验收三个模块，包括档案前处理、目录录入、图像扫描、自动质检、人工质检、成品提交六大工序，各环节环环相扣，我们利用 workflow 技术贯穿整个过程，实现了数字化加工过程的有序控制，使数据流、信息流、业务 workflow 在内部合理快速流动，不仅能够有效地组织档案加工处理过程，也可以大大提高数字化档案加工的速度和效率。

2、数字化加工软件 workflow 技术应用

(1) workflow 管理系统应用。

在数字化加工软件开发过程中，课题组引入 workflow 管理系统，重点关注 workflow 引擎的灵活性、基于 workflow 引擎的流水线配置的

方便性、与加工工具集成的便利性以及工作流的监控等。研究内容包括：工作流引擎、工作流管理集成机制、建模工具、协同工作机制、流程设计器和流程监控工具等。主要功能是通过计算机技术的支持去定义、执行和管理工作流，协调工作流执行过程之间以及群体成员之间的信息交互。依靠工作流管理系统来实现真正的“人-机”交互系统，用户是系统中的基本角色，是直接的任务分派对象，用户可以直接看到电脑针对自己列出的“任务清单”，跟踪每一项任务的状态，或继续一项任务，而不必从一个模块退出，进入另一个模块，搜索相应任务的线索。在工作流管理系统中，用户的任务分派和任务的完成状态，可以被最大程度地电脑化和受到控制。

通过工作流引擎，管理员可灵活设置双机摘录、条目质检、人机交互、任务派发等流程节点，通过电脑和人工多次质检确保数字化加工的质量，克服了传统模式的弊端，减少人为差错和延误，提高了劳动生产率，缩短数字化加工运营周期，提升了档案行业的标准化加工流程，实现了多人协同工作，优化并合理利用了资源。

加工软件工作流程图：

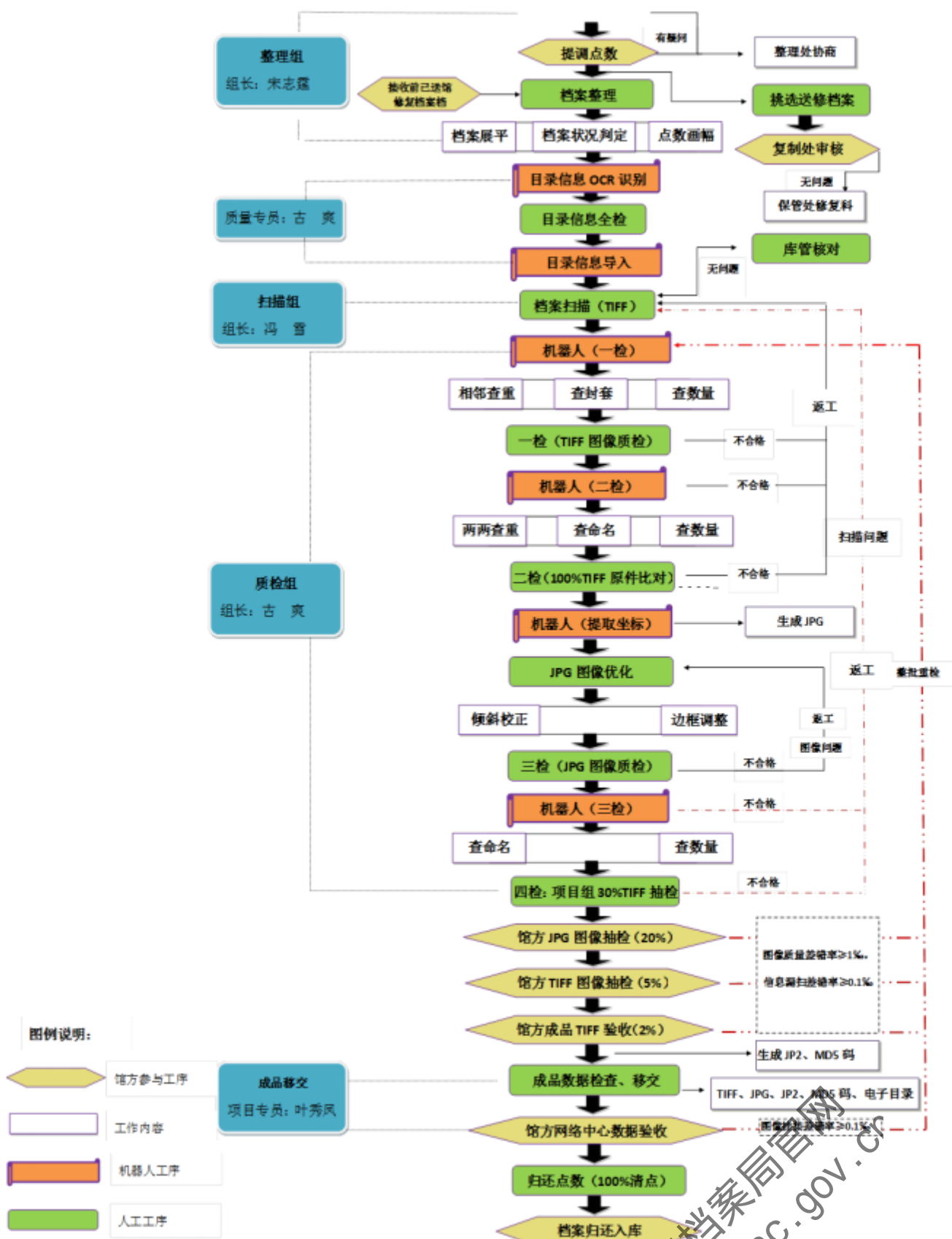


图 8-加工软件工作流程

(2) workflow 中的数据管理。

在数字化加工软件中，workflow 涉及到的数据包括流程管理数据和加工数据。流程管理数据包括 workflow 的实例流水线数据、工序数据、任务数据、参与者数据等，其控制着 workflow 的流转并记录 workflow 的实际运行过程，为了便于查询、分析、统计，我们将流程管理数据存储到数据库中；加工数据是指由加工工具操作的数据，反映的是具体业务的实际内容信息，它是在整个流程中进行流动，为了便于操作和修改，加工数据以文件形式集中存储于服务器上，并且每个工序的输出文件都在服务器上单独存放。当进行数字化加工时，从服务器获取待加工数据，在经过本地加工工具处理后，再将加工结果数据上传到服务器上集中存放。

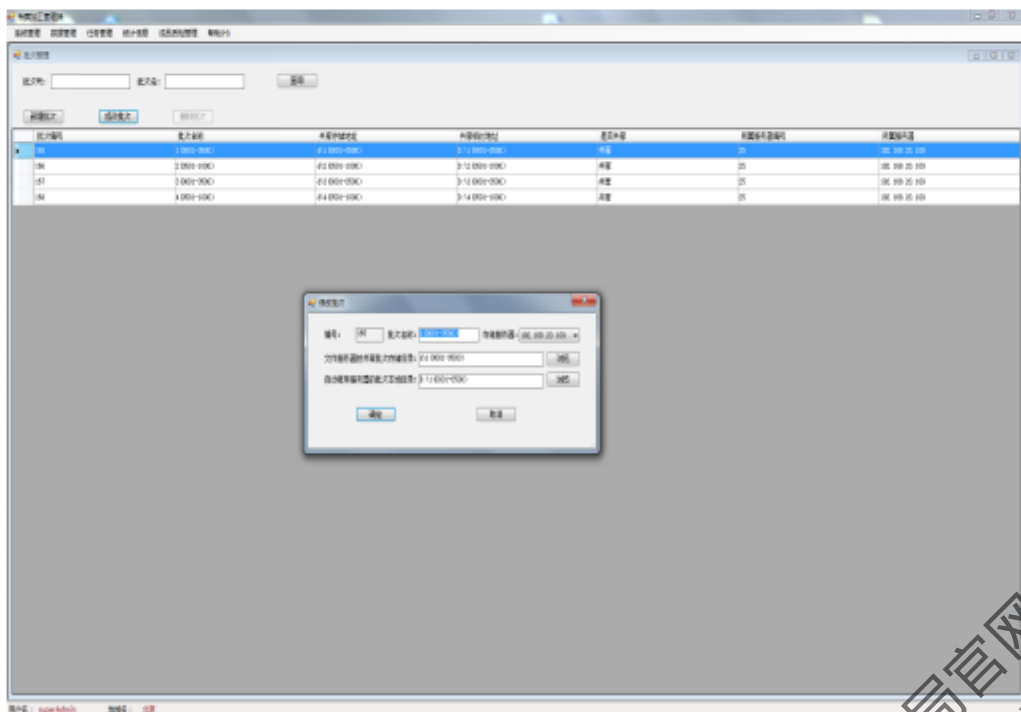


图9-批次数据管理

(3) workflow 与加工工具集成。

数字化过程中常常需要多个加工工具配合来完成加工任务，加工工具的种类繁多、来源不同，如何将不同的加工工具集成到 workflows 中，并保持加工工具与 workflows 的松散耦合，是 workflow 技术应用研究的一项重要内容。

为了保证 workflow 对工具的集成具有足够的开放性。我们通过将工具映射到 workflow 工序节点的方式进行集成，工序与工具的映射是自由的，工序可以与任意一个工具进行映射，但一个工序只能与一个工具进行映射。同时为了使工具能够正确使用 workflow 中的加工数据，并且在加工完成时能将加工完成后的结果、状态等信息反馈给 workflow，我们定义了标准数据接口协议，该协议主要包括启动协议文件和加工完成协议文件。启动协议文件中记录数据的本地组织方式、输入输出类型、加工类型等，加工完成协议文件中记录加工的工作量和加工完成标志等。

workflow 接到加工任务，根据任务的工序配置，下载待加工数据到本地，然后生成启动协议文件，根据映射查询工具并启动对应的工具，工具根据启动协议文件加载数据；之后当工具完成加工后，加工工具生成加工完成协议文件，workflow 根据加工完成协议文件上传加工数据和工作量等信息。这样，所有符合标准数据接口协议的工具有都可以接入 workflow，我们既可以集成人机交互工具，也可以集成自动处理工具；既可以集成加工制作工具，也可以集成全检、抽检、返工等工具。加工工具与 workflow 的集成为数

