

Automating PDF Imports from Spreadsheets with Import Agent & Workflow

Brought to you by the City of Fort Worth Laserfiche Team

- Presented by Joey Chapa

Today's Approach

- I'll assume you are an intermediate to advanced user – but newbies are welcome!
- Quick overview of our Laserfiche implementation and the tools we used in this project
- The problem, a solution, and the obstacles we encountered along the way
- Parting advice
- Q & A

CFW's Laserfiche Implementation

- Laserfiche 11 (Rio) – on premise installation
- Full suite of applications, but not Forms or Quick Fields
- Licensed users typically opt to use the desktop Windows client
- WebLink is in widespread use a CSO public portal
- Leverage both templates and workflows to file our records
- Generally, we convert images files to TIF format during import

Tools - Templates & Workflows

- Templates – custom collections of fields that collect and present metadata
- Workflows – our main automation tool, the Laserfiche equivalent of a macro

Tools - Import Agent

- Monitors folders for incoming documents
- Moves them into the repository and applies any included metadata
- Filing and post-processing options via profiles
 - Color space
 - Image resolution
 - Optical character recognition

The Challenge: Water's First Laserfiche Project

- Historical Project Files~12,000 PDFs on Water's servers
- Inventoried in an Excel spreadsheet
- The Problem – how to get these into Laserfiche without duplicating effort

Template – WATER-Construction Files

- About twenty fields, several of which are multi-valued
 - *Project Name*
 - *Project Number (Water)*
 - *Project Number (Sewer)*
 - *Contractor*
 - A variety of names, numbers, and dates
- Serves as a pilot project to explore Laserfiche features and to act as a demonstration for future Laserfiche projects

Data Required for Filing

- Example from our scan shares

☐ Name	Date modified	Type	Size
 Pre-Laserfiche Scanning_Import003_SeqNum_02141_HOLLY PUMPING PLANT INSTALLATION OF 2-27.pdf	4/17/2023 8:59 AM	Adobe Acrobat D...	119,222 KB
 Pre-Laserfiche Scanning_Import003_SeqNum_02141_HOLLY PUMPING PLANT INSTALLATION OF 2-27.xml	6/26/2024 9:38 AM	XML File	3 KB
 Pre-Laserfiche Scanning_Import003_SeqNum_02142_EAST SIDE SUPPLY LINE, CONTRACT 1.pdf	8/15/2023 11:16 AM	Adobe Acrobat D...	161,768 KB
 Pre-Laserfiche Scanning_Import003_SeqNum_02142_EAST SIDE SUPPLY LINE, CONTRACT 1.xml	6/26/2024 9:38 AM	XML File	3 KB
 Pre-Laserfiche Scanning_Import003_SeqNum_02143_WEST SIDE WATER FEEDER MAINS, CONTRACT 2.pdf	4/17/2023 9:03 AM	Adobe Acrobat D...	134,862 KB
 Pre-Laserfiche Scanning_Import003_SeqNum_02143_WEST SIDE WATER FEEDER MAINS, CONTRACT 2.xml	6/26/2024 9:38 AM	XML File	3 KB


```
<?xml version="1.0" encoding="utf-8"?>
<LF:importengine xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://laserfiche.com/namespaces/importengine ImportEngine.xsd"
  xmlns:LF="http://laserfiche.com/namespaces/importengine" version="1.0">

  <LF:toc on_document_conflict="unique" on_folder_conflict="unique">

    <LF:document name="Pre-Laserfiche Scanning_Import003_SeqNum_02143_WEST SIDE WATER FEEDER MAINS, CONTRACT 2" template="WATER-Construction-Projects">
      <LF:metadata>
        <LF:field_values>
          <LF:field_value name="Constr-Proj Document Type">Construction Projects</LF:field_value>
          <LF:field_value name="Constr-Proj Document Description">Pre-Laserfiche Scanning</LF:field_value>
          <LF:field_value name="Constr-Proj Project Name" position="0">WEST SIDE WATER FEEDER MAINS, CONTRACT 2</LF:field_value>
          <LF:field_value name="Constr-Proj Project Number Water" position="0">86W-3M</LF:field_value>
          <LF:field_value name="Constr-Proj Project Number Sewer" position="0"></LF:field_value>
          <LF:field_value name="Constr-Proj Contractor" position="0"></LF:field_value>
          <LF:field_value name="Constr-Proj Developer" position="0"></LF:field_value>
          <LF:field_value name="Constr-Proj Consulting Engineer" position="0"></LF:field_value>
          <LF:field_value name="Constr-Proj Street-Intersection"></LF:field_value>
          <LF:field_value name="Constr-Proj M+C Number" position="0"></LF:field_value>
          <LF:field_value name="Constr-Proj City Secretary Contract Number"></LF:field_value>
          <LF:field_value name="Constr-Proj City Project Number"></LF:field_value>
          <LF:field_value name="Constr-Proj Department of Engineering Number (DOE)"></LF:field_value>
          <LF:field_value name="Constr-Proj Content From Date"></LF:field_value>
          <LF:field_value name="Constr-Proj Content To Date"></LF:field_value>
          <LF:field_value name="Constr-Proj Accession Document Number"></LF:field_value>
          <LF:field_value name="Constr-Proj Documents" position="0"></LF:field_value>
          <LF:field_value name="Constr-Proj Box Number Originating">14-B-4</LF:field_value>
          <LF:field_value name="Constr-Proj Import Batch">Import003</LF:field_value>
          <LF:field_value name="Constr-Proj Sequence Number">2143</LF:field_value>
          <LF:field_value name="Constr-Proj Date of Scanning"></LF:field_value>
          <LF:field_value name="Retention Start Date"></LF:field_value>
        </LF:field_values>
      </LF:metadata>

      <LF:electronic_document content_type="application/pdf" extension="pdf">
        <LF:fileref ref="\\laserwf\ScanShares\WATER_Construction\Pre-Laserfiche Scanning_Import003_SeqNum_02143_WEST SIDE WATER FEEDER MAINS, CONTRACT 2.pdf"/>
      </LF:electronic_document>
    </LF:document>

  </LF:toc>
</LF:importengine>
```

