



managing input across the enterprise

Industry: Government

Captiva Software Helps New Jersey Division of Revenue Cut Tax Return Processing Costs

New Jersey State Treasurer James Di Eleuterio faced the challenge of his department receiving approximately 3.8 million personal income tax returns in a single year. His 300 data entry operators would have to key return data into their aging CMC/Unisys terminals that were barely handling the existing volume. He decided it was time to modernize New Jersey's revenue collection operation by automating its tax return processing system.

Di Eleuterio selected systems integrator Northrop Grumman to implement a LAN-based forms automation solution based on intelligent character recognition (ICR) software from Captiva Software Corporation. The New Jersey Document Processing System, called DPS, was already using Captiva's ICR software to process 95% of its tax returns, which consist of 1040 individual tax forms and HR (Homestead Rebate) 1040 forms.

Today, the Captiva solution automates approximately 95% of New Jersey's tax return processing, while 5 percent of the returns are diverted for manual data entry. Nick Manocchio, Chief of Technology Services at the Division of Revenue said, "Our job is to correctly capture what the taxpayer puts on the form in the fastest, lowest cost manner."

SECRETS OF SUCCESS

According to Joe Roose, Assistant Director at the New Jersey Division of Revenue, the DPS project succeeded because New Jersey was able to optimize Captiva's ICR technology. First came the issue of form design, since New Jersey's forms were not ICR-friendly. Roose had special marks printed on the forms to improve scanner registration and data fields were repositioned for easier

location. His department changed the envelope size to allow returns to be folded only once, resulting in faster document preparation and better scanning results.

The division also added graphic symbols to the returns to better audit their work as forms went through various processing stages. Roose reported, "We printed a different format barcode on each document within the tax return, which we use to check the completeness of the tax return and the order of documents."

REENGINEERING THE BUSINESS PROCESS

To maximize its return on the state-of-the-art scanning and recognition technology, the New Jersey Division of Revenue reengineered some of its business processes. "We learned as we went along and made a lot of changes to improve our processes. Some were small things, some large," Roose recounts.

Roose outlined the procedures currently employed by the division:

1. Staff opens envelopes and sorts forms into three categories: hand printed, machine printed and rebate applications.
2. Attached checks are removed and processed to ensure timely receipt of funds coming to the state.
3. Returns are batched by barcoded batch control sheets placed before each return. To reduce labor, barcodes are printed at the top, bottom and sides of the batch control sheet to allow document preparation worker to very quickly insert the form.
4. The four-sided hand printed submissions (with an average of four to five attachments such as W2s) are scanned.



5. The 10 to 11-page machine printed submissions are scanned.
6. Automated processing is handled by Kodak IL923 high speed paper scanners controlled by a Sun workstation. Scanning is done in portrait mode to ensure a smoother flow of W2 forms, which are of various sizes and printing types.
7. Barcoded batch controls are recognized automatically and define the start and stop of each return.
8. Barcodes on each page perform a check for completeness and a check on whether the paper was properly registered.

SPEED AND ACCURATE THROUGHPUT

Once forms are scanned, images are routed in batches of 50 to FormWare's ICR engine, which is controlled by a UNIX-based E4000 Sun server. Data recognition is handled by 14 Pentium PCs interconnected through Ethernet cards, running 24 hours a day during peak processing times. At an average of 350 to 400 characters per return, DPS achieves a recognition throughput of approximately 1 million characters per day per PC.

Manocchio reports that field accuracy rates on the hand printed forms are 90% to 92%, while the average is 97% to 98.7% for machine printed forms. With average data fields of seven to eight characters, FormWare reaches an impressive character-level accuracy of over 98% on handprint and over 99.6% on machine print. Human accuracy is 99.5% per character.

Accurate throughput is a product of more than careful scanning and excellent ICR. Also contributing equally are rules and procedures that tightly define the characteristics of a given data field and use ICR results to cross check fields against each other whenever possible. For this purpose, Captiva developed special programming tools to allow users to easily set up complex data validation routines.

Verification of the recognized data is managed on 33 workstations, then reconfigured into the large record format that historically has been output from manual key entry stations. It is merged with exception items that have been keyed on ASCII-based screens and sent to New Jersey's mainframe for processing. After capture, the images are reformatted and passed to a FileNET Imaging System for storage and retrieval, so after scanning, paper can be boxed up, shrink-wrapped and sent offsite for storage. And because all data for a customer inquiry is available on the Division's new imaging system, the Revenue Department has never needed to access the paper storage area.



© 2002 Captiva Software Corporation. Captiva, FormWare, InputAccel, PixTools, ISIS, ClaimPack and InvoicePack are trademarks of Captiva Software Corporation. The other company and product names mentioned may be trademarks or registered trademarks of the related companies. All specifications are subject to change without notice.