Industry Paper

# Historic Document Conversion

Artificial Intelligence & Image Analysis
www.A2iA.com

*Unlock the information trapped in historic archives. Locate, extract, and index data for search and reporting.*

Award-winning and driven by research and development, A2iA, Artificial Intelligence and Image Analysis, is recognized as the worldwide leading developer of handwritten and machine printed text recognition, information extraction and intelligent document classification. By operating the world's largest research center dedicated to extracting information from everyday paper documents, A2iA has been proactive in its developments to meet the needs of the market place, and producing a visible ROI for its users for more than 25 years.

## DID YOU KNOW

**Industry analysts have explained that the best use for A2iA's Proprietary Intelligent Word Recognition, IWR, technology is to eliminate a high percentage of the manual entry of handwritten data that otherwise could be keyed only by humans**

The digitization of these archives has been identified as a necessity to preserve their integrity and overcome the scarcity of space, as well as to make them available for public view once specific information has been removed, or declassified. But how can the content of these records be accessed once they are imaged? How can specific information be found quickly, and intelligence gained, once in digital format?

Difficult documents to automate. While ordinary documents today can be classified and indexed with relative ease, automating this process with hundred-year-old images in old-fashioned writing styles can be quite complex. These archives have characteristics that make them challenging to work with, and make them unsuitable for existing document management solutions:

- The documents are typically handwritten.
- The writing-style is old-fashioned and varied.
- The images are found in inconsistent formats and do not follow a predefined template.
- The images are of poor quality because of their age and their storage conditions.

*a2ia DocumentReader*, a toolkit designed specifically for locating, extracting and indexing complex data including cursive handwriting, is proven around the globe to eliminate many of the pain points associated with unlocking the data trapped inside complex archives that contain old-fashioned writing. By analyzing the scanned images on a holistic level, looking at both their geometry and content, *a2ia DocumentReader* gains an understanding of the documents being interrogate. The software then performs a literal transcription of the handwritten and/or typewritten areas. In the following steps, it extracts key words or expressions that can be used for redaction, declassification, and indexing.



Americas: +1 917 237 0390 | EMEA +33 0 44 42 00 80
marketing@a2ia.com

## *a2ia DocumentReader™*

- Locates and extracts even the most complex information, including handwriting, both modern and old-fashioned styles.
- Indexes complex documents (those that are old or unstructured, or that occur in varied formats).
- Declassifies confidential information and makes the documents available for public search and view.
- Allows information to be searchable and reportable, even from century-old archives.
- Extracts all types of information, whether printed or handwritten, alpha, numeric, or alpha-numeric handled solely by the customer and financial institution. Because of its small footprint, *a2ia CheckReader* leaves computer resources available for other applications and operations.

## Advanced Document Classification and Complex Data Extraction

By going further than common OCR and ICR engines that are unable to recognize cursive handwriting, and other classification engines that use only layout to route a document, *a2ia DocumentReader* automates the entire incoming workflow. This new generation of technology allows all-types of paper documents and archives to be processed automatically, with little to no human intervention.

All proprietary to A2iA, *a2ia DocumentReader* contains no third-party technology. And, by operating the world's largest research center dedicated to extracting information from paper, with a focus on handwritten and unstructured contents, A2iA's R&D team is able to adapt the technology to meet the demands of the users.

## *A2iA Processes 1.3 Million WWI Archives, Saving Nearly 70% in Data-Entry Costs*

To give every citizen the opportunity to access a piece of history, the French Ministry of Defense partnered with A2iA to process 1.3 million files of "Those Who Died for France" during the war of 1914-1918. These documents were handwritten in an old-fashioned style, and those that contained medical history were to remain confidential.

As the only technology that can read old handwritten archives, *a2ia DocumentReader* automatically located and extracted records that contained medical data, protecting their confidential character. By declassifying these documents, A2iA ensured they were ready for public view.

