

Crowdsourcing supported questionnaire processing

MPS IntelliVector

MPS IntelliVector is an award-winning data extraction and workforce transformation solution, especially tailored for the financial, insurance and government sector. MPS IntelliVector transforms traditional paper document and forms processing by utilizing multi-level automation, microtask-based processing and secure enterprise crowdsourcing, to radically improve all major aspects of paper processing. The implementation of MPS IntelliVector results in: cheaper and faster processing, reduced human involvement, and exceptionally high data confidentiality and accuracy even in case of outsourcing or crowdsourcing!

About MPS

Multipass Solutions is an independent software vendor and consultancy with broad experience in ECM. We offer flexible award winning software solutions for document and forms processing, secure data entry, system integrations and business data clean-up to raise business efficiency, lower operating costs and create a unified, secure business IT environment.

Customer challenge

The goal of the project was to set up a solution and to process ~4000 highly confidential hand-filled questionnaires (~48000 pages) each containing 280 fields (check boxes and fields containing handwritten numbers and text) in less than one month. Our client had limited internal resources for the project, so they were looking for a solution, that can both, automate the processing without decreasing the data accuracy and allow to utilize external resources for the data entry, without compromising the confidentiality of the processed data.

Main requirements:

- Configure the scanners for the digitization
- Automate the processing of app. 4000 questionnaires (48 000 pages, app. 1,1M field)
- Guaranteeing full confidentiality of the data
- Provide a simplified web interface for the manual data entry and quality control
- Enable the usage of external resources without compromising data confidentiality

The solution

After an extended vendor evaluation period, MPS IntelliVector was chosen due to its unique capabilities, which allow to automate and speed up processing, while the microtask-based approach allows to utilize cheaper, external workforce for the manual parts of the processing, yet still guarantee 100% data confidentiality.

After client employees scanned all the questionnaires, using pre-configured scanners, MPS IntelliVector automatically recognized the questionnaire types and extracted the parts that required processing. During the extraction the documents were broken down into small individual fields, or microtasks, which were then processed by a combination of automated and manual recognition.

- **in case of check boxes:** using only automated OMR (Optical Mark Recognition)
- **in case of handwritten numbers:** cross-checking results of automated ICR (Intelligent Character Recognition) and microtask-based manual data validation
- **in case of free, multi-line handwritten text:** usually for handwritten text MPS IntelliVector uses a combination of ICR and manual microtask-based data validation (similarly to one

used for the handwritten numbers), but in this case traditional manual data entry was used.*

** (MPS IntelliVector uses some of the best recognition engines on the market, however there is no automated recognition technology that can guarantee 100% accuracy for this type of text, so in this case manual data entry was more rational in terms of time and cost.)*

Microtask-based data entry

In contrast to the traditional approach offered by other vendors, where data entry users type in the data looking at full-page images of the documents, MPS IntelliVector recognizes the type of the incoming documents, breaks them up into small, individual, microtasks for the data entry. This way, data entry users see only small, anonymous pieces of information, without their original context, so the confidentiality of the initial data is fully preserved. This approach allows to securely utilize external, even crowdsourced workforce for data entry or data validation even in case of highly confidential data.

Each microtask (in this case a field containing multi-line handwritten text) was typed in by two data entry users and only if the two results didn't match, then it was sent to a third data entry user. Microtasks went to quality control only if there were no matching results out of three data entries. A total of 7 internal users and 10 external part-time users were doing the data entry. They were complemented by 4 quality controllers.

Results

- Project finished in one month
- **48 000 pages scanned and processed** (3858 questionnaires)
- **1 106 386 fields (microtasks) processed**
- **In 35%** of cases replacing manual processing with **fully automated processing**
- **In the remaining 65%** of the cases achieving:
 - **200% faster processing** in case of fields containing numbers
 - **20% faster processing** in case of fields containing text
- **Huge cost savings** due to the involvement of external resources for the data entry
- **100% data confidentiality** throughout the data entry (even with crowdsourcing)