字化加工的流水线作业提供了重要保证。

(4) 加工工具配置文件的灵活配置。

为了满足不同的加工质量、加工内容等需求，对大多数的加工工具进行了参数化设计，并将参数值记录在配置文件中，在使用时，加工工具可以根据配置文件中的不同的参数值实现不同的加工效果。这样的设计给工艺制定人员带来了极大的方便性。加工人员可以根据项目需求，很方便的对工具进行参数化配置，从而满足项目目标。例如扫描工具可以根据不同的 dpi 参数扫描出不同质量的图像，文字工具根据不同的错误率判断是否合格等。

加工工具的参数化设计中，工具和配置文件是分离的，是一对多的关系，为了在加工流水线中实现工具与配置文件分离，我们引入工序的概念，我们将流水线中的需要处理数据的地方称为一个工序，一个工序对应一个工具和一个配置文件集，而一个工具可以对应多个工序。也就是说一个工具和一个配置文件集组成了一个具体的工序。若干个工序组合形成流水线。当不同的工序使用同一个工具而使用不同的配置文件集时，加工本地只有一份工具，在启动工具前，工作流会根据工序的不同为工具下载对应的配置文件集并放入合适的位置。这样的设计实现了工具与配置文件分离，加工人员可以根据项目需求为工具灵活配置不同的配置文件。

(5) 加工工具的版本管理。

在数字化加工中的加工工具并非是一成不变的，会根据实际的生产需求进行功能扩充和性能优化，加工工具的研发是一个不断迭代的过程。在这个过程中会不断有新的版本发布到生产中，这样就会产生多个版本，经常会造成版本使用混乱。如何解决加工工具的版本混乱问题是 workflow 管理系统中的重要一环。

为了实现加工工具的版本控制，我们通过版本检测和强制升级的方式来实现。当加工工具有新的版本时，将新版本加工工具上传服务器并对版本进行登记。当需要打开加工工具时，我们将加工工具的本地版本与服务器版本进行比对检测，当有新的版本时，强制对新版本加工工具进行下载并更新，更新完成后才可以启动加工工具，这样便避免了因版本升级导致加工人员所用加工工具不一致的现象。

(6) 自动处理工具的调度。

在数字化加工过程中，我们将耗时的可自动化处理的功能独立出来，形成自动处理工具，自动处理工具不受人工干扰，可以高效稳定的执行，是提高加工效率和质量的一个重要手段。常见的自动处理工具有自动查重工具，自动图像优化工具，自动切分识别工具等。

自动处理工具不会自动运行，需要一个管理自动处理任务的工具来调度。该管理工具实时检测是否有自动处理加工任务，当有加工任务时会先下载加工数据，然后调用自动处理工具完成处

理，最后将自动处理结果上传。该管理工具实现了对自动处理任务的全自动化加工，极大地提高了生产质量和效率。

界面示例如下：

任务ID	名称	启动状态
5	自动提取图像坐标	已启动
7	自动处理图像	已启动
9	扫描自动转70°(一级)	已启动
10	扫描近工合并(一级)	已启动
11	自动图像重直(一级)	已启动
14	扫描近工合并(二级)	已启动
17	扫描自动转70°(二级)	已启动
18	自动图像重直(二级)	已启动
19	自动图像重直(三级)	已启动
22	扫描近工合并(三级)	已启动
23	扫描自动转70°(三级)	已启动

任务ID	工序名	包名	开始时间	结束时间	任务状态	失败原因
1361626	自动图像重直(三级)	05-02-001-003071-0001	2014-10-10 01:48:55	2014-10-10 01:49:53	已完成	
1362293	自动图像重直(三级)	05-02-001-003071-0001	2014-10-10 01:49:51	2014-10-10 01:50:48	已完成	
1362295	自动图像重直(三级)	05-02-001-003046-0023	2014-10-10 01:50:48	2014-10-10 01:51:02	已完成	
1361641	自动图像重直(三级)	05-02-001-003071-0016	2014-10-10 01:51:00	2014-10-10 01:51:13	已完成	
1362345	自动图像重直(三级)	05-02-001-003071-0016	2014-10-10 01:51:13	2014-10-10 01:51:26	已完成	
1361596	自动图像重直(三级)	05-02-001-003071-0031	2014-10-10 01:51:26	2014-10-10 01:51:48	已完成	
1362352	自动图像重直(三级)	05-02-001-003071-0040	2014-10-10 01:51:48	2014-10-10 01:52:23	已完成	
1361582	自动图像重直(三级)	05-02-001-003071-0045	2014-10-10 01:52:23	2014-10-10 01:52:45	已完成	
1362395	自动图像重直(三级)	05-02-001-003071-0045	2014-10-10 01:52:45	2014-10-10 01:53:25	已完成	
1362369	扫描自动转70°(一级)	05-02-001-003014-0050	2014-10-10 01:53:25	2014-10-10 01:53:47	已完成	
1362413	自动图像重直(一级)	05-02-001-003014-0050	2014-10-10 01:53:47	2014-10-10 01:54:14	已完成	
1362422	自动图像重直(一级)	05-02-001-003071-0058	2014-10-10 01:59:14	2014-10-10 01:59:39	已完成	
1362763	自动处理图像	05-02-001-003029-0050	2014-10-10 02:04:38	2014-10-10 02:05:10	已完成	
1362780	自动处理图像	05-02-001-003029-0067	2014-10-10 02:05:10	2014-10-10 02:05:28	已完成	
1362799	自动处理图像	05-02-001-003029-0084	2014-10-10 02:05:28	2014-10-10 02:05:37	已完成	
1362815	自动处理图像	05-02-001-003029-0016	2014-10-10 02:05:37	2014-10-10 02:05:46	已完成	
1362834	自动处理图像	05-02-001-003029-0032	2014-10-10 02:05:46	2014-10-10 02:06:02	已完成	
1362837	自动处理图像	05-02-001-003029-0035	2014-10-10 02:06:02	2014-10-10 02:06:21	已完成	
1362853	自动处理图像	05-02-001-003030-0002	2014-10-10 02:06:21	2014-10-10 02:10:10	已完成	
1362888	自动处理图像坐标	05-02-001-003044-0004	2014-10-10 02:10:11	0001-01-01 12:00:00	进行中	

图 10-自动处理工具界面

(7) 工作流程中数据流转优先级

在数字化加工中，常常会遇到需要紧急加工某个工序、某包数据或某个任务的情况，我们通过优先级管理策略来解决这个问题，首先我们对工序、包、任务分别设置优先级，在任务调度的时候，我们的获取任务策略是首先考虑工序优先级，然后包优先级，最后任务优先级。即优先级高的工序先获取，然后获取处于该工序状态的具有高优先级的包，最后确定相同包优先级下具有高优先级的任务。通过优先级管理，管理员可以随时调整工序、包、任务的优先级，满足生产需求。

(8) 工作流程流程监控

workflows. The process management data is stored in the database, recording participants, procedures, operation types, and participation times. We can monitor the process based on this information. Through progress monitoring, we can understand the processing progress in time, analyze task bottlenecks, and provide a reference for合理安排生产、调整和优化加工过程. Through monitoring the pass rate, we can discover abnormal pass rates in time, analyze and summarize the reasons in time, and carry out training in time to improve the pass rate. Task status diagram example:

工序名称	员工姓名	工作包	数量	任务状态	分配员工	分配时间	任务开始时间	任务结束时间
录入资料	superAdmin	04-15-011-000208-0008	4	已完成	系统	2016/5/10 17:00	2016/5/10 17:00	2016/5/10 17:00
扫描图像	张恒	04-15-011-000208-0008	4	已完成	superAdmin	2016/5/7 17:04	2016/5/8 8:17	2016/5/8 9:17
扫描自动转JPG(一级)	jgr105	04-15-011-000208-0008	4	已完成	系统	2016/5/8 9:17	2016/5/8 9:17	2016/5/8 9:17
自动图像直量(一级)	jgr105	04-15-011-000208-0008	4	已完成	系统	2016/5/8 9:17	2016/5/8 9:17	2016/5/8 9:18
扫描质检(一级)	胡文月	04-15-011-000208-0008	4	已完成	古爽	2016/5/8 14:14	2016/5/8 14:33	2016/5/12 8:51
扫描质检(二级)	韩悦	04-15-011-000208-0008	4	已完成	古爽	2016/5/12 13:01	2016/5/12 13:30	2016/5/12 13:53
自动图像直量(三级)	jgr121	04-15-011-000208-0008	4	已完成	系统	2016/5/12 13:53	2016/5/12 13:54	2016/5/12 13:55
扫描质检(三级)	superAdmin	04-15-011-000208-0008	4	已完成	系统	2016/5/12 13:55	2016/5/12 13:55	2016/5/12 13:55
自动提取图像坐标	jgr121	04-15-011-000208-0008	4	已完成	系统	2016/5/12 13:55	2016/5/12 13:55	2016/5/12 13:55
检查图像坐标	即敏强	04-15-011-000208-0008	4	已完成	古爽	2016/5/12 15:40	2016/5/12 15:44	2016/5/12 17:13
自动处理图像	jgr120	04-15-011-000208-0008	4	已完成	系统	2016/5/12 17:13	2016/5/12 17:13	2016/5/12 17:13
图像处理质检	郭冉兰	04-15-011-000208-0008	4	已完成	古爽	2016/5/13 8:28	2016/5/13 8:37	2016/5/13 9:51

图11-任务分配

5.2.2 数字化前处理环节质量管理

扫描前处理主要工作内容包括档案点数，记录档案实体状况并生成数字化加工目录，对档案进行除霉、展平及送修处理。一是确保档案交接档案数量准确无误，包括提档点数及还档预点数；

