opentext™

Government Agency Secures Data and Achieves Faster PCI Compliance with OpenText

Customer uses OpenText™ Redact-It™ Enterprise with Optical Character Recognition to ensure millions of document pages don't include credit-card data

Public-facing government entities tend to work with high volumes of data, so compliance with regulations around that information can prove challenging. OpenText helped one agency become compliant with a solution to search and classify its millions of stored documents, redacting any credit-card information discovered.

A large government agency identified a large subset of its legacy documents that contained credit-card numbers needing to be remediated to meet Payment Card Industry Digital Security Standards (PCI-DSS). A spot-check review then uncovered that there were many different document types, multiple versions of forms with varying content, and that no associated metadata existed that could be used to classify the documents for remediation. Because of the volume and number of document types, a solution was required to automate the process to meet set deadlines for legacy content.

PCI-DSS was developed to encourage and enhance cardholder data security and facilitate the broad adoption of consistent data security measures globally. Its 12 requirements include "to protect stored cardholder data" and "to restrict access to cardholder data by business need-to-know." Failure to comply with all requirements could lead to hefty fines and/or penalties, depending on the payment card brand involved. Also, because the agency is high-profile, non-compliance would have been damaging to its reputation.

The customer examined two enterprise-class OpenText™ products that could perform the high-quality Optical Character Recognition (OCR) that would make classification and redaction—the complete removal of all the credit-card information—possible. The solution needed to be able to integrate with the OpenText™ Content Suite Platform and needed to pull all suspect content from Content Suite and return redacted content back to the repository seamlessly. After proof-of-concept sessions test results were analyzed, the agency ultimately decided on OpenText™ Redact-It™ Enterprise with OCR. OCR enables the images and scanned documents to be text searchable, while the Redact-It solution had the best capabilities for finding and redacting the right data.

INDUSTRY

- *Government*

CUSTOMER

- *Government agency*

CHALLENGES

- *Exposure of credit-card information, risking data breaches*
- *Non-compliance with PCI-DSS, potentially leading to hefty fines*
- *Differing forms, formats, languages, and non-searchable, low-resolution images, as well as a lack of metadata*
- *Massive volume of documents collected over the years*

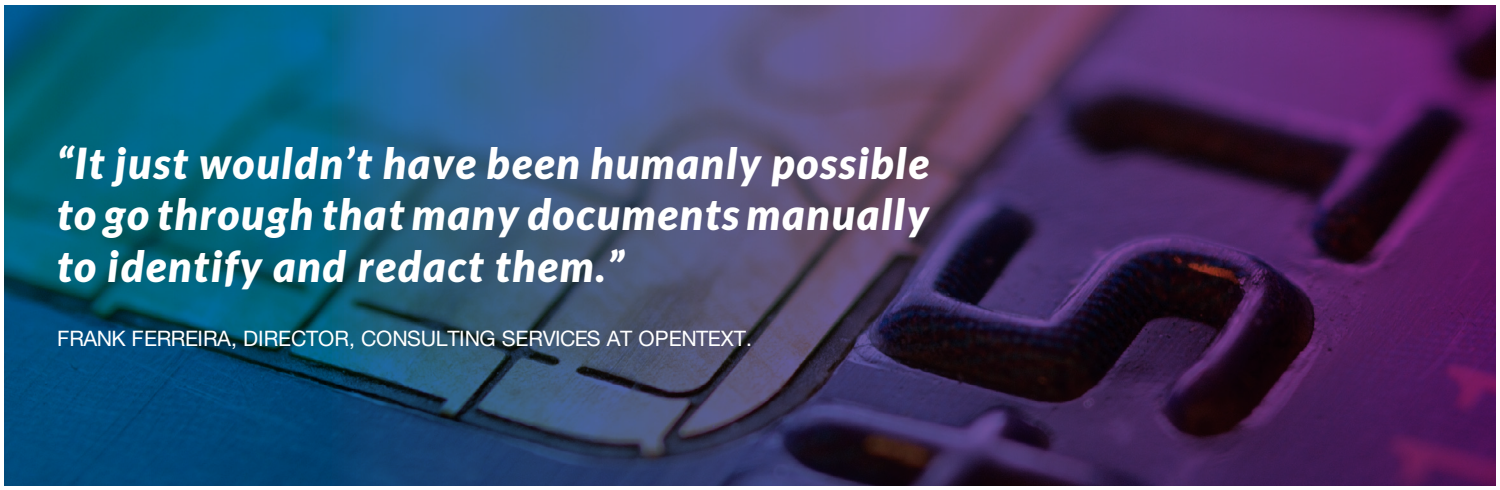
SOLUTION

- *OpenText™ Redact-It™ Enterprise*

BENEFITS

- *Exceeded client requirements of 80% success rate*
- *Helped client achieve PCI compliance for legacy content*
- *Achieved 87% success rate on 2.3 million documents processed*
- *Reduced project duration by years*

opentext™



Identify the Right Documents

After initial classification and rationalization of suspect content, approximately 2 million of the documents stored in OpenText™ Content Server were identified as possibly containing credit-card data. To make matters even more challenging, the documents came in many different varieties comprising various languages, format types, and versions—with no metadata to classify each document. Many of the documents even contained multiple forms scanned together.

“It just wouldn’t have been humanly possible to go through that many documents manually to identify and redact them,” says Frank Ferreira, Director, Consulting Services at OpenText.

Redacting the Right Way

Redaction was performed in two ways. The first pass was done using Redact-It’s predefined macros for credit-card number patterns. Some initial testing found many false positives, with some forms containing other numbers in patterns similar to credit-card numbers. The team modified the search string to account for the other types of numbers and was quickly able to exceed the targeted success rate.

Many of the forms also contained data that were of low-quality resolution (less than 300 dots per inch), askew, or had the wrong orientation, and couldn’t be identified by OCR. For those forms, the client helped the OpenText team determine search strings of other text to help identify and classify each page of the documents as a particular type of form. This was then followed by using zone-based (or coordinate-based) redaction templates to redact the areas that were known to contain the credit-card information for each form type. “Excellent results were achieved with respect to form recognition by page, which meant that zonal redaction could be reliably applied once forms were identified,” says Mary-Lou Baird, OpenText Project Manager. Once redaction was completed in PDF renditions of the originals, the redacted versions were saved back into OpenText™ Content Server and the older, pre-redacted versions were purged—to further restrict access to cardholder data and therefore ensure PCI compliance.

Overall, OCR was conducted on 2 million documents (or almost 9 million pages). Of those, 8 percent were identified as requiring redaction. Despite some delays due to infrastructure provisioning, the entire project took less than nine months to complete. And throughout the process, the OpenText team supplied regular and frequent status updates to help the agency demonstrate to the audit team that it was making reasonable attempts to remediate the situation. Ferreira adds, “The customer was very pleased with the results they achieved, and said they were better results than they even thought were capable.”

www.opentext.com/contact