Configuring

Setting up to process sets of files containing PDF documents and data

To process sets of files containing one or more documents in a PDF file and additional data in an auxiliary input file, you set up an input device that uses one of the batching methods for sets: **Number of sets**, **Pages in sets**, or **Sets by time**. Then, you set up a PDF workflow. When the **IdentifyPDFDocuments** step runs, it produces a single PDF job that contains all the individual documents. The step includes the additional data in the document properties file (DPF) for use by other steps in the workflow.

The step uses two properties to identify the additional data:
- The value of the **Auxiliary input file extension** property identifies the file that contains the data.
- The value of the **Headers file** property identifies the file that specifies which data in the auxiliary input file to add to the DPF.

This procedure uses an example to show how to process sets of files containing PDF documents and data.

- In the example, an insurance company uses an application to produce letters to customers. Each letter is in a separate PDF file. The letters produced by the application do not contain the name, phone number, and email address of the agent for the customer.
- A separate application produces comma-separated (CSV) files with the agent information and other data that helps the company compose and route the letters. The two applications output 100 pairs of files at the same time.
- To optimize the processing of the letters and to add the contact information for the agents, the company batches 50 letters into a single PDF job with the 50 CSV files required to read the agent data into the DPF that the other steps in the workflow use.

Before you set up a workflow and an input device:

- Make sure that the auxiliary input file meets these requirements:
  - The header line contains the database names of document properties, separated by commas, for each data value that you want to add to the document properties file.
  - The file has one data line for each document in the associated PDF file.
  - On each data line, the values for the document properties are separated by commas.
  For example, this auxiliary input file is associated with a PDF file that contains one document. The file contains a header line and one data line with five data values:
  ```
  Doc.Custom.AgentName,Doc.Custom.AgentPhone,Doc.Custom.AgentEmail,Region,AgentCodeKelly Lopez,1-800-
  555-1234,kelly.lopez@insurancebiz.com,Southeast,B475
  ```

- Make sure that the headings for all the data values you want to use in the workflow are defined as RICOH ProcessDirector document properties.
  In the example, you want to use the values for `Doc.Custom.AgentName`, `Doc.Custom.AgentPhone`, and `Doc.Custom.AgentEmail` in a step based on the **EmailDocuments** step template.
  You define **Doc.Custom.AgentName**, **Doc.Custom.AgentPhone**, and **Doc.Custom.AgentEmail** as custom document properties.

  > ⬇ **Note:**
  > - We recommend that the names of custom document properties start with `Doc.Custom`.
  > - If you do not want to use a data value in the workflow, you do not need to define the heading for the data value as a RICOH ProcessDirector document property. In the example, you do not define `AgentRegion` and `AgentCode` as RICOH ProcessDirector document properties.

- Create a headers file that lists the database names of the document properties whose values you want to add to the DPF. Each database property name is on a separate line.
  For example, you create a `headers.txt` file with this content:
  ```
  Doc.Custom.AgentNameDoc.Custom.AgentPhoneDoc.Custom.AgentEmail
  ```

When the **IdentifyPDFDocuments** step in your workflow processes the set of files in the example, it creates a document properties file with data extracted from the letter and data from the auxiliary input file. For example, the company has mapped customer name and customer email address data in the letter to document properties in the Identify PDF control file. The **IdentifyPDFDocuments** step creates a DPF with these values:
```
Doc.Custom.CustomerName Doc.EmailAddress        Doc.Custom.AgentNameDoc.Custom.AgentPhone
Doc.Custom.AgentEmailChris Smith         chris.smith@myisp.com   Kelly Lopez1-800-555-1234
kelly.lopez@insurancebiz.com
```

To set up to process sets of files containing PDF documents and data:

1. Click the **Administration** tab.

2. In the left pane, click **Devices ⇒ Input Devices**.

3. Add or copy a hot folder input device.

    For example, click **Add ⇒ Hot Folder**, and name the input device **PDFInputFromSets**.

4. On all the tabs, fill in the required and optional properties that you need to adjust to match your environment.

5. Click the **General** tab.

6. For the **Child workflow** property, select the name of the workflow that you are modifying to process sets of files.

    For example, select **PDFDocumentsFromSets**.
    If you are creating a new workflow, use the default value. After you save the new workflow, display the properties for the input device and select the workflow as the value of this property.

7. Click the **Batching** tab.

8. For the **Batching method** property, select **Number of sets**, **Pages in sets**, or **Sets by time**.

9. Specify values for other properties associated with the batching method you selected.

    For example, you want the input device to batch and submit files to the workflow after it receives 50 sets of PDF files and CSV files. For the **Number of files to batch** property, enter `50`.

10. Specify a value for the **Matching pattern for sets** property, or use the default value:
    `(.+?)\.[^.]*$|$`
    The default value tells RICOH ProcessDirector to add files whose names are identical except for their extensions to the same set.
    For example:
    `customer_1.pdfcustomer_1.csv`

11. For the **Data patterns** property, enter: `.*pdf$`

12. Enter property values for the file pattern that identifies an auxiliary input file.

    For example, type these values for an auxiliary input file with a CSV file extension:
    - **File pattern**: `.*csv$`
    - **Spool file usage**: `auxinput`
    - **Spool file type**: `csv`
    - **File pattern required**: Yes
    - **File pattern sequence**: `1`

    RICOH ProcessDirector lets you use any value for the **Spool file usage** property that is not a RICOH ProcessDirector keyword. Keywords include `control`, `overrrides`, and `print`.

13. Click **Add**.

14. When you finish setting property values for the input device, click **OK**.

15. Click the **Workflow** tab.

16. Open a workflow that you want to modify, or create a new workflow.

    For example, you copy and modify the **EnhancePDFDocuments** supplied workflow. You name the copied workflow **PDFDocumentsFromSets**.

17. Add or modify a step based on the **IdentifyPDFDocuments** step template.

18. Set values for the properties of the **IdentifyPDFDocuments** step:

    1. For the **Identify PDF control file** property, specify the full path or symbolic name of the control file that you created using RICOH ProcessDirector Plug-in for Adobe Acrobat.
       The default Identify PDF control file defines each PDF file as a single document. Use RICOH ProcessDirector Plug-in for Adobe Acrobat to create a custom control file if:
       - Your PDF files contain two or more documents.
       - You want to add markup to the documents.
       - You want to map data in the documents to document properties.

    2. For the **Auxiliary input file extension** property, enter the file extension of the auxiliary input files.
       Make sure that this value matches the value of the **Spool file type** property for the auxiliary input file pattern that you defined on the input device

       In the example, the auxiliary input files have a CSV extension. Enter: `csv`

    3. For the **Headers file** property, enter the full path and name of the file that lists which values to copy from the auxiliary input file to the DPF.

       For example, enter `/aiw/aiw1/aux_input/headers.txt` (Linux) or `C:\aiw\aiw1\aux_input\headers.txt` (Windows).

19. Make other changes to the workflow as needed.

20. Save the workflow.

21. Test the input device and workflow:

    1. Enable the workflow.

    2. Enable and connect the input device that sends jobs to the workflow.

    3. Submit sets of PDF files and CSV files to the input device until you reach the limit at which the input device batches the files and submits them to the workflow.

For the example input device, submit 50 sets of PDF and CSV files.

When the **IdentifyPDFDocuments** step runs, it generates this output:

- A single PDF file with all the documents from all the sets.
- A `sets` directory that contains a subdirectory for each set of files.
- A document properties file that contains the values for any data mapped to document properties in the Identify PDF control file and the values of data from the auxiliary input files.

Parent topic: [Configuring the PDF Document Support feature](#)