二是点数并记录档案画幅数量、档案基本状况，为数字扫描、质检提供电子目录。该环节有专门前处理工作记录单，用以记录档案画幅数、档案实际状况和修复信息，加工公司前整理记录数据（扫描修复时间、尺数、人员、时间，扫描展平时间、尺数、人员，扫描除霉时间、人员）等，作为数字化加工目录元数据保存。

1. 质量管理举措——对实体档案进行扫描前处理。

明清档案形制复杂，实体情况差异悬殊，部分档案无法满足直接数字化扫描的技术要求。为保证数字化加工图像质量，课题组组织修复专业技术人员对业务人员进行了档案拆装、修复、展平等多项专业技术培训，并提供除尘桌和软毛刷，撤潮纸和去筋面粉浆糊等专门用具，由专人实时指导和监管，在档案正式扫描前对实体进行去褶、除霉、修复、展平等多项处理工作，确保档案实体符合数字图像扫描标准，为数字图像扫描加工打下基础。

如图：



图 12-残破档案示例



图 13-扫描前处理—展平效果展示

2. 引入表格登录识别技术提高目录录入质量。

课题组根据质检规范制定了档案状况判定基本标准，制作了目录录入表格（即前整理工作单），并将表格定义为表格登陆识别模板，前整理人员依据标准在表格内详细记录档案状况和相关信息，通过目录录入形成档案数字化加工目录元数据，并上传至数字化加工系统，为下一阶段的数字图像加工提供扫描和质检依据。

表格登录识别系统由两部分组成，包括表格登录模块、表格识别模块。完成的功能为：将扫描（单张扫描或批量扫描）得到的图像或已有的图像文件，通过识别找到对应的已登录的模板，并按照模板的设置进行识别，并将最终识别结果保存在数据库中。课题组通过创建表格模板，将表格图片分解成多个不同类型的元素，逐个对元素进行识别，分别将数据提取出来，以指定格式保存。本方法可以轻松实现表格的智能录入，提高效率，提高质量。在制作表格模板时，提取表格拓扑结构特征，并提取表格标题信

息。

技术原理：

(1) 表格登录模块

a 模板登录：即往模板数据库中增加模板。选择一个空的表格图像，给该模板命名，并自动分配一个 ID 号。自动进行倾斜矫正、自动提取表格的拓扑结构线段。将图像名称、模板名称、ID 号以及表格线段保存在模板数据库中。

b 模板删除。从模板数据库中删除一个或多个模板。模板在删除时，必须保证所有的模板文件没有引用该模板，否则，删除失败。

c 模板文件登录：建立一个新的模板文件，给模板文件命名，并初始化模板文件为空。所有模板文件放在一个特定的目录下。自动调模板文件的编辑界面。

d 模板文件的删除。删除一个模板文件必须保证所有的工作没有与此模板文件关联。否则，删除失败。

e 模板文件的编辑。模板文件的编辑功能包括增加、修改、删除、复制、粘贴功能。

f 模板文件的保存。当模板文件的编辑得到确认后才能真正修改模板文件。

(2) 表格识别模块

a 生成工作文件：在工作文件目录下生成一个工作文件，给该

工作文件命名，指定一个模板文件，并指定该工作文件将包含的图像的档案号，初始化工作状态。工作文件的描述包括该工作对应的模板文件、工作状态、工作包含的图像。图像可以是扫描得到或已有的图像文件。

b 浏览当前的工作及其模板文件。可以浏览该工作的所有图像及识别结果，可以浏览该工作对应的模板文件内的所有模板图像及识别设定。

c 工作的修改。可以增加或删除工作所包含的图像文件；删除工作中的图像时，该图像的所有临时文件（识别过程中生成的）也一起删除；当需要更改工作所关联的模板文件时，工作中所作的识别全部取消，所有临时文件删除，工作置为初试状态；可以修改工作的档案号。

d 图像操作：包括自动倾斜矫正、手工倾斜矫正。

e 表格类型识别。自动提取待识别表格图像的拓扑结构，然后与模板库中的模板进行比对，选择一个可信度最大且大于一定阈值的一个模板作为表格类型结果。

f 表格识别。按照模板的设置进行自动字符识别。具体实现是，将二值图像和该类型的模板（包含偏移量）作为参数传进来，识别核心将识别结果保存在临时文件中。

h 识别后的结果按照模板的设置，逐条显示在识别结果区域内。可以对其进行校对。

i 识别结果的入库。对全部处理完的工作，可以将结果保存在数据库中。数据库的结构根据模板的定义自动生成。

(3) 基于深度学习的 OCR 技术

深度学习是机器学习研究中的一个新的领域，其动机在于建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像，声音和文本。深度学习的概念源于人工神经网络的研究。含多隐层的多层感知器就是一种深度学习结构。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示。

卷积神经网络（CNN）是一种特殊的深层的神经网络模型，它的特殊性体现在两个方面，一方面它的神经元间的连接是非全连接的，另一方面同一层中某些神经元之间的连接的权重是共享的。它的非全连接和权值共享的网络结构使之更类似于生物神经网络，降低了网络模型的复杂度，减少了权值的数量。卷积神经网络常用于图像相关处理，文字识别是其一个典型应用。

本系统的 OCR 识别，采用 LeNet-5 模型（见图 14），LeNet-5 共有 7 层，不包含输入，每层都包含可训练参数（连接权重）。

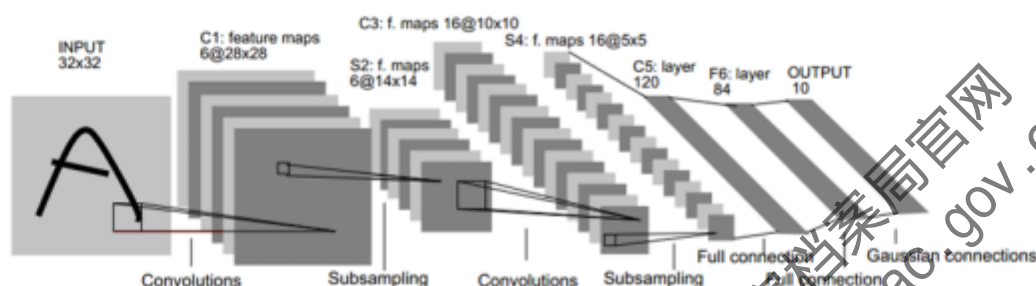


图 14—LeNet-5 模型

在本系统中，采用 OCR 技术，将经过表格定位提取的档案信息字符识别，用于数字化系统的质量控制。在本系统中，档案信息包括档案数目和档案状况。其中，各种档案状况可以枚举出来，用序号表示，如：“1”表示档案状况良好，“2”表示档案破损，“3”表示档案水浸……因此，本系统中的 OCR 解决的是手写体数字识别。

我们利用汉王公司在各个实际项目中采集得到的 100 万手写体数字和印刷体数字混合集，其中 90 万为训练集，10 万为测试集。经过训练，在测试集上得到 98.90% 的正确率。在系统实际使用中，由于规范操作的书写，得到了 99.99% 的正确率。

应用方法：

首先对纸质的元数据信息单进行扫描，然后通过元数据识别工具对元数据纸质表单进行扫描、OCR 识别，最后通过合并工具将单页识别的元数据信息进行合并，形成元数据 Excel 文档，人工核对无误后，再将电子版元数据信息通过档案导入工具批量的导入到加工系统中。

(1) 整理备注：为整理组人员手写内容，如：断扣一起扫等直接以数字代替，在目录录入时直接转换成相应文字内容。然后利用基于深度学习的 OCR 技术进行识别。

识别过程如下图：

公司：		前处理 档案整理记录单																
项目	C9006-5	卷号	003114	盒数	1	面幅		整理员	王敬然	整理时间	2014. 8. 4							
件号	整理数量					整理状况										卷号	整理备注	
	文种	正面	背面	浮帖	夹条	不符	破裂	碎片	粘白	粘字	小开	首粘	尾粘	断扣	不明	折痕		簿册
1		3	1															
2		5																2
3		5																
4		2					✓											
5		2																
6		5			1													
7		2																
8		5					✓											
9		4																
10		4																
11		5																
12		7																

图 15-前整理记录单手写信息

```

C:\Windows\system32\cmd.exe
result is "11::2::00000000000000"
result is "12::2::00000000000000"
result is "13::2::00000000000000"
result is "14::2::00000000000000"
result is "15::2::00000000000000"
result is "16::2::00000000000000"
result is "17::2::00000000000000"
result is "18::2::00000000000000"
result is "19::2::00000000000000"
result is "20::2::00000000000000"
result is "21::2::00000000000000"
result is "22::2::00000000000000"
result is "23::2::00000000000000"
result is "24::2::00000000000000"
result is "25::2::00000000000000"
result is "26::2::00000000000000"
result is "27::2::00000000000000"
result is "28::2::00000000000000"
result is "29::2::00000000000000"
result is "30::2::00000000000000"
result is "31::2::00000000000000"
result is "32::2::00000000000000"
result is "33::2::00000000000000"
result is "34::2::00000000000000"

```

图 16-数据自动识别过程

(2) 识别结果质检、导入示例：

编号	文种面幅数	正面面幅数	反面面幅数	浮贴面幅数	夹条面幅数	其他面幅数	备注	整理状况
05-02-001-003114-0001		3	1					
05-02-001-003114-0002		5					有浮贴、	有浮贴、
05-02-001-003114-0003		5						
05-02-001-003114-0004		2						破裂、
05-02-001-003114-0005		2						
05-02-001-003114-0006		5			1			
05-02-001-003114-0007		2						
05-02-001-003114-0008		5						破裂、
05-02-001-003114-0009		4						
05-02-001-003114-0010		4						
05-02-001-003114-0011		5						
05-02-001-003114-0012		7						
05-02-001-003114-0013		8					断扣一起扫、	漏粘、断扣、断扣一起扫、、
05-02-001-003114-0014		3						断扣、
05-02-001-003114-0015		5		1				
05-02-001-003114-0016		16						
05-02-001-003114-0017		5						断扣、
05-02-001-003114-0018		5						
05-02-001-003114-0019		6						
05-02-001-003114-0020		3						
05-02-001-003114-0021		3	1					
05-02-001-003114-0022							超厚簿册不处理、	超厚簿册不处理、