The Ministry of Defense insisted on high-quality detection with a maximum rate of error of 0.5%. A2iA **exceeded** the ministry's expectations with a 0.2% rate of substitution for the processed files and **data input savings of nearly 70%.**



Americas: +1 917 237 0390 | EMEA +33 0 44 42 00 80
marketing@a2ia.com

## Complex archives contain important information that can help uncover the past and even unlock a piece of history

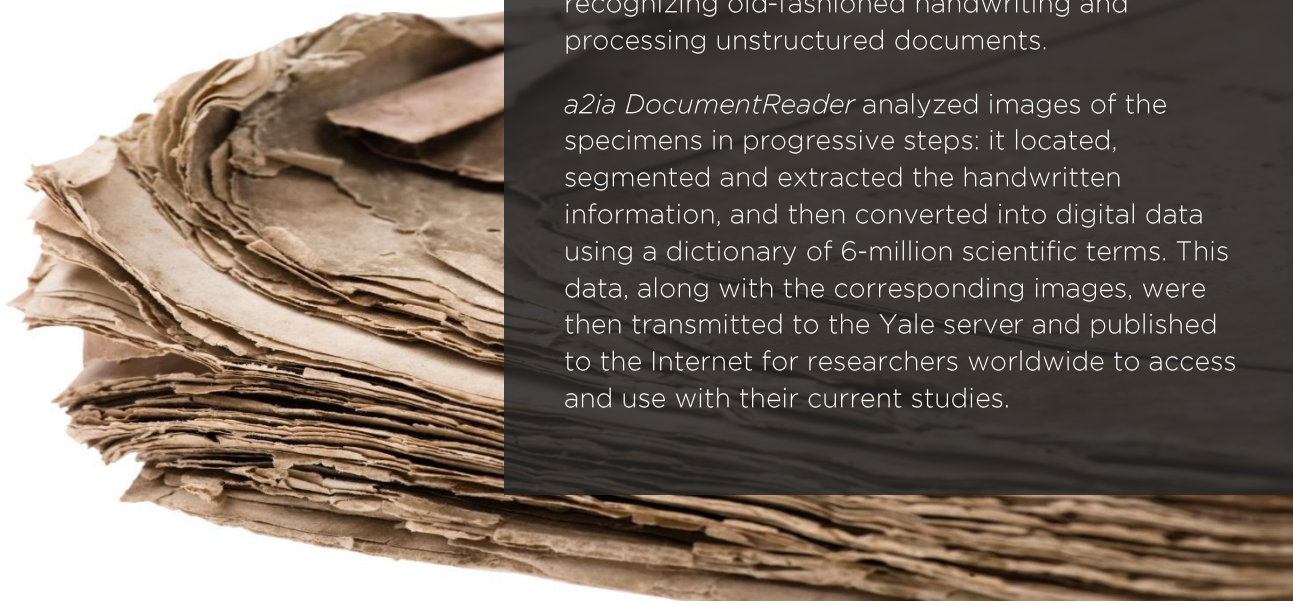
A2iA's proprietary IWR, Intelligent Word Recognition, technology lies at the core of the company's software toolkits. IWR recognizes entire handwritten words similarly to that of a human, matching them to a user-defined dictionary. Instead of looking at words letter-by-letter, IWR performs a deeper analysis and allows users to recognize, classify and index all cursive and complex documents with the same level of flexibility as printed or digital data.

## *a2ia DocumentReader™* Makes Yale University's Botanical Archives Available Worldwide

To allow scientists from around the world access its collection of herbaria, Yale University sought to upload several thousand botanical specimens to the Internet. The task was far from easy: some documents were almost 160-years old, most of them handwritten, and their format and placement of information varied.

Yale University chose *a2ia DocumentReader* to process the handwritten information from the specimens, as it is the only software capable of recognizing old-fashioned handwriting and processing unstructured documents.

*a2ia DocumentReader* analyzed images of the specimens in progressive steps: it located, segmented and extracted the handwritten information, and then converted into digital data using a dictionary of 6-million scientific terms. This data, along with the corresponding images, were then transmitted to the Yale server and published to the Internet for researchers worldwide to access and use with their current studies.



a2ia | Image Processing | Document Classification | Data Extraction



Americas: +1 917 237 0390 | EMEA +33 0 44 42 00 80
marketing@a2ia.com