File Specifications – Handling Delimited Files

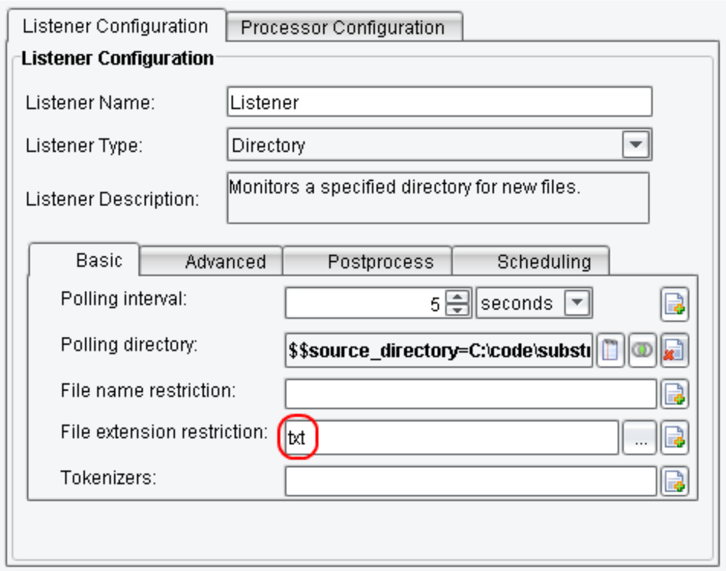
Overview

In this tutorial we'll cover using the “Delimited and Fixed-Width File” Transformer and the related File Specification Editor to convert a delimited file into XML. This tutorial assumes you're familiar with the concepts demonstrated in “Using Transformers” and extends them by introducing an additional Transformer.

Steps

Create a new Route with a Directory Listener / Transport pair, configured in the same way as those in “Creating a Simple Route.”

The Directory Listener's extension should be “txt”:



Listener Configuration

Listener Name: Listener

Listener Type: Directory

Listener Description: Monitors a specified directory for new files.

Polling interval: 5 seconds

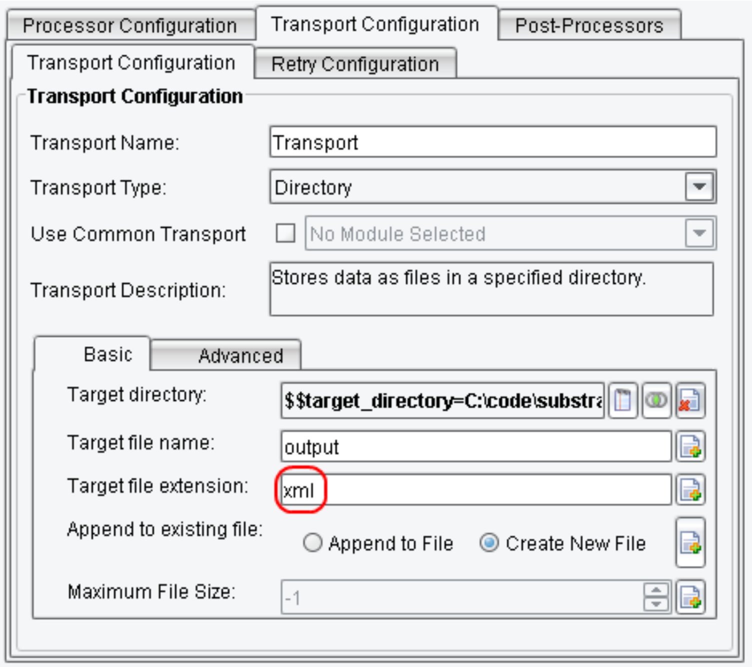
Polling directory: $$source_directory=C:\code\substi

File name restriction:

File extension restriction: txt

Tokenizers:

We'll make the Directory Transport's extension “xml”:



Transport Configuration

Transport Name: Transport

Transport Type: Directory

Use Common Transport □ No Module Selected

Transport Description: Stores data as files in a specified directory.

Target directory: $$target_directory=C:\code\substra

Target file name: output

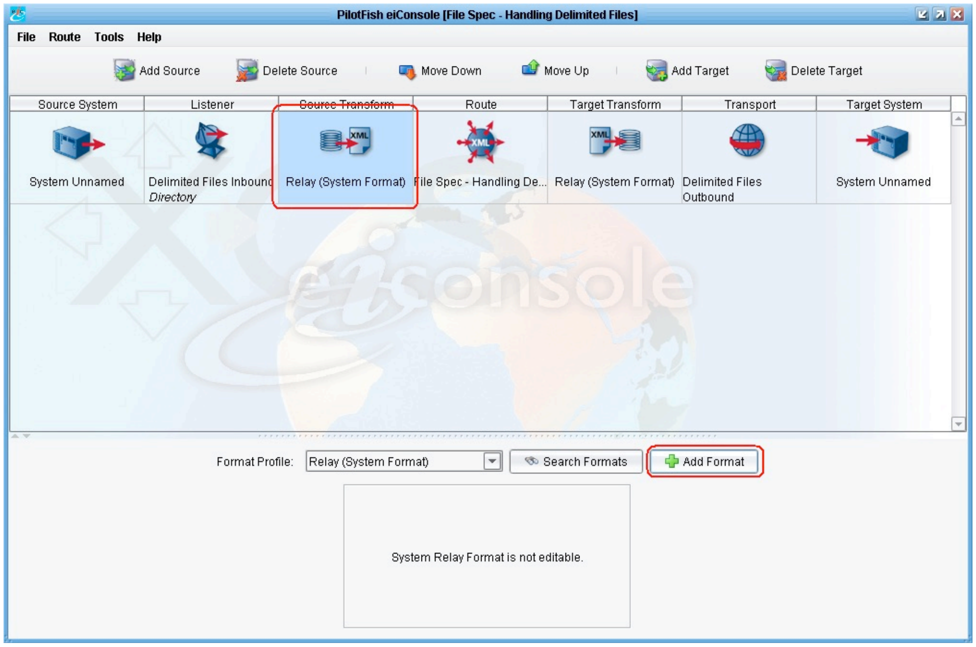
Target file extension: xml

Append to existing file: ○ Append to File ● Create New File

Maximum File Size: -1

We'll now need to add our Format. One important concept to note is that the direction of a transformer is determined by where it appears in the Route. If we configure our Transformer on the Target Transform stage then it will expect XML as input and convert that to the described delimited format. Conversely, configuring the Transformