



FTA Awards Nomination/Entry Form

Person who led this effort or project

Name Keith Gross

Agency Name WI Dept of Revenue

Phone Number (608) 264-6865

Email Keith.Gross@wisconsin.gov

About your program, idea, or project

Name your program, idea, or project: Data Capture Modernization

What is the problem that you wanted to solve?

Wisconsin Department of Revenue has been using scanning and automatic data capture for tax documents since the early 90s. In the last decade we've expanded this to include all inbound mail, making this central to our operations. But the software we've been using has not evolved with the times. We still use IBM Intelligent Forms Processing, which we purchased in 1993. Though cutting-edge for its time and the support from the development group is still excellent, the product is not being upgraded and is now far behind current trends in data capture.

Fifteen years ago, we bought in Kofax at great expense to modernize our data capture system. The initial contractor delivered a setup for scanning Informational Returns which worked well enough but after repeated attempts to apply it elsewhere we concluded the software was too complicated to setup, difficult to keep running and the vendor seemed to be moving in a different direction.

Last year, we decided to pick a new direction for both scanning and data verification. Something that would really boost productivity in both areas and offer the opportunity of handling some scans without human review.

Who was involved in addressing the problem?

The initial work to develop a project proposal was done within IT. As the project rollout began, it was quickly realized how much impact even minor differences in the capture system could have, so the testing and review effort for further phases was expanded to include processing system developers and end users.

How did they go about finding a solution?

It was suggested that we look at new AI-based solutions that make form setup easier. We decided to run a pilot with one form using Google's Document AI and a low code application from a contractor experienced in working with Google cloud. We worked with them for five months to train a form and test the application.

In the end, we felt that Document AI had great potential. The OCR was superior to either of our current systems and the setup and training were simple and straightforward. But the low code application environment couldn't really deliver a particularly good data verification environment for human in the loop. So, the decision was to move forward with an in-house developed application built on our standard environment and simply use Document AI for the OCR and data extraction.

Describe the outcome. What is the new idea, approach, program, or activity?

At the end of the pilot, we decided the low code environment didn't produce a good data capture

experience and the cost of running it was quite high, but Document AI seemed to have real potential. It could produce very good results with very low form setup effort and could be adapted to minor changes in the forms without requiring a significant retraining effort. This could greatly simplify our support for vendor-produced forms saving significant staff time every year.

What has changed since this was implemented? How have your operations improved? Include any data, analytics or metrics that would show the value of your program. Don't forget management advantages such as improved morale.

We just completed the first season of Informational Return processing. Productivity stats for people doing data verification on these showed a 300% increase in documents per hour. Due to budget constraints, we had to reduce our temporary workforce and decided not to hire employees to work on Informational Returns. This decision along with the productivity gains, resulted in labor cost savings of over 56%. Full-time staff could handle their existing workload and informational returns new system despite staff needing to learn the new system and deal with early glitches. We expect even better results next season as we do further Document AI training and refine the data verification interface based on staff feedback.

Based on that experience we project once we're fully switched over that the cost to run the system will be between 5-15% of what the older systems were costing us. And since it's an in-house development system we have full visibility into how the system works and support can easily be handled by our existing development group.

Data verification staff really like the new system. The improved accuracy of the OCR greatly reduces the number of adjustments the data validator must make. The feel of the work is more of read and review. This creates opportunities for more automation based on high confidence where we may not even need to have humans review certain data fields if Document AI provides a high enough confidence in its accuracy.

Is there any component of your program that makes it workable only in your state or city?

Another agency with a similar tax type or this kind of problem could adapt or adopt this idea.

Is this an in-house project, or did you partner with an outside vendor or service-provider?

Our idea, but we used a publicly available software or service for at least part of the implementation

Additional information or comments about your usage of outside vendors or service providers.

The application is 100% in-house developed but makes use of Google's cloud AI solution (Document AI) for the OCR and data extraction.

What comes next – will you be adding to your program, rolling it out more widely, trying additional approaches?

We are currently in the process of converting all our existing data capture to this new system. First year addressed all our Information Returns (W2s and 1099s) as well as the scanning of inbound mail and is expected to be complete by June 2025. All told this involves between 400,000 and 500,000 pages of data capture and another 1,000,000 pages of related imaged pages.

Over the next two years we'll convert the scanning of all our paper tax forms. This involves about 250,000 filings composed of over seven million pages of data capture.

Additional Optional Materials