图 17-数据自动识别结果

导入之后为：展平、应修已扫、未修未扫、已修已扫、超厚簿册不处理、超 A2 不处理；断扣一起扫、有浮贴、有夹条、夹条原位置无法确定、内容粘连无法打开、超大画幅、边距不足、背面无信息未扫描、无信息页未扫描、无信息页不扫描、文字方向不一致、不拍不录、信函不破拆，不扫描。

加工软件显示结果如图：

类型	内容
编号	05-02-001-003124-0095
档案类型	
责任者	广德御银库
文种	密文
文字种类	汉
残缺现象	残缺、破损
残缺程度	轻度
待修	
面数	5
时间	
客户备注	
总幅数	3
档案状况	原件残缺
文种面幅数	0
正面	3
背面	0
浮贴揭开	0
独立小条	0
其它面幅	0
备注	一起扫
整理状况	断扣、破裂
加工提示	顺序及命名、压字、露字、折角、文字完整性

此框为元素据 OCR 自动识别后，再结合人工目录质检合格后，导入系统，供后续工序依据，如图显示内容。

国家档案局官网
WWW.SAAC.GOV.CN

图 18-加工软件目录显示

手写记录单转换为加工目录整体过程展示：

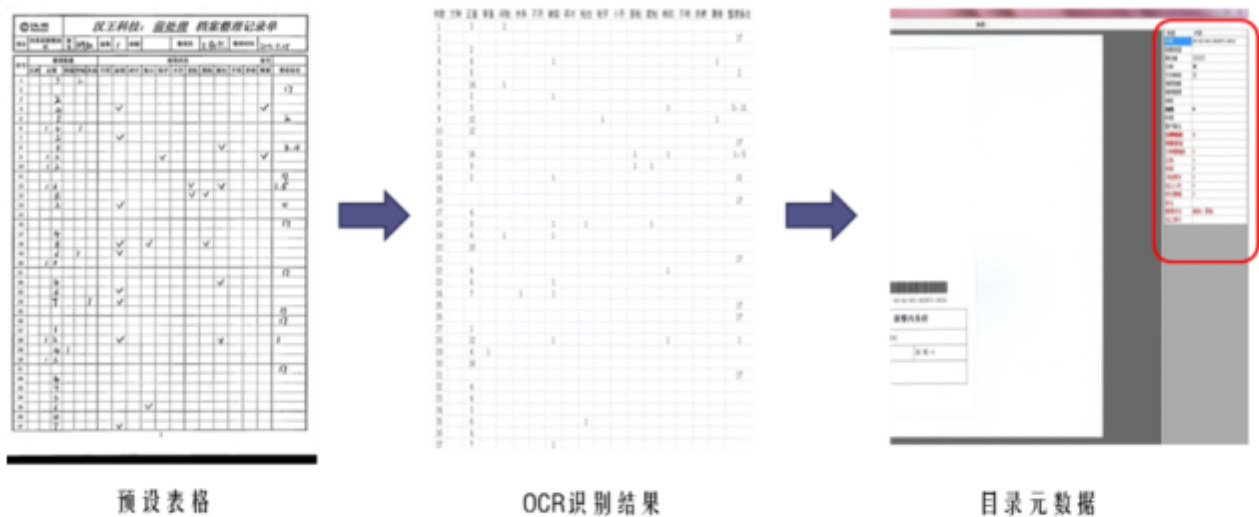


图 19-整体过程展示

应用效果：

基于一史馆的数字化加工项目，我们前期进行了约 35 万条数据测试，后期进行了 8 万条数据检验。目前，在配备快速扫描设备的情况下，表格登陆识别技术能够录入数据 30 条/分钟，准确率 99.9%。通过该技术应用，比人工数据录入增加 26 条/分钟，准确率提高了约 1%，为海量目录数据高效录入提供了质量保证。

5.2.3 数字图像扫描环节质量管理

1. 数字图像扫描质量控制

数字图像扫描人员从数字化加工系统中接收任务，依据前整理提供的数字化加工目录信息以及馆方制定的数字化加工规范，对档案按既定要求顺序扫描，形成原始 TIFF 格式图像，并对图像进行命名核对。加工人员以卷为单位从库管领取档案，加工完成后统一归还。在此过程中，其只能通过工作流程接收任务以及进行

数据提交，无法对提交数据进行修改。同时，为确保档案加工质量，现场指定专人定期对数字化加工设备进行巡查，并要求加工人员在每卷扫描前进行色彩校正，对前整理提供目录信息进行验证，发现问题报现场管理人员统一修订，确保数字化加工信息前后一致。

由于档案原件装于封套中，档案本身没有任何页码或其他标记，在本环节的项目管理中，课题组通过规定加工人员行为规范，实行“一件一扫”，首先确保档案图像与封套、目录对应一致性。其次，设定加工人员工作系统权限，加工人员在扫描过程中无法对目录标记信息进行更改，当加工人员发现问题时只能通过专人进行修改，借此对档案扫描人员和目录信息进行二次核查，确保扫描准确性；第三，通过档案流转单，提前预知本卷档案存在的特殊情况或难点，结合工作流系统目录信息，能够有效避免扫描问题出现，为数字化加工质量提供了保证。

2. 基于命令流的图像处理技术应用

在本环节中，课题组尝试引进自动图像处理工具，对加工人扫描上传档案图像进行命名核对、数量检查及封套检查。经图像处理算法计算得到图像倾斜角度、版心区域等数据，对档案图像进行纠偏、自动裁切处理等。但由于档案图像尺寸较大，分辨率较高，原始图像文件大小在50-60M以上，应用级图像文件大小也在5M以上。有时候一件档案包含几百幅图像，需要流转的图像包

大小在 1G 以上，在图像处理的各道工序之间流转，严重影响流转效率，为此必须要解决效率问题，自动图像处理工具才能够真正发挥作用。

通过分析，课题组发现在人工干预的各个环节，即使在图像质量大幅降低的情况下，也不影响操作人员正确判断图像情况，做出正确的操作。因此，可以在各个操作人员环节，流转低质量图像，并将操作员的操作记录为操作命令流传回服务器，在原始高质量图像上自动实现这些命令，保证操作后的图像质量。

技术原理：

课题组设计的基于命令流的图像处理系统包括管理端和工具端两大部分。管理端负责图像数据包的上传下载、分发流转等调度管理工作；工具端由 5 个图像处理工具组成，分别为大图转小图、自动图像处理、人工图像处理、图像处理质检以及成品图像生成工具。

命令流是由一条条的命令组成的一个命令组合，这些命令按照一定的顺序排布，能够完成一定的图像处理功能。

这种基于命令流的图像处理系统，对图像数据包的处理包括以下步骤：

a 通过管理端的图像导入功能将扫描图像上传至图像处理系统的服务器，形成数据包，等待进行图像处理。

b 管理端将图像数据包下载至大图转小图工具所在 PC 机，启

动大图转小图工具，将数据包中的每页图像缩小为原始扫描图像 1/16 的小图像，形成小图数据包，管理端将小图数据包上传至服务器进行管理。

c 管理端将小图数据包下载至自动图像处理工具所在 PC 机，启动自动图像处理工具，经图像处理算法计算得到图像倾斜角度、版心区域等数据，记录至命令流文件，形成命令流数据包，管理端将命令流数据包上传至服务器进行管理。

d 管理端将小图数据包和命令流数据包一并下载至人工图像处理工具所在 PC 机，操作人员可通过人工图像处理工具预览命令流作用于小图图像呈现的图像处理效果，也可修改图像处理效果，其修改保存至命令流文件，操作人员确认完成人工图像处理，管理端将新的命令流数据包上传至服务器进行管理。

e 管理端将小图数据包和步骤 4 中产生的命令流数据包一并下载至图像处理质检工具所在 PC 机，与人工图像处理工具相同，操作人员可检查并修改图像处理效果，其修改保存至命令流文件，管理端将新的命令流数据包上传至服务器进行管理。

f 管理端将步骤 1 中形成的原始扫描大图数据包和步骤 5 中产生的命令流数据包一并下载至成品图像生成工具所在 PC 机，启动成品图像生成工具，该工具会根据命令流对扫描图像进行图像处理，产生最终的成品图像，由管理端上传至服务器进行管理。