An example
XML file

How Import Agent Processes a File

- It opens an XML file and finds the PDF it will import
- It copies the PDF into Laserfiche and converts it to TIF images
- If it fails (for any reason), the XML file is moved to an *IAError* folder
- It applies the template to the file and populates it with the proper values
- If desired, it runs the OCR process on the images

Opting to Use Visual Basic for Applications

- Why VBA?
 - Ubiquitous
 - Something I know
 - Capable of file management and the creation of text files
 - Self-serve mentality

The VBA Project

Has three main sections

- One procedure – one main process that assembles the XML
 - Specifies the columns where the metadata is located
- Two functions
 - One to create one or more lines of XML for each field
 - Another to swap out any reserved characters with character codes (this will ensure that all field values populate correctly in the template)

Benefits of this approach

- Adding metadata is flexible and familiar
- Side-step any drop-down fields
- Easy to test changes and to file items in both test and production
- Ownership of the process

The Downsides

- Spreadsheets are easy to break
- Some knowledge of VBA is required
- Assumes “import-friendly” images

Complications

- Aged records – microfilmed, then converted to PDF
- Inconsistent resolution, color spaces, and cropping
- Incorrect page sizes
- *Discovered at different times*

Big issues

- Some items were importing at ten to twenty times the size of the original PDFs.
- Files had mixed resolutions & color spaces *within the same file*
- Files refused to import
- Files refused to convert to TIF
- OCR failures – partial processing, app lockups and crashes

Workarounds

- Sort by DPI, adjust the import profile
- Sort by color space, adjust the import profile
- Import manually, copying metadata from the XML file

The Payoff

- 19,000+ construction files now in Laserfiche
- Fully searchable – all images OCR'ed
- New Import method via spreadsheet

To-Do Items

- Generalize the macro to work with any inventory, any repository
- Move the metadata mapping *outside* of the code
- Investigate doing the same thing with workflow
- Create or find a separate tool to inspect images before import

Parting Advice

- Laserfiche is indispensable, but mind the whole process
- Image quality is paramount
 - Know the quality of your images
 - Constantly monitor ongoing scanning projects
- Be practical
 - You've got what you've got - call it a win
- One feature request – per page image conversion in Import Agent

Any questions?

