

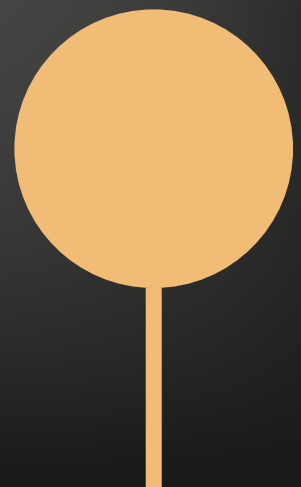


With a Bold Implementation
of Artificial Intelligence,
the U.S. Nuclear Regulatory
Commission Becomes the
First Federal Agency to
Integrate Data from Millions
of Documents with Virtually
no Human Interaction

A Shockingly
Powerful
Government
Innovation
Disrupts the
Future of



DOCUMENT
WORK
FLOW



ORCHESTRATING DATA MIGRATION AND DIGITIZING HISTORICAL RECORDS

If you look at the history of nuclear reactor accidents, the United States has an excellent track record. In all reported accidents since 1961 only one resulted in a very small radiation leak. The U.S. Nuclear Regulatory Commission (NRC) is the agency that protects the U.S. population from nuclear-related accidents.

Formed in 1974, the NRC regulates all commercial nuclear power plants, reactors, and radioactive materials through licensing, inspection, and policy enforcement.

Just how big a job is it? They are responsible for the safe operations of 58 nuclear power plants, 96 nuclear reactors, and the approval process for constructing / decommissioning nuclear facilities. History has proven the destructive power of nuclear reactor failure and there is no margin for error.

Increasing regulation and oversight, and the Freedom of Information Act introduced challenges that pushed the limits of NRC technology and threatened to put safety at risk.

Two primary records management regulations were at the epicenter of NRC's incredible workflow innovation. These were put into place by the National Archives and Records Administration (NARA), and carried a soon-approaching deadline. Here are the requirements:

-  1. Ensure that all Federal records are created, retained, and managed in electronic formats with appropriate metadata
-  2. Consistent with records management laws and regulations, develop plans to close agency-operated storage facilities for paper, and other analog records, and transfer those records to Federal Records Centers operated by NARA or commercial storage facilities

As you might imagine, the NRC maintains a massive amount of public and classified information. They must balance efficient, accurate, and secure data workflows with the demands of budget constraints and transparency.



MULTIPLE CHALLENGES

THE PROJECT: MICROFICHE AND APERTURE CARD DATA MIGRATION

NRC staff maintains access to data from over 39 million pages of records stored on microfiche and aperture cards. The NRC is driven to not only comply with Federal records mandates, but also to provide the best possible experience for the way NRC stakeholders and the public interact with records.

With over 2.5 million documents containing over 39 million pages of information stored on film, finding a solution to integrate the data contained within them while preserving an electronic copy was imperative for NRC CIO David Nelson. Previous attempts at processing a fraction of their records took 5 years to accomplish and produced poor results that required immense manual data entry.

Many factors make data extraction from microfiche and aperture cards difficult.

One is that documents on film are contained in frames. In cases where sensitive information needed to be removed, entire frames were scratched from the surface of the film. Another difficulty is that film usually contains physical defects introduced during scanning. Environmental factors like dust create blobs and blemishes on document images. And in many cases age and normal wear and tear make accurate optical character recognition (OCR) a difficult task. "I envisioned a software that could just migrate all this data," said Nelson. "An artificially intelligent software that would orchestrate the migration and integration of data with our systems. As a music fan, I imagined us directing a symphony of data where information flowed seamlessly from documents into our database. It was a dream I wasn't sure would come true."

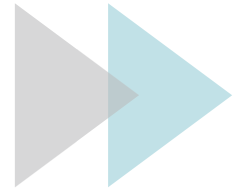
Knowing there had to be a more intelligent and cost-effective way to process archived microfiche and aperture cards, NRC went in search of a modern document processing solution and discovered Grooper.