这种基于命令流的图像处理系统的优点有：

a 图像处理操作以命令流的格式保存至文件，在服务器和 PC 端之间上传下载的数据为命令流文件，而不是经图像处理后的结果图像文件。与图像数据相比，命令流文件数据量小，便于快速安全的进行网络传输。

b 图像处理系统具有质检功能设置，并且质检工具检查和修改的是命令流数据，相比于直接修改图像数据，所做操作便于撤销和重做。

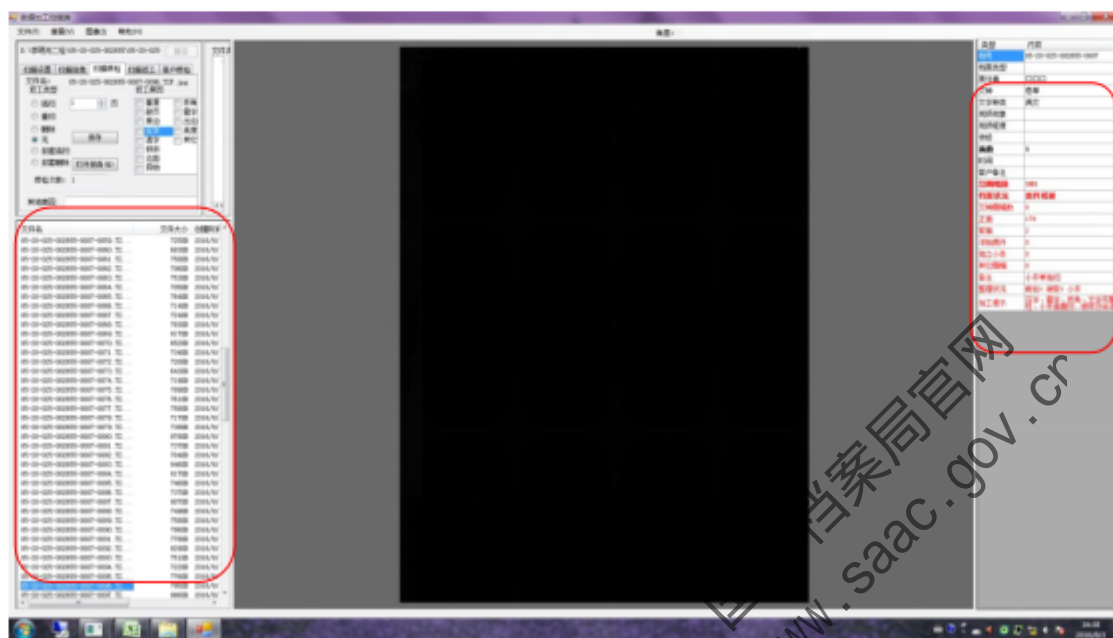
c 引入大图转小图处理，将原始的扫描图像先缩小为 1/16 的小图像，后续的图像处理工具只需从服务器下载小图像，基于小图像完成图像处理操作，大大提高了图像数据传输的速度。

应用界面展示：

a 数据检查工具

对加工过程中的数据进行检查，验证图像的名称、数量等信息。并将与扫描相关的信息进行提示，重点信息以红色提示，并根据前整理信息给出扫描注意提示，避免扫描错误。

图 20-1 数据信息提示



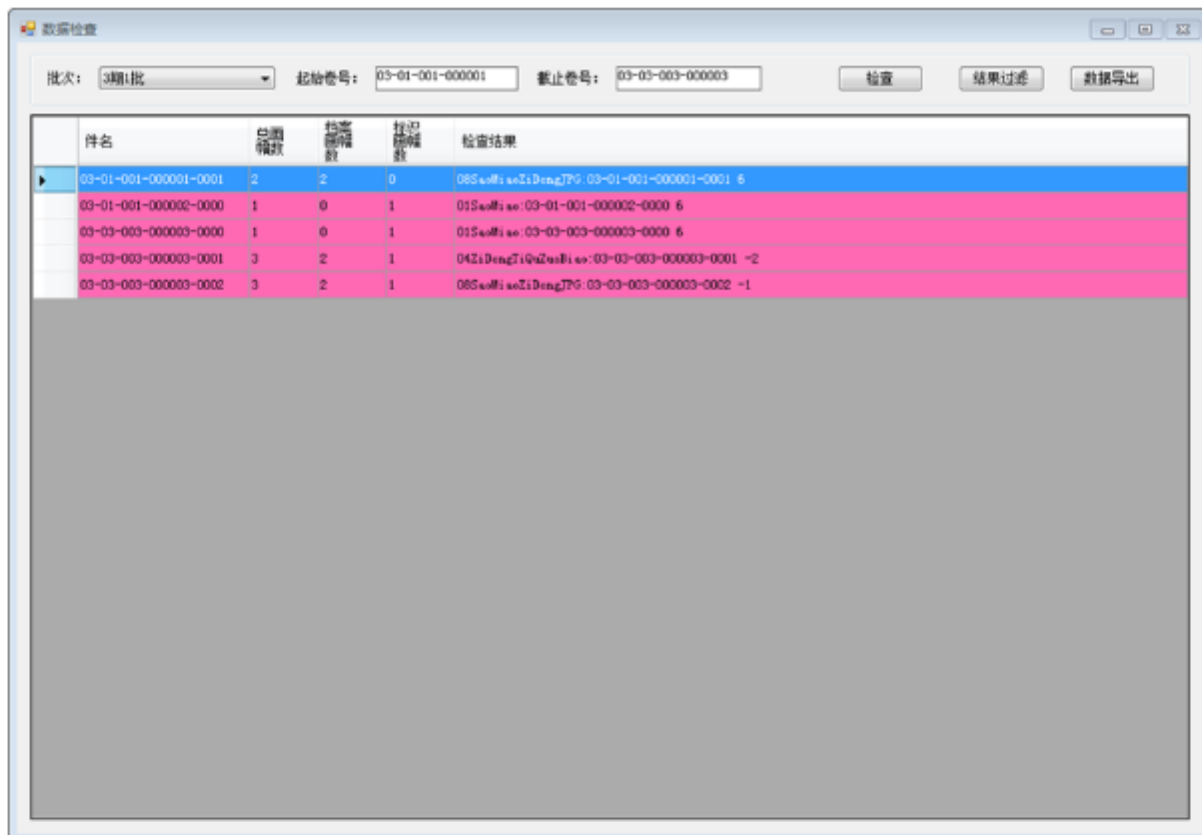


图 21-图像数量检查

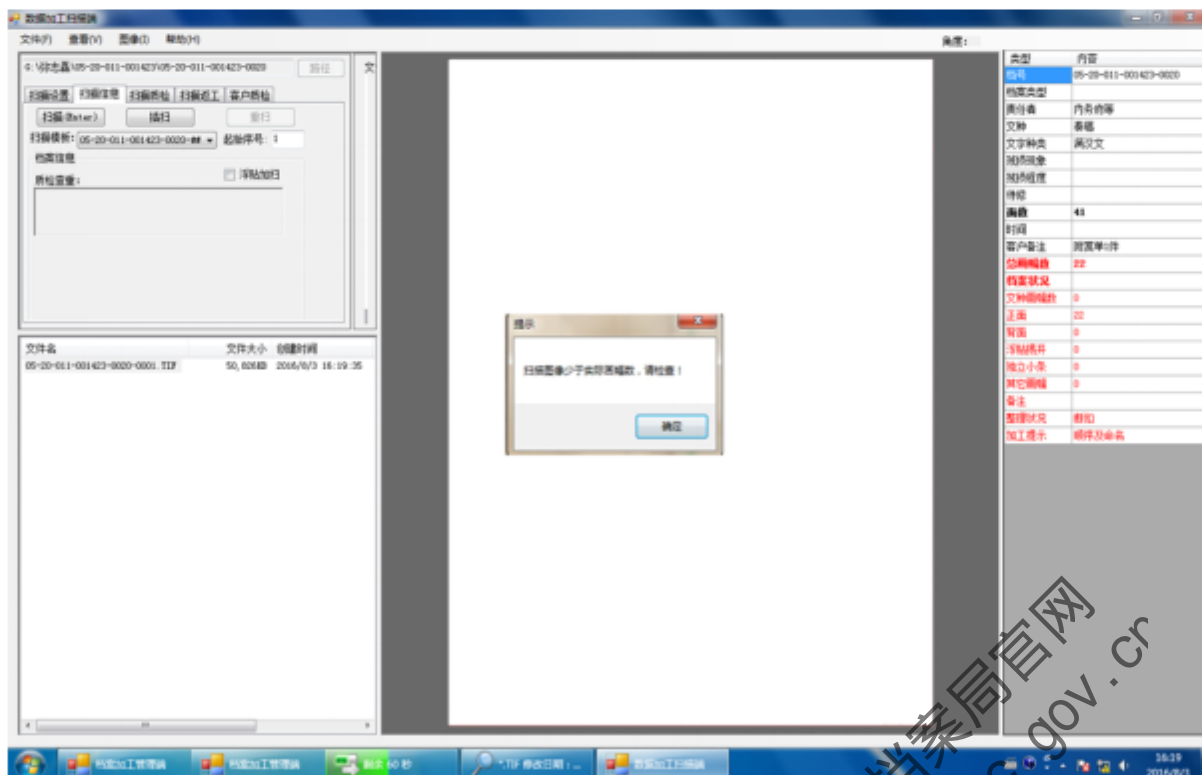


图 22-检查结果提示

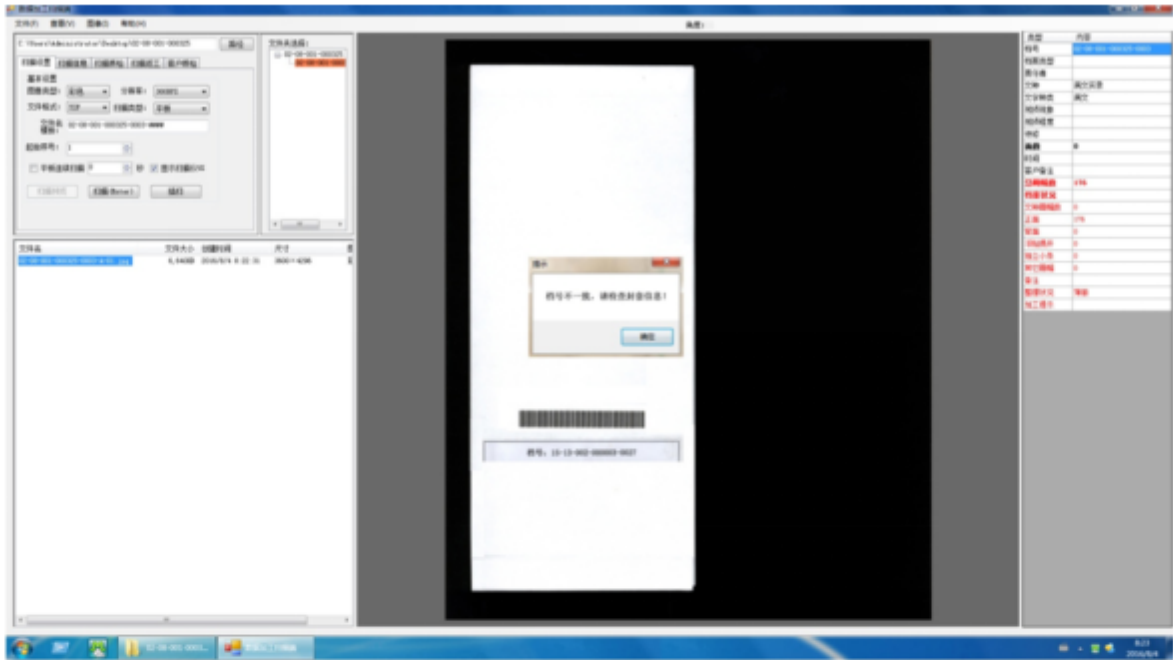


图 23-封套检查-条形码识别

b 图像自动处理工具

自动图像处理工具通过图像处理算法，计算得到图像倾斜角度、版心区域等数据，操作人员可通过人工图像处理工具预览命令流作用于小图图像呈现的图像处理效果，也可修改图像处理效果，操作人员确认完成人工图像处理 after，再将新的数据包上传至服务器进行管理。

