

### NYSDEC Improves OCR Efficiency by 85% with PdfCompressor

#### Business Challenge

The New York State Department of Environmental Conservation (NYSDEC) needed to find all relevant documents utilizing a search query based on text inside and across all files. These documents included complex PDF files - electronic, image, and hybrid documents. Generic OCR technology required the PDFs to be rendered to images and then re-OCR'd, losing critical information such as bookmarks, tables of contents and hyperlinks. The state government agency needed to ensure that all text within a document was indexed as well as preserve key PDF attributes in order to maintain the integrity of their vital documents.

#### Solution

In order to index all documents as well as preserve the information contained in the agency's documents, NYSDEC turned to CVISION's PdfCompressor. PdfCompressor was used to OCR all of the agency's documents in their repository (OpenText Documentum®). With this solution in place, all NYSDEC content were made fully text-searchable with no loss of information. Even further, the OCR'd documents were fully indexed once placed back into their Documentum environment.

#### Results

CVISION's PdfCompressor was able to solve the state government agency's problems of information loss caused by the previous, generic OCR process while creating fully indexed files. As experts in PDF technology, CVISION built a solution that would avoid the need to render PDFs to image before processing, thus preventing any loss of key attributes such as bookmarks, metadata, PDF/A, hyperlinks, etc. Without unnecessary rendering, PdfCompressor was also able to reduce processing time to just 15% of the previous, generic OCR process. This solution helped NYSDEC preserve all of its vital information while greatly improving processing speed. In addition, PdfCompressor was implemented quickly and easily into the existing workflow.

#### Objective

#### Benefits Achieved

<ul style="list-style-type: none"> <li>Eliminate information loss caused by the previous, generic OCR process</li> </ul>		<p>CVISION's PdfCompressor was able to OCR the agency's documents while preserving key attributes such as bookmarks, metadata, PDF/A, hyperlinks, etc.</p>
<ul style="list-style-type: none"> <li>OCR all of the agency's documents, including complex PDFs</li> </ul>		<p>PdfCompressor made the NYSDEC repository fully text-searchable, with no loss of information.</p>
<ul style="list-style-type: none"> <li>Avoid unnecessary rendering of PDFs to image before processing</li> </ul>		<p>By avoiding unnecessary rendering, PdfCompressor was able to reduce time to OCR to just <b>15%</b> of the previous, generic OCR process.</p>
<ul style="list-style-type: none"> <li>Integrate seamlessly into the existing workflow environmentprocessing</li> </ul>		<p>CVISION's PdfCompressor integrated smoothly into both the Documentum and Captiva environments.</p>