Merit Award Application

Jurisdiction: Texas

Submitted by: Diane Silver, PMP, Project Manager, Information Technology Section, Office of the Secretary of State - Texas

Date Submitted: 3/20/08

Brief Description:

Project Title: **Corporations and UCC Filing Document Image Redaction Project – Texas Office of the Secretary of State.**

Summary:
During calendar year 2007, the Texas Office of the Secretary of State undertook a project to redact all social security numbers from images of Corporations and UCC Filing Documents viewable from their web portal, SOSDirect. The entire project (from information gathering to completion) was done in seven months and resulted in the redaction and replacement of over 1,025,306 images without disruption of service.

Detailed Description:

In April, 2007 the Corporations and UCC Sections of the Texas Office of the Secretary of State began a project to redact social security numbers from images of filing documents that were available for online viewing from the agency web portal application SOSDirect. Although there was no legal requirement to do so, the agency had the authority to redact and elected to do so voluntarily. The project was funded from the sale of public records (as opposed to appropriated funds). Several other Business and Public Filings automation projects were temporarily put on hold in order to make this project the top priority.

The first order of business was for Corporations and UCC to decide which documents, and which fields within the documents might contain social security numbers. After some research, the following wording was included in the RFO and provides a description. "Images to be reviewed will consist of two types: Uniform Commercial Code (UCC) financing statements and public information reports for corporations and limited liability companies. The UCC financing statements are filed using various forms. The forms have changed over time. The social security number may appear in an identified field on a form or as part of the text of the collateral description or filer reference fields.   The public information report is a standard form with an identified field for SSN. However, the SSN may also appear on an attachment to the standard form. The format of the form changed in 2004 to eliminate the SSN field, but some forms with 2004 file dates utilized the old format. Generally, the information is typed, but some handwritten data is presented on some forms."

The following briefly outlines the approach that was taken:

__Redaction software and services selection:__
A team comprised of key resources from the UCC/Corporations and IT sections of the agency began researching software currently available for scanning and redacting documents. It was decided that due to the large estimated number of documents requiring review for redaction, the processing (review for redaction) of existing images of non-electronic filing documents would need to be out-sourced. Information gathering sessions were set up with several vendors who had experience using redaction (rules based) software and providing this service. After weighing the options, a decision was made to award a contract to *mobilis* (the company that had been providing services in support of the agency's ViewStar imaging and workflow system for several years that was partnered with *Mentis Technology*, a company experienced in providing redaction services (with subcontractor *Doc-2-e-file)* using *aiRedact* software. Having extensive knowledge of the imaging and workflow system made *mobilis* uniquely qualified to make changes to existing processes as well as to provide the needed integration of the *aiRedact* software with the existing system.

__Redaction of Non-Electronic Filings:__
Once the contract was awarded, the internal IT developers wrote programs to extract approximately 4,130,095 documents, or a total of 6.3 million pages (Tiff images) and loaded them along with corresponding index files to encrypted external drives for transfer to *Mentis Technology.* UCC and Corporations worked with *mobilis/Mentis Technology* to document requirements to be used in building the rules engine portion of the redaction software to allow it to search required fields for SSN pattern matching. A representative sample of 100,000 images was first provided to *mobilis/Mentis Technology* for use in developing the rules engine and testing. Based on findings that filings from certain submitters contained SSNs embedded as part of other identification or tracking numbers that appeared on the filing, refinements in the rules were made. After testing was completed, the remainder of the images were shipped for processing.

The IT developers also wrote programs and revised existing workflow processes to so that redacted TIFF images, when returned, could be imported into the system. The existing images in PDF format (used for viewing from the web) were deleted and new PDFs generated. The non-redacted TIFF images were replaced by those that had been redacted, but the non-redacted originals were retained for viewing (if needed) by internal users only.

Only the redacted images were returned by *Mentis Technology* to SOS IT on the encrypted external drives. The first shipment was received the first week in August. Approximately 16% of the 6.3 million images that were sent to *Mentis Technology* required redaction. Over the course of the next two months the IT team worked to import 1,000,306 redacted images into the workflow system-- a process that included deleting and regenerating the non-redacted images in PDF format. (261,288 of the images were PIR documents, and the remaining 739,018 were images of UCC Filings.) Work was completed by the first week of October.