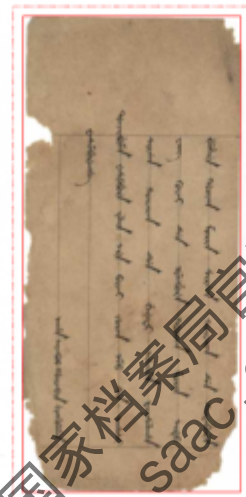
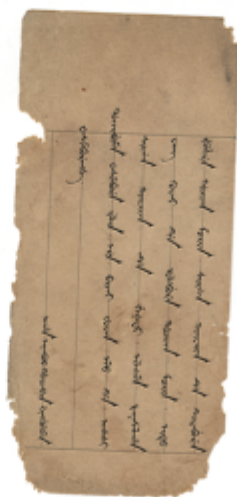


图 24-图像自动纠偏

该技术通过对成品画幅数自动核查、图像命名核对、封套图像核对等功能，大大提高了人工检查效率和准确率，同时通过图像自动处理，实现格式自动转换、图像自动纠偏、裁切等功能，很大程度上避免了人工操作造成的失误，提高了质量控制能力。

5.2.4 图像质量检查环节质量管理

作为明清档案数字化图像加工的最后一环，质量检查工作起着至关重要的作用，是项目的生命线。明清档案的珍贵性和不可再生性要求我们在对待档案原件时慎之又慎。档案原件从库房到加工场地，提调涉及部门较多，环节交接登记详细，加工一次极为不易，这就要求我们数字化加工必须一次完成，一次达标，减少重复提调和重复加工。

1. 数据的质量检查与验收

(1) 严格公司三道质检工作流程。

为确保图像质量，馆方严格监督外包公司方三道质检程序执行以及公司管理层 30%原档抽查。在日常现场安全巡查和监管中，时刻关注外包公司质检流程是否按要求执行，不定期检查其质检工作记录等。公司三道质检通过计算机技术与人工核查相结合，各环节环环相扣，各有侧重，通过后一环节对前面工作环节的检查，最终获得质量合格的成品。

(2) 严格按照验收流程对公司提交数据进行验收。

数字化加工外包为流水化生产，日均加工量可达 8 万画幅左右，

数据量较大。为及时发现并解决问题，项目监管组抽检和项目组验收采取的是数据批次提交验收方式，以每周为1批次，根据公司规模、监管人员数量，按照其加工能力，设定批次数据提交验收总量，约定数据提交时间，严格按照质量标准进行检查，做好批次检查记录，留存问题图像及返工后图像，出具批次抽检报告和验收报告，作为项目整体验收依据。项目整体验收时，项目组按照验收办法认真组织，准备项目过程文档和成品数据，确保验收程序科学严谨。通过此种方式，一方面能够有效降低风险成本，另一方面可促进数字化外包管理工作例行化，保证数字化加工外包项目整体平稳运行。

(3) 建立严厉的问责惩罚措施。

为确保数字化加工过程中档案安全和质量，同时能够及时追责，一方面，在合同中明确规定了相关处罚机制，另一方面，建立了及时纠错及责任回溯机制。在合同中，对于公司出现质量问题规定严格的惩罚条例。在公司提交的批次数据中，经质检验收发现目录录入准确率、图像差错率、漏扫画幅数、目录与图像挂接正确率超出规定标准的，第一次不合格，整批次退回并返工，第二次不合格，即可终止合同，不再支付项目剩余资金，并可要求公司赔偿损失，尤其对图像漏扫更是零容忍，发现1画幅即扣除罚金1万元。

为确保合同执行，做到及时追责。在数字化加工各个环节制

定详细的登记、交接和数据流转登记制度，明确数字化加工过程中各个环节责任人，实行档案流转登记核查，加强数字化加工过程执行监督，确保及时发现问题、可追溯到个人、能够及时修正。通过严格的惩罚措施和有效地回溯追责机制，有效保障数字化加工的优质、高效。

2. 基于 SIFT 特征的图像匹配技术应用。

本项目中的明清档案没有编号，且部分为满文等其他文字档案，操作人员在发生重复扫描或者漏扫描时，因为没有明显标志，不易发现错误。虽然在扫描前，我们已经录入了每件档案的画幅数目，如果扫描画幅数目与档案画幅数目不一致，系统是可以自动发现的。但是如果数目对上了，但是漏扫了一幅而重扫了另外一幅，如何发现错误就很困难了，尤其是在操作人员不理解档案内容的情况下。

为此课题组引入了图像匹配技术，也称图像查重技术，主要是关注图像的局部特征，提取两幅图像中的各个稳定的特征向量，并进行对比。如果对比发现相同或相似的特征向量达到一定比例，则认为两幅图像重复。在本项目中，采用的是图像的 SIFT 特征。

技术原理：

SIFT 特征是基于物体上的一些局部外观的兴趣点而与影像的大小和旋转无关。对于光线、噪声、微视角改变等容忍度也相当高。基于这些特性，它们是高度显著而且相对容易提取，在母数

庞大的特征数据库中，很容易辨识物体而且鲜有误认。

SIFT 特征检测主要包括以下四个基本步骤：

a 尺度空间极值检测：搜索所有尺度上的图像位置。通过高斯微分函数来识别潜在的对于尺度和旋转不变的兴趣点。

b 关键点定位：在每个候选的位置上，通过一个拟合精细的模型来确定位置和尺度。关键点的选择依据于它们的稳定程度。

c 方向确定：基于图像局部的梯度方向，分配给每个关键点位置一个或多个方向。所有后面的对图像数据的操作都相对于关键点的方向、尺度和位置进行变换，从而提供对于这些变换的不变性。

d 关键点描述：在每个关键点周围的邻域内，在选定的尺度上测量图像局部的梯度。这些梯度被变换成一种表示，这种表示允许比较大的局部形状的变形和光照变化。

SIFT 特征向量匹配：当两幅图像的 SIFT 特征向量生成以后，下一步就可以采用关键点特征向量的欧式距离来作为两幅图像中关键点的相似性判定度量。取第一幅图的某个关键点，通过遍历找到第二幅图中的距离最近的两个关键点。在这两个关键点中，如果最近距离除以次近距离小于某个阈值，则判定为一对匹配点。