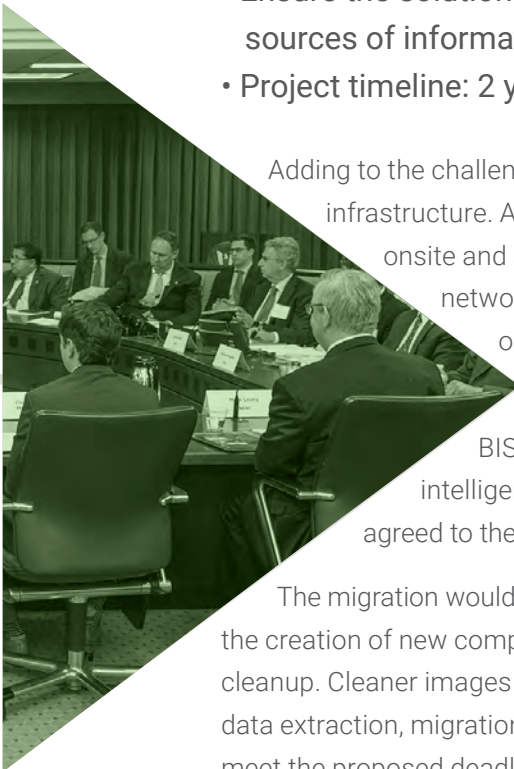
BUT WAS GROOPER UP FOR THE CHALLENGE?

THE REQUIREMENTS

The requirements were straightforward and demanding



- Automate data collection on forms
- Process 39 million pages of historical information
- Provide highly accurate OCR on all documents
- Compare extracted data with internal NRC databases
- Link all information to NRC Agency-Wide Document Access and Management System (ADAMS)
- Ensure the solution would scale to process new sources of information
- Project timeline: 2 years



Adding to the challenge, data could not leave NRC secured infrastructure. All scanning work was to be performed onsite and all data processing within a secured network. All scanning equipment and operators would have to pass robust security clearances and controls.

BIS, parent company of the Grooper intelligent document processing platform agreed to the challenge.

The migration would require immense automation and the creation of new computer vision algorithms to aid in image cleanup. Cleaner images would help streamline and automate data extraction, migration, and metadata creation in order to meet the proposed deadline.

THE PLAN

After issuing a purchase order for Grooper, a software license was transferred to the NRC.

Using a subset of NRC archives as a test, multiple data extraction models were configured within Grooper. Models were trained using machine learning and natural language processing to identify complete documents and extract the necessary metadata.

Document images and pertinent data would be integrated into NRC databases and document management systems.

4 Mekel MACH7 high-speed film scanners were selected for the project because they were the fastest film scanners available. They were delivered and installed within NRC secured infrastructure to pass security inspection.

Because microfiche contains up to 96 document pages per card, workers would need to process approximately 20,000 microfiche / aperture cards every month to stay on schedule. 9 workers would be used:

- **126 worker-hours per day**
- **6 scan operators**
- **3 image quality reviewers**

After all equipment and workers passed security clearances, work would begin.

96
DOCUMENT
PAGES PER
CARD

The Solution to Drastically Reduce Timeframe, Cost, and Cycle Times

NRC traded legacy document scanning for Grooper, a shockingly powerful high-efficiency intelligent document processing and data migration platform.





All scanning and hardware configurations were fully controlled by Grooper. By offloading image processing to Grooper, NRC Mekel MACH7 high-speed film scanners were able to run at the hardware's maximum speed.

For microfiche and aperture cards, Grooper computer vision was configured to automatically power through any physical defect found on microfiche and aperture cards. Image processing algorithms produced two copies of the documents; one for pristine human-readable PDF document images, and the other for best-in-class OCR.

Patented Grooper OCR processes were used to identify every single word on every scanned document. This was accomplished by using multiple OCR engines, and pre-built data overrides.

Natural language processing and classification engines grouped and classified pages that spanned multiple physical cards.

