

FAIRFAX COUNTY AUTOMATED REDACTION CASE STUDY

Fairfax Circuit Court leverages Extract Systems' automated redaction software (ID Shield™).

The screenshot displays the Extract Systems' ID Shield software interface. The main window shows a document titled "COMMONWEALTH OF PENNSYLVANIA DEPARTMENT OF TRANSPORTATION Bureau of Driver Licensing Mail Date: JANUARY 1, 2015". The document text is partially obscured by redaction boxes. A sidebar on the left contains a table with columns for Text, Category, Type, Page, and Exemptions. The table lists several redaction items, including "02020202" (MCData, DLN, Page 2), "[No text]" (Manual, Page 3), "90909090" (HCData, DLN, Page 3), and "01/01/1985" (HCData, DOB, Page 3). The bottom of the interface shows a status bar with "Click and drag to draw a rectangular redaction." and "OCR available | Zoom: 37% | 300 x 300".

Text	Category	Type	Page	Exemptions
02020202	MCData	DLN	2	
[No text]	Manual		3	
90909090	HCData	DLN	3	
01/01/1985	HCData	DOB	3	

extract

EXTRACTSYSTEMS.COM

EXTRACT'S SOFTWARE HELPS BALANCE PUBLIC ACCESS WITH PERSONAL DATA SECURITY IN FAIRFAX COUNTY, VIRGINIA.



BACKGROUND:

Fairfax County is the most populous jurisdiction in the Commonwealth of Virginia with a population of 1.1 million people. They have a large need for removing Social Security Numbers from important documents while protecting the people of their county's private information.

Challenges:

BALANCING DATA ACCESS WITH PERSONAL DATA PRIVACY:

Fairfax County desired to balance data access with personal data privacy by redacting records available via its secure, remote access database.

The county issued a national RFP to select a redaction service provider to process **44.5 million pages of historical documents dating from 1742 for the removal of Social Security Numbers**. Additionally, the county sought automated redaction software to process **2.5 million pages of day-forward documents annually**. The software needed to integrate with the county's custom in-house system, several third party applications, and database.

The requirements set forth in the county's RFP were technically complex, including functionality to meet future redaction mandates and the ability to **achieve a 99.5 percent accuracy rate**. As an additional challenge, the backfile project needed to be completed within six months.

CONFIGURABLE SOFTWARE REDACTION RULES:

One of the most unique technical aspects of the project was the need to simultaneously process documents to completely redact Social Security Numbers alongside documents that required the truncation of Social Security Numbers, that is, redaction of the first five digits only. The court's Judgment documents required that only the first five-digits of Social Security Numbers be redacted, leaving the last four digits visible. Fairfax Circuit Court handles different types of documents, so they needed software with advanced document classification capabilities to apply the correct redaction rules to specific document types.

DOCUMENT TYPES INCLUDED:

- Marriage Licenses
- Charter Documents
- Financing Statements
- Certificates of Trade Name
- Guardianship & Trust Orders
- Wills
- Liens
- Fiduciary Bonds
- Deeds
- Asset/Estate Inventories

FAIRFAX COUNTY'S VISION:

To find a redaction service provider that has a high accuracy rating and no data breaches on private personal information, that would work on historical and day-forward documents in less than a six month timeframe.

"THROUGH OUR PARTNERSHIP WITH EXTRACT SYSTEMS, WE WERE ABLE TO REDACT 44.5 MILLION BACKFILE AND GAP IMAGES AND INTEGRATE REDACTION WITH OUR DAY-FORWARD PROCESSES – ALL A MONTH AHEAD OF SCHEDULE."

**-BRENNA SAWYER
PROJECT MANAGER
FAIRFAX CIRCUIT COURT**

Extract Solution:

ACTION:

After an extensive review process, Fairfax County selected Extract Systems to provide redaction services for its backfile. All images were processed through ID Shield Advanced Redaction Software combined with a manual review of the pages identified as potentially containing sensitive information. Though the requirement was to deliver the project within six months, Extract delivered the project in **five months**, and achieved **an accuracy rate of 99.5%**.

RESULTS:

ID Shield Redaction Software has been setup to perform a hybrid redaction process where documents are **automatically redacted** if they contain high confidence data items. Automatically redacted documents are swept into Fairfax County's system and made available **immediately**. ID Shield is configured to queue files with medium or low confidence data items into a verification queue. Files in the verification queue are reviewed by a set of operators using the ID Shield desktop client software. Lastly, Extract worked with Fairfax County developers to implement a connection to its API to enable launching of the verification user interface directly from the county's web browser-based application. This allows operators to review specific documents in the queue on an ad-hoc basis.

Benefits:

Extract Systems helped Fairfax County by implementing ID Shield Advanced Redaction Technology. Fairfax County was looking for a solution that could redact historical and day-forward data in a time period of less than six months. Extract was able to get through all of their data in just five months, a month ahead of schedule. Not only did they accomplish their task ahead of time, they were able to seamlessly integrate with Fairfax County's document management system

Software Detail:

EXTRACT REDACTION SOFTWARE (ID SHIELD):

Extract Systems redaction software starts with the industry-leading OCR engine, deploys machine learning to classify documents, and layers additional recognition technology into the software to find data within complex unstructured documents.

Extract's state-of-the-art data detection rules find and redact all sensitive data types including: Social Security Numbers, dates of birth, credit card numbers, CVV codes, account numbers, check numbers, expiration dates, names, postal and email addresses, and virtually all types of personal information within structured and unstructured documents. This core competency is a proprietary formula that incorporates pattern recognition logic, phrase context proximity and keywords into the software based on the data that needs to be captured and secured.

EXTRACT
DELIVERED
THE PROJECT
IN FIVE
MONTHS,
AND
ACHIEVED
AN
ACCURACY
RATE OF
99.5%.

ID SHIELD
HAS
PROCESSED
MORE
THAN
FOUR BILLION
DOCUMENTS
FOR OVER
400
CUSTOMERS
WITHOUT
A SINGLE
DATA
BREACH.

Redaction zones are burned in, making it impossible for identity thieves to retrieve private information. All hidden meta data associated with the original document is eliminated. Depending on customer requirements, the original document always remains unaltered, but the copy (or copies) of the original can be redacted and saved to a database, or can be indexed, then redacted. Redactions can be applied to the image, or the location of the sensitive information can be noted for redaction at a later time.

THE EXTRACT SYSTEMS™ PLATFORM:

Extract is a leading provider of advanced data capture and redaction solutions that drive operational efficiency and secure private information for government, healthcare, and other commercial sectors. Extract's advanced redaction technology has processed more than 4-billion documents for over 400 customers without a single breach in data protection. Extract's data capture products automatically find and incorporate manual data entry fields into designated information systems as structured data. Automatic capture of information trapped in paper-reliant workflows reduces costs and optimizes the intelligent collection of data.