

Global Financial Organization Creates Digital Treasure Trove of Key Documents

INNODATA ISOGEN CASE STUDY

Innodata Isogen Digitizes Extensive Collection of Legacy Publications

CHALLENGE

A global financial organization charged with overseeing and monitoring foreign exchange rates, balance of payments and government economic models has consistently recorded its past and ongoing activities and viewpoints in an extensive publications library. These publications not only serve as documents of record, but also help financial analysts and other observers shape government economic policy in countries worldwide.

In the past, many of these documents existed only on paper. However, the organization recently decided to maximize the value of this content repository by converting 25 years' worth of publications to an electronic format. This would enable the content to be searched and retrieved on-line, providing greater ease of use and vastly expanding the potential audience.

This was hardly a straight-forward digitization task. Many of these documents had been printed in different formats – from hard and soft cover books to document-style reports. Some 1,500 of the 7,000 publications were bound books, printed in a variety of formats – 6" x 9", 8.5" x 11", and 10" x 7". Ranging in length from 60 to as many as 85 pages, the multi-lingual publications included photos, charts, tables, and graphs. In addition, the organization required each of the 7,000 paper documents to be generated in six different electronic files.

SOLUTION

To convert the diverse documents and their content into a searchable electronic format, the organization turned to Innodata Isogen, which has carried out some of the world's largest data conversion efforts.

Accuracy is of paramount importance to the agency, which commands a singular position as a leading monitor of global financial policies. Innodata Isogen's proven capabilities for achieving accuracy rates beyond the base requirement of 99.95% or higher was a key factor in the agency's final decision.



CHALLENGE

Digitize collection of 7,000 legacy publications from paper to digital format, at lowest possible cost

SOLUTION

OCR and scan these publications with 99.95 percent accuracy

BENEFITS

Agency now has virtual content treasury of its key documents, easing use for researchers and other organizations

IMPLEMENTATION

Innodata Isogen began by assessing the appropriate technologies required to handle the volume and complexity of the source volumes. Using proprietary imaging technologies and processes, Innodata Isogen began processing the body text of the documents, along with the photographs, charts and other graphic elements. In all, Innodata Isogen was required to provide six output files for each document:

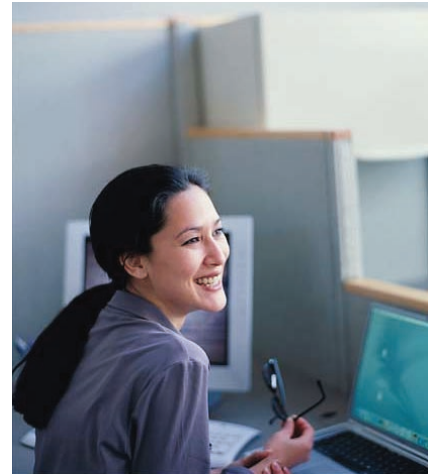
- A searchable PDF file of the document's interior content for printing
- A searchable PDF file of the cover for printing
- A Web quality, merged PDF file of the cover and interior of the document for Web viewing
- A 24-bit .TIFF file, in color, of the cover
- A print-quality PDF file of the entire publication for printing
- A 300 dpi .TIFF file of the cover

To improve search capabilities, the team also set up a disciplined process for extracting targeted metadata. The six files for each publication were assigned a common file name structure with different characters to identify which type of file it was. Each publication has a base identifier that depicts information about its type, as well as the naming schema.

Both Innodata Isogen and the customer conducted strict quality checks, ranging from basic tasks such as checking proper pagination, order, and image quality for PDF and .TIFF files. The images needed to be clear, free of extraneous marks, and positioned properly on the page.

IMPACT

When the project was finished, the agency rapidly began making new digital content available to economic officials, journalists, students and others who rely on this unique content collection. As a result, the organization can now showcase its thought leadership capabilities and economic leadership through this new dynamic content repository.



The organization can now showcase its thought leadership capabilities and economic leadership through this new dynamic content repository

FIND OUT MORE

Learn how our solutions can help your company achieve success, talk to an Innodata Isogen representative today

CONTACT US

Innodata Isogen

Three University Plaza

Hackensack NJ 07601

(201) 371-2828

www.innodata-isogen.com