Up to 20 individual metadata attributes were identified, added to documents, and compared against NRC internal databases to match documents to the correct ADAMS database Accession Number.

Finally, automated workflows dealt with duplicate forms, implemented classification for over 400 individual document classification codes, and provided manual review to match any documents that didn't automatically correspond with an Accession Number or classification code.

**20
METADATA
ATTRIBUTES**

“It’s happening around us, we’re leaning forward.” Nelson said “To the vendors, please really focus on the values of your offerings, help us with new contracts, work with us, and be a partner, help us solve the problems and our real challenges, not necessarily what you think might be holding us back. We are moving rapidly to modernize government.”

THE RESULT

**39 MILLION
PAGES OF
EXTRACTED
MICROFICHE
AND
APERTURE
CARDS
IN JUST
9 MONTHS**

NRC successfully migrated data from 39 million pages of extracted microfiche and aperture cards in just nine months, well on track to meet the proposed deadline.

"Quality of document images, including OCR results, compared to previous conversion work was 100 times better," said NRC Archivist Thomas Wellock.

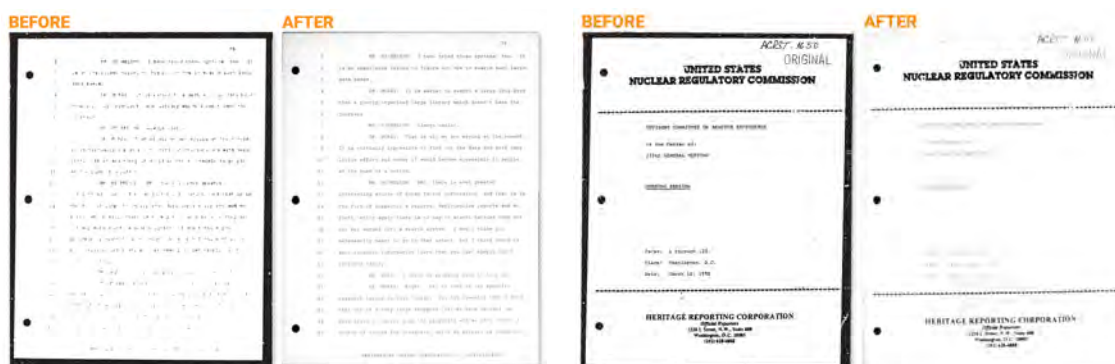
Forms processing workflows, costs, and timelines were radically improved.

Because NRC purchased software instead of services, they empowered employees in their document processing center to focus on more productive work instead of manual data entry. Future projects include integrating data from hundreds of thousands of paper records.

With a single-line-item software purchase, NRC proved to the world the value of automating data integration with A.I.

If you think your records conversion project is too expensive, or too complex, or simply not achievable, Grooper intelligent document processing is the solution for rapid, accurate, and cost-effective document data integration.

MeriTalk, a publication focused on improving the outcomes of Government IT, interviewed NRC CIO David Nelson after a meeting he attended at the White House regarding tech vendors. "It's happening around us, we're leaning forward." Nelson said "To the vendors, please really focus on the values of your offerings, help us with new contracts, work with us, and be a partner, help us solve the problems and our real challenges, not necessarily what you think might be holding us back. We are moving rapidly to modernize government."



View document images up close, visit <https://go.bisok.com/nrc>.

GROOPER

EMBEDDING INTELLIGENCE WHERE
DOCUMENTS AND DATA MEET.



GROOPER was built from the ground up by BIS, a company with 35 years of continuous experience developing and delivering new technology. Grooper is an intelligent document processing and digital data integration solution that empowers organizations to extract meaningful information from paper/electronic documents and other forms of unstructured data.

The platform combines patented and sophisticated image processing, capture technology, machine learning, natural language processing, and optical character recognition to enrich and embed human comprehension into data. By tackling tough challenges that other systems cannot resolve, Grooper has become the foundation for many industry-first solutions in healthcare, financial services, oil and gas, education, and government.

WWW.GROOPER.COM