应用示例：

目前，该技术应用在同一件档案中可实现相邻画幅两两查重两种工作方式，即对相邻两个画幅或任意两个画幅进行查重。

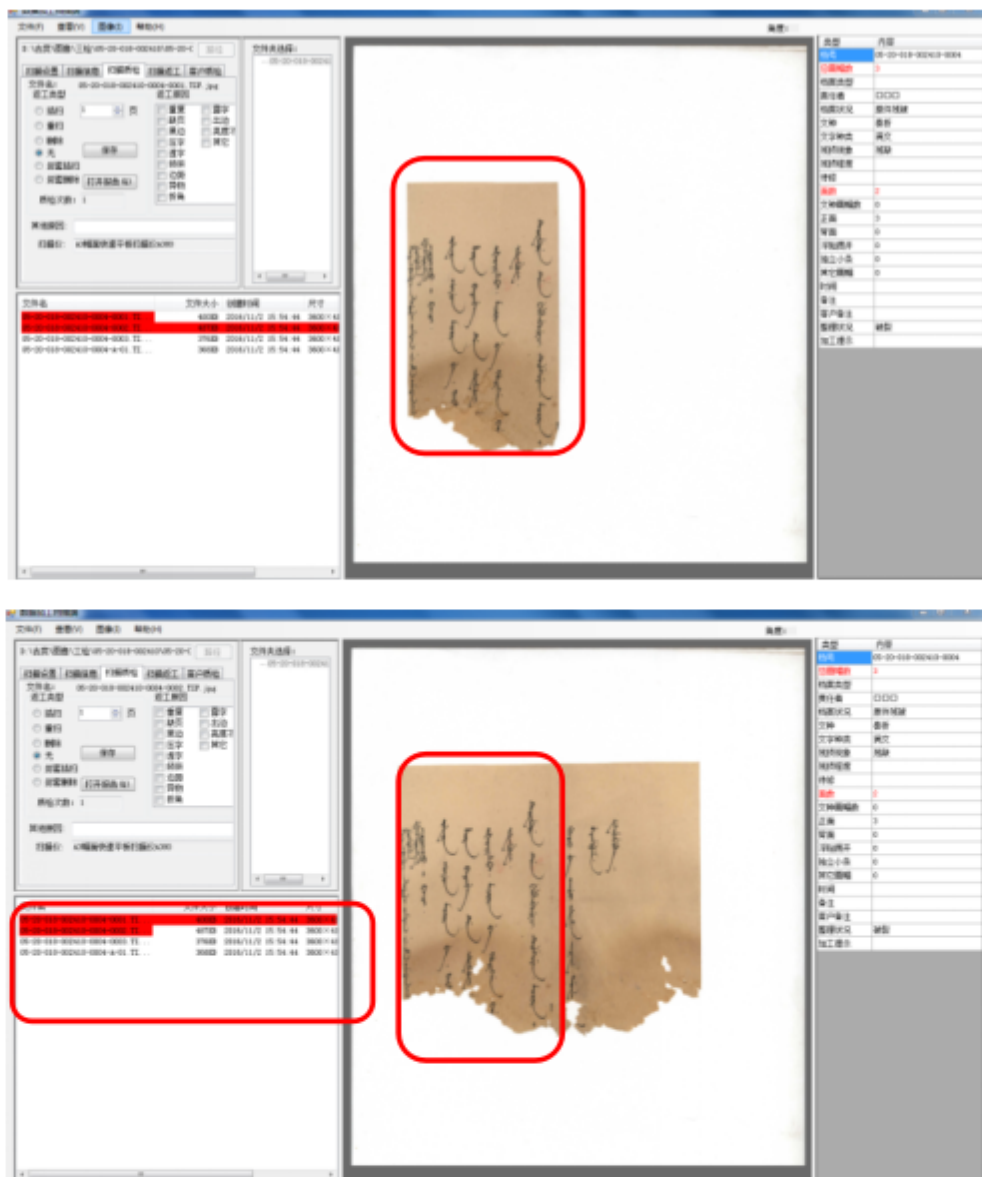


图 25-1 相似图像比较及软件红色警示

如上图所示，红色方框内图像为同一件档案的相同部分，两个画幅有大部分重复，此时系统通过图像查重，将对两相同图像目录进行标识，如图左侧红色方框。此时，该图像将无法作为成品数据提交馆方验收，必须返回上一道工序进行修改。查重结果不受图像平移的影响、不受图像缩放的影响、不受图像旋转的影响，比如扫描时图像出现任意旋转或者出现褶皱，查重软件都能准确的做出标记。查重工序由系统工具在服务器后自行运算，

依据档案特点、生产计划工具查重能力可随时扩展。通过测算，目前图像匹配度在 85%及以下，基本可实现完全查重，再结合人工质检，有效控制了图像的重复加工和漏扫。

5.3 质量管理体系风险控制

由于历史档案特殊的文化历史价值和文物价值，在海量数据加工过程中，档案原件安全和数据安全控制是必须高度重视的两个环节，这也是整个质量管理体系中亟需进行控制的两个风险点。我们在项目实施过程中，将档案原件安全放在了工作的第一位，同时严格控制数据流转和保管，风险控制举措比较完备，最终保证了档案原件安全和数据安全。

5.3.1 档案安全控制

1. 档案安全管理手段

本项目的安全目标是保证档案实体“一件不丢、一件不坏、一件不乱”，将档案安全问题放在第一位。为实现项目安全目标，我们主要从人力、实体、技术三个方面设置防范手段，其中以人力防范措施为主，具体如下：

(1) 人力防范：

馆方与外包公司签订“安全保密协议”，并对参加本项目的所有员工均进行政治审查，并要求其与外包公司签订“安全保密协议”；

馆方与外包公司均设立专职安全管理人员，负责实体档案和

电子数据的安全；同时负责加工现场人员管理、出入管理、监控管理、消防管理等；指定设备采购、维护等相关人员，负责现场设备安全与维护；严格工作登记制度，在执行过程中做好相关记录工作。

（2）实体（物）防范：

加工环境要求清洁、安静、防火、防盗、防震、防水，光线适当通风良好、无杂物（包括但不限于水或无关事物等）；加工设备保持稳定，定期维护保养；工具洁净，摆放整齐，专人专用；加工工位采用专用防水箱存放档案，加工场地附近设置合理数量的消防器材；

提供专门房间（通风良好）存放档案，使用专用档案柜（防水）存放待处理和已处理的档案（二者分柜保管），库房附近设置相应消防器材；提供专门工具（防水）运输档案；档案库房和加工区设置温湿度计，保持环境温度湿度符合标准。

（3）技术防范：

针对数字化工作的各工位安装监控设施，作为辅助手段保证加工过程中实体档案安全；使用软件加密算法对图像数据加密，保证加工过程中数据安全，加密过的图像数据在数字化加工系统环境内可显示，离开加工系统环境，图像数据则无法打开及阅读；加工系统按照国家保密局要求设计了标准的安全管理机制，设置有管理员、保密员、审计员三个管理岗位，相互制约控制整个系

统能够安全保密运行，有效保证数据安全；加工系统可通过管理端设置加工端空闲锁定时间，当人员离开工位超过所设置时间时，工位客户端系统自动锁定，人员回岗后需重新登录才能进入系统，有效保证客户端数据安全；加工电脑与外部交换数据的接口（如USB接口、红外线、蓝牙、火线接口、SCSI接口、光驱等）必须全部封断。

2. 纸质档案实体安全管理措施

本着安全第一的原则，我们主要从档案原件的交接管理、加工现场保管、数字化加工三个方面强化对纸质档案实体的全案保护，具体如下：

（1）档案原件提调安全

档案提调涉及保管处、数字化项目组及外包公司三方，三方均设置专职档案提调管理人员，档案提调须由三方提调管理人员全程监管；档案原件出库前要进行点数确认，出库要核对档案数量及目录，填写《中国第一历史档案馆档案原件出库交接单》，签字确认后方可出库；档案运送过程中要做到“车车有押运”，每辆车必须有馆方管理人员及外包公司专职工作人员共同在场；押运路线必须在馆内监控设备监控范围之内。

档案进入加工区库房前外包公司要点数确认，核点档案实际数量后，三方在《数字化项目档案原件提调交接单》上签字确认后方可进入加工区库房；档案原件归库前外包公司交馆方保管处

均需进行点数确认，核查归还档案数量无误，填写《数字化项目档案原件归还交接单》并签字确认后方可归库；档案原件归库后需填写《中国第一历史档案馆档案原件归库交接单》，并由馆方单位签字确认。

（2）加工区保管安全

加工区档案保管应按有关规定进行温、湿度平衡调整，安放消防设施，设置监控设备。档案周转库钥匙由馆方指定人员掌管，外包公司设置两名库管各掌管一张门禁卡（严禁将钥匙和门禁卡转交他人使用）须馆方及外包公司三人同时开门入库。除馆方指定人员和加工方两名库管外，原则上其他人员不得入库。任何人出入周转库，必须填写《库房出入登记表》并签字。

在档案库房中，实体档案应按“待整理”、“已整理”、“待修复”、“已修复”和“已扫描”分别保管；档案库房使用专用铁皮柜存放“待整理”、“已整理”、“待修复”、“已修复”和“已扫描”的实体档案，库房附近需设置相应的消防器材，铁皮柜需架高放置以防水。

每个加工岗位配有专用档案存放箱，用于临时存放扫描加工的实体档案，加工人员离开工作岗位档案存放箱要上锁；加工区设置实体档案加工流转柜，临时存放加工流转中的实体档案。流转柜中档案必须每天入库，不得过夜存放；实体档案当日领出当日归还，加工工位上档案不得过夜留存。

(3) 数字化加工现场安全

加工区钥匙及门禁卡由馆方 2 名工作人员分开保管，每天开关门须由馆方 2 名工作人员及外包公司安全管理人员同时在场方可。关门之后馆方工作人员需将钥匙放入专用保险柜中保存并登记。

加工现场馆方和外包公司设置专职安全管理员，对加工现场全程巡查监督，保障加工现场档案安全；点数、展平、扫描等各项工作需在监控设备运行状态进行，确保档案实体不离开监控范围。

外来人员进入加工区需填写“外来人员出入登记表”；工作人员进出前整理工作区需登记；严格工作流程，馆方现场管理人员须填写工作日志，记录加工现场每天工作情况。

外包公司工作人员填写“档案数字化工作流程单”，明确档案流转过程中各个环节责任人；保障加工档案领用有序安全。由库管及加工员应按规定数量，将档案装入可完全封闭的物流箱（防水），并根据档案信息填写“档案数字化工作流程单”，随档案实体及物流箱一同流转。工具上标明编号，使用人员需登记领用，原则上一人一具，不得混用；

做好档案数字化加工登记。修复档案交接需填写“档案修复登记表”，补扫档案交接需填写“档案补扫登记表”，目录册交接需填写“目录册出入库登记表”；库管将“档案原件调交接单”、

“档案出入库登记表”、“档案修复登记表”、“档案补扫登记表”、“目录册出入库登记表”扫描和录入项目管理系统，以便查询和管理，项目结束后，将上述表单电子数据导出移交馆方。

制定标准的实体档案操作规范，包括运载、翻阅、拿取、摆放等，严格要求所有工作人员执行制定的统一标准；“扫描前处理”工作环节必须在馆方工作人员指导下按照要求完成，以保证档案原件安全、完好无损；加工过程中，注意对纸质档案的保管和保护，注意防火、防水、防光；不准将与工作无关的物品带入工作间内，如食品、水、油渍物品等，要让档案与机器设备远离污染源，严格保管与维护；加工完成后，加工现场要清场，清场后由技术主管在“安保员”和“项目经理”的监督下对工位和设备进行除尘工作。

5.3.2 数据安全控制

数据安全控制方面，我们采取物理控制和人为控制相结合的方式，对软硬件进行严格控制，结合流程控制和工作规范，确保了数据安全。

1. 物理控制

全部加工、管理设备处于独立局域网环境下作业，必须与互联网和办公网完全物理隔断；加工设备与加工员一对一绑定，加工员禁止操作除自己加工设备以外的任何设备；加工电脑与外部交换数据的接口（如不必要的USB接口、红外线、蓝牙、火线接

口、SCSI 口、光驱等) 必须全部封断。

进入加工现场不得携带任何电子设备(包括但不限于手机、相机、笔记本、游戏机、MP3/4、U 盘、移动硬盘、软盘、光盘等)。

利用数据上传功能, 实时将扫描的电子影像上传到一史馆的服务器和盘阵中。数据库管理系统的自动备份功能, 采取“每日备份、多重复制、异地存放”的策略, 保证数据万无一失。

