

U.S. DIGITIZATION PROJECT

JUST THE FACTS



PROJECT

Digitization of paper documents of varying age, size and density

STATISTICS

40 million documents, some more than a century old, scanned in less than four years

SOLUTION

- Dedicated on-site scanning staff
- Custom software programs for project streamlining
- Delivery of 300 dpi color TIFF uncompressed files

EQUIPMENT

InoTec 400-series document scanners

InoTec Document Scanners Prove Archival Digitization and High Throughput Can Coexist

Crowley Imaging recently completed a high-volume, on-site digitization project scanning nearly 40 million archival documents in just under four years using InoTec 400-series document scanners. Known for their 24/7 daily duty work cycles, high throughput and low maintenance requirements, the InoTec document scanners are generally considered a tool for records managers, service bureaus and health care systems. In this instance, the need for speedy production paired with a requirement for archival image quality delivered by InoTec's exclusive line array CCD and precision



lenses made the InoTec scanners an unusual and ideal choice.

Key features for selection included:

- Gentle, no maintenance belt transport system
- Reliable paper output
- Focused cool LED lighting
- Glassless paper guide
- Perfect Document Technology (PDT)

Project Overview

The conversion project included the digitization of 40 million documents of various sizes, quality control and digital content delivery. At the highest point of production, there were six (6) InoTec 400-series scanners on-site: four for scanning; one dedicated to processing any necessary rescans; and one for emergency use.

Project Specifications

The project required images to be saved as 300 dpi color TIFF uncompressed files. In a

somewhat unusual request, the client required an exact image of each original with no image enhancement. Each document was scanned in duplex.

Project Challenges

As expected, a project of this volume created and encountered several challenges, all of which were well-met by the programming expertise of the Crowley Imaging software engineers, the scanning experience of the on-site technicians and the robust construction of the InoTec scanners.

Dust and Debris: The first batch of documents was fairly uniform and in good shape but years of sitting in one place, inadequate archiving and general abuse via public use led to a variety of quality issues. Because of their glassless guide system, the InoTec scanners handle dust and debris very well. As with any paper scanning effort, operators did have to clean debris and paper flakes from the scanner, but the built-in traps prevented much of this from falling onto the InoTec's bottom mirror. Additionally, InoTec scanners include a fan in the base to pull dust down and out of the scanner.

Prep and Sorting: Many of the documents were stapled together as sets. The client performed most of the prep duty, removing staples and flanking each set – front and back – with a barcoded yellow separator sheet. The separator sheets were scanned in sequence with the rest of the batch. A custom software program was written to digitally identify the barcodes later in the naming process as a set of images that belonged together.

Capture Border: The client requested the capture of an additional 2mm around each image to ensure that the entire item had been captured. InoTec settings were used to create this additional scan space.

Storage: For fast data transport, InoTec scanners connect to a PC via gigabit ethernet cards. Each computer on the project was also connected to a gigabit switch which in turn fed into a makeshift server that handled all image storage and processing. The server was outfitted with

six terabytes (6TB) of internal storage and held all originals and derivatives of the scanning. All images were transmitted to the server in real-time via the network connection.

Naming: Each image name was comprised of a three-letter static prefix. The first letter represented the scan station; the second and third letter represented the initials of the scan technician. This was followed by a 5-digit sequential number and an 'A' or 'B' for front/backside recognition.

Material Condition: The oldest documents dated from the late 1800s and were brittle and in poor condition. The client, concerned about further damage from scanning, requested the use of a straight-path scanner for these older materials. However, when scanned side-by-side, it was quickly demonstrated that the InoTec's rotary system – which uses the combination of glassless guides and gentle polyester belts – posed no more risk to document integrity than that of a straight-path scanner. For material that was aged, damaged or fragile, the operators were able to slow the scanner speed to ensure that the integrity of the document was not compromised.

Even while operating at slower speeds, the InoTec document scanners **increased throughput by 400%** as compared to the straight path scanner, translating to hundreds of dollars per hour in labor savings before post-processing. From an image, gentle-handling and financial standpoint, the decision to stay with InoTec document scanners on the entire project was easily made.

Speed: Each batch averaged 1500-1700 documents. When the material was recent or in good condition, a batch was scanned in 20 to 30 minutes. At the highest point of production, 80 batches were scanned per day – an estimated 120,000 documents/240,000 images.

Custom Software Solutions: Due to the project's high-volume and the exact archival standards of the client, the following programs were created by Crowley Imaging's software engineers for efficiency and image quality:

- A filename detection program to track all items mentioned in the naming section (scanner, scan tech, image count) that managed each item on a batch basis from the time it was checked into the system to the time the deliverables were accepted by the client.
- A rainbow detection system. No matter how sterile an environment, dust and debris will always be present with documents, particularly with older material. Occasionally, small particles of dust or flecks of document corners would appear in the images, manifesting as rainbow streaks – sometimes faint, sometimes pronounced. Although this did not affect text or image capture, operators were instructed to locate and repair even the slightest and most translucent occurrences. A software program was



created which checked for color variations of brilliant and neon reds, blues, greens, and yellows. All images were run through this processing program before quality control inspection; offending images were re-scanned.

- Once manual quality control inspection was complete, a custom program prepared the images for digital content delivery for ingestion into a content transfer system. Automated preparation included applied metadata, a complex naming schema, the generation of an MD5 text file for each image and other various text files filled with requested info.

The combination of a knowledgeable and dedicated staff familiar with archival preservation digitization and the high-production capabilities and image quality of the InoTec 400-series document scanners ensured that 40 million documents were digitized and delivered correctly and on time.

For more information about Crowley Imaging or the InoTec family of document scanners, please call **(240) 215-0224** or email studies@thecrowleycompany.com.

About The Crowley Company

Incorporated in 1980, The Crowley Company is a leading digital and analog film technologies company headquartered in Frederick, Md. with manufacturing divisions (Mekel Technologies, Wicks and Wilson, Extek and HF Processor brands) in California and the United Kingdom. With over 100 employees, The Crowley Company provides an extensive number of digital document and film conversion services to the publishing, commercial, government and archive sectors. It also manufactures, sells, and services high-speed microfilm, microfiche and aperture card scanners, microfilm duplicators, film processors and micrographics equipment.

www.thecrowleycompany.com



thecrowleycompany



Thecrowleycompany



the-crowley-company



crowleyscans



@CrowleyScans



CrowleyScans

CROWLEY

5111 Pegasus Court, Suite M, Frederick, Maryland 21704 USA • 240.215.0224
www.thecrowleycompany.com • studies@thecrowleycompany.com