

Aspen Systems Corporation
Noah Katz
Development Analyst
Adobe Certified Expert: Acrobat 4,5,6

Business case:

Background: Aspen Systems is producing documents for search and review purposes in an investigation. The TIFF images, full text OCR, and field data abstracted from the documents (title, author, addressee, copyee, documenttype, date, etc.) are being loaded into a Concordance database. There is also a delivery of a subset of the documents, for use in an independent – even portable – system. These are provided in PDF format, and each DVD is indexed with Acrobat Catalog to allow searching the collections with the free Adobe Reader. Some of the fielded data is placed in the PDF Document Properties fields, both standard (Title, Author), and custom (Addressee, Copyee, DocType, DocNo, Date).

The data, from SQL database, is ‘injected’ into the PDF files with Apago PDF Enhancer as part of the enhance stage of the workflow processing. The specific data for each document is merged with a template configuration file (.PEC, an XML file specifying the settings for PDF Enhancer) already populated with the specified image compression settings and password security settings.

For quality control, random PDF files are inspected. I placed an AcroButton on those workstations to expose the 7 data fields - with 1 click and no clutter – to the user (available for Reader or Acrobat). In the event of exception handling for those files that failed processing through PDF Enhancer (usually timed out due to file size of thousands of Searchable Image pages), I put another button on the workstations (full Acrobat only) that prompted the user to input the data for each of those fields (and the custom fields were created by JavaScript on the fly as part of this data entry process). The user avoided going to the Adobe interface, and in fact need not know or care where the document properties were stored or accessed in a PDF file.