Electronic (EDI and Web) Filings:
Applications to handle the redaction of existing and incoming UCC electronic filings (EDI and web filings) were also developed by in-house programmers. One application provided for searching selected fields stored in database tables to identify the existing electronic filings that needed to be pulled for review and provided an interface for viewing and redaction if required. (Images of filings submitted electronically are generated dynamically using the data input in the filings process.) Changes were made to some of the existing imaging and workflow processes to allow the original (non-redacted) data and or document to be retained for in-house retrieval only while replacing the copy of the image available for public viewing on SOSDirect. A second application was developed to scan designated data fields of all in-coming electronic filings (EDI and web) and place candidates for review in a queue for manual review and redaction if necessary.

Starting the end of July, existing electronic UCC filings, identified as candidates for review for redaction, were placed in a work queue to be processed. Some refinements were made to the extract program to pull all filings submitted by certain secured parties once it was determined that the SSN was embedded within other tracking numbers appearing on the UCC form. Over the course of the next four weeks approximately 120,000 electronic filings were reviewed, 25,000 of which were redacted. At the same time, the application used to screen incoming EDI and web filings was moved to production and internal (online) reports put in place to alert UCC document processors of incoming web or EDI filings, identified as candidates for redaction that required review.

Going Forward UCC Redaction Processing:
Going forward, Corporations filing documents required no redaction as only the old PIR filing documents contained SSN's. However, incoming UCC paper filing documents required continued scanning for redaction candidates. To meet that requirement, the *aiRedact* software was loaded to a special server and modifications made to the existing scan review processes for paper documents so that candidates for redaction were brought into a queue for review. Changes to workflow again made provision for retaining the non-redacted original for internal viewing only but generated web-viewable (PDF) images from the redacted image.

Additional Steps:
In addition to redaction of existing documents and integration of new automation into the UCC filing processes, actions were taken to reduce the number of incoming UCC filings that contained social security numbers. Notices were sent to secured parties with over 10 filings requiring redaction. Information about the redaction project was posted on the SOSDirect and agency websites and articles were sent for publication in various organization newsletters.

UCC document processors kept track of submitters that filed documents containing this sensitive data and contacted them to request that they discontinue this practice. This follow-up procedure was done for both electronic and non-electronic filers and has substantially reduced the number of candidates for redaction review going forward.

Cost-Benefit:

Despite the fact that there was no legislative or judicial requirement that the agency remove personal information (SSN's in this case) from documents displayed on the web, it is felt that being proactive in removing sensitive information had the benefit of addressing concerns that might be raised about data obtained from images displayed on SOSDirect being used for purposes of identity theft.

The cost to the agency for purchasing the *aiRedact* software and associated services provided by the vendors and subcontractors was $264,776. This cost was more than offset by a much more significant cost-avoidance. Since the agency began providing support through SOSDirect for submitting filings and placing orders for copies in August 2001, the percentage of orders for UCC copies (including searches for documents) that is handled strictly by automated processes has risen to 99%. The impact on the UCC division has been significant in that employees have been re-deployed to other sections of the agency or been assigned different tasks to better balance the workload. The fact that offices in several other states have been forced (due to external pressures) to take similar web sites down until sensitive information was redacted had become a very real concern. The UCC division is no longer staffed to handle the workload if it became necessary to resume the mostly abandoned, manual paper-bound processes. The cost to the agency of staffing the UCC division back up to pre-2001 levels would have far outweighed the cost of the redaction project. Or, if the needed staff increase was not possible, the UCC section would have been unable to meet its performance measures (turn around time) with resulting negative financial impact to the agency.

Lessons Learned:
As is generally the case, there is learning that goes on during the process of putting new automated processes into place. Adjustments and fine tuning had to be done as filings were being examined to improve the accuracy rates of identifying SSNs. However, as a collaborative effort involving both the internal team with representatives from Corporation/UCC/IT and Administrative Services as well as the vendor teams, we were able to complete this large and complex project in a little over seven months. Being pro-active and doing the project, while expensive both in terms of actual dollar cost as well as in negative impact on other automation initiatives, allowed us to alleviate concerns over any potential prospect of being forced to take down our online SOSDirect system.