2. 人为控制

工作电脑只能安装操作系统, 以及杀毒软件、必要的设备驱动和数字化加工软件, 严禁安装除此以外的任何其它软件; 工作电脑出现故障, 必须由馆方指定人员维护。

各工种操作人员使用自己的工号和密码登录加工系统, 严禁对外透露个人的工号和密码, 严禁使用他人的工号和密码登录; 同时个人的密码将定期进行更换。

使用软件算法对影像数据加密, 保证加工过程中数据安全, 即使“非法”得到数据也无法打开及阅读; 各工种加工人员离开自己的加工岗位半小时以内的情况, 必须退出加工系统操作界面; 离开加工岗位半小时以上的情况, 必须关闭计算机。严禁出现软件运行着但无人操作的情况。

为了保证加工的顺利进行, 加工现场需准备一定数量的备用机, 对损坏的设备要及时进行修理(所供设备维修时间均低于 1 天)。注重设备的日常保养, 每天都由专人进行维护工作。加工现

场配备必要的不间断电源和漏电防护等设施。

成品数据提交后，清除加工端数据。（注：清除数据采用格式化硬盘的方式）成品数据验收后，清除服务器和盘阵数据。（注：清除数据采用格式化硬盘的方式）。

工作人员不得擅自在加工电脑上使用外来设备，外来设备必须由项目经理指定的专人（技术主管）检查和调试安装。

6. 课题成果推广应用范围

6.1 课题特点概述

本课题在广泛研究档案数字化管理国内外研究成果的基础上，结合明清历史档案数字化的特殊要求，制定了《数字化加工项目手册》、《数字化加工及管理基本准则》、等 10 余项管理制度、技术标准和工作规则，建立了一整套明清历史档案图像数字化加工质量管理体系，具有较强的理论创新价值和实际推广应用价值。

课题基于明清历史档案图像数字化加工质量管理体系，研发了一套明清历史档案数字化加工软件。该软件引进了 workflow 技术、表格登录识别技术、图像匹配技术及图像自动处理技术，实现了大规模图像数字化加工中计算机信息处理技术的应用及融合，技术特点鲜明，具有较强的创新性和开拓性。此外，课题组还发表了《谈明清档案数字化外包管理的难点与对策》《档案数字化加工外包过程质量控制研究》《浅谈明清档案数字化图像加工的若干思考》等多篇学术论文，对明清历史档案数字化加工项目管理的理论、方法、手段等进行了有效的总结，学术价值较高。

整体来看，本课题有如下特点：

1. 质量管理由点及面，贯穿始终，呈现良好的系统性及科学性。

该体系下的质量管理，包括组织体系建设、流程控制及业务能力建设、制度标准建设、设备运维管理、数字化加工自

自动化技术应用等方面，涉及数字化加工各个方面，由点及面，系统性强；质量管理要求贯穿明清档案图像数字化生产全流程，在前处理、著录、扫描、自动质检、人工质检、成品提交六大工序中，均形成了专门的质量控制手段，并实现后一工序对前一工序结果时时验证，不同工序间可进行相互监督。基于此体系，合作双方组织层级清晰，权责明确，沟通顺畅，加工人员数量及业务水平相对稳定，制度标准及操作规范针对性强，易操作，有效解决了数字化加工过程中各种问题，实现了加工质量的实时控制，整体运转顺畅，呈现了良好的系统性及科学性。

2. 自动化技术应用特点明显，针对性强。

一史馆明清历史档案具有较高文物价值，极其珍贵，但年代久远，纸张状况较差，残损程度较高，对档案安全及质量要求极高。课题组通过引入 workflow 技术、表格登陆识别技术、图像匹配技术以及图像自动处理技术，实现了档案加工流转全程定位跟踪，在服务器上自动对工序、人员、任务执行时间进行详细记录，确保档案可随时追溯，问题有据可查。同时，解决了档案目录元数据信息录入效率低、录入错误率高等问题，攻克了明清历史档案不标记页码扫描易漏、易错这一难题，实现了档案图像处理自动裁切、自动检查命名、封套信息等，技术应用效果明显，针对性较强。

3. 现实指导意义较强，具备一定的推广应用价值。

该体系针对性较强，形成了一系列高要求的历史档案数字化图像质量标准，对一史馆档案数字化指导作用明显，能够很好的起到填补领域空白、引领行业趋势、规范行业标准的效果，对于明清档案数字化事业具有重要意义。同时，该体系中质量管理的方法及技术应用亦适用于其他类型档案数字化工作。如：体系中表格登陆识别技术可应用于组织机构代码和身份证号识别、银行和医院各类单据识别工作中，图像匹配技术也可应用到各档案馆、图书馆和出版社的数字化工作中，辅助人工质量管理，避免出现画幅重复、漏扫等问题。该体系的质量管理方法和技术应用案例参考适用范围较广，具备一定的推广应用价值。

6.2 课题可推广应用范围

1. 可直接应用于明清历史档案抢救保护工作。目前，一史馆尚有约 400 万件档案未开展数字化，不但给这些珍贵历史档案的安全工作带来很大隐患，且严重地制约着档案的利用和开放，制约着明清历史档案的社会效用，因此，迫切需要对这些档案进行整理和数字化抢救保护。该项目的顺利实施将能够为明清历史档案图像数字化加工提供有效支撑，为国家珍贵文化遗产的保护和传承做出应有贡献。

2. 为历史档案数字化加工提供有效借鉴。通过对明清历史档案图像数字化加工质量管理体系的研究和实践，为其他档案馆开展明清历史档案数字化提供借鉴和指导，可以推广到各种历史档

案数字化工作中进行使用。如，图像匹配技术就可在藏文档案数字化、外文档案数字化等小语种文字档案数字化工作中发挥重要作用。本项目工作模式还可为其他专业档案保管机构、图书馆以及文化单位等提供借鉴，进一步推动全国各行业数字资源建设步伐，为振兴国家文化事业做出贡献。

3. 树立明清历史档案图像数字化加工行业标杆。通过项目有效实施，形成体系化、标准化、模块化的质量管理体系，形成一系列高规格的历史档案数字化图像质量标准，引领国内外明清历史档案图像数字化加工开展，为推动明清历史档案图像数字化加工国家标准及行业标准的建立提供有效支撑。

4. 为其它行业自动技术应用提供参考。

以课题中 workflow 技术应用为例，其不仅可以运用在档案数字化加工管理中，在信息化网络建设，财务管理、协同办公、项目管理、供应链管理、订单管理、物流业务平台、人力资源管理、绩效管理等不同的领域都有其相应的应用方法、应用软件集成。可推广应用的行业包括制造业，消费品行业，电信服务业，银行、证券、保险等金融服务业，物流服务业，物业服务业，物业管理，大中型进出口贸易公司，政府事业机构，研究院所及教育服务业等，特别是大的跨国企业和集团公司。通过 workflow 系统对应用系统集成、对质量文档进行管理，使数据流动起来，在流动中增值。

7. 总结

在明清历史档案数字化加工方面，本课题属首次采用外包项目管理模式与现代计算机技术相结合的流水式数字化加工。加工效率、质量标准在国内外历史档案数字化加工领域处于领先地位。本项目能够很好的起到填补领域空白、引领行业趋势、规范行业标准的效果，对于明清档案数字化事业具有重要意义。这也是明清档案数字化领域首次大规模引进社会力量，通过技术与人才上的合作，实现明清档案数字化工作模式上的突破。一方面，其实实现了对传统历史档案图像加工模式的传承，填补了明清历史档案海量数字化图像加工领域质量管理体系建设的研究空白。另一方面，项目顺应信息化时代要求，实现了大规模图像数字化加工中计算机信息处理技术的应用及融合，创新性和应用性较强，具有良好的研究价值和现实指导意义，对国内外其他历史档案数字化工作起到了借鉴作用。

参考资料:

- 【1】史江 秦慧 赵德喜 历史档案数字化进程中的业务外包问题探索 档案学通讯 2010年第1期
- 【2】中国档案学会档案信息化技术委员会 关于档案数字化质量控制体系的研究 2014年档案事业发展研究报告集
- 【3】毛建军 明清档案数据库的建设与思考 档案与建设 2010年第11期
- 【4】黄丽华 宋 华 档案数字化外包安全管理研究 中国档案 2015.6
- 【5】赵彦昌 王红娟 中国流失海外的满文档案文献及其追索 山西档案 2010年第6期
- 【6】赵彦昌 韩 娜 美国档案数字化工作的现状及启示 国际档案界 2008年第4期
- 【7】潘维艳 历史档案的数字化、网络化研究 兰台世界 2011·10月中旬
- 【8】褚瑞霞 国外图书馆、档案馆的数字化及启示 《枣庄师专学报》第18卷第4期 2001年8月
- 【9】戴小蕙 历史档案的数字化技术保护策略 兰台世界 2013·6月中旬
- 【10】史 江 秦 慧 赵德喜历史档案数字化进程中的业务外包问题探索 档案学通讯 2010年第1期
- 【11】赵彦昌 论满文档案数字化的两个问题 兰台世界 2012·2月中旬
- 【12】赵彦昌 王红娟 满文档案数字化的措施及其方法 兰台世界 2012·5月增刊
- 【13】华 林 少数民族文字历史档案的数字化建设 中国档案 2005.11
- 【14】葛 俊 档案数字化业务外包存在的问题与对策 档案建设 2015.4
- 【15】孙 宇 国内外数字档案馆比较研究 信息化建设 2015.10
- 【16】钟国文 童 霞 提高档案数字化外包质量的策略探析 机电兵船档案 2012年第1期
- 【17】王玉瑾 宋魏巍 我国历史档案数字资源体系建设研究 黑龙江史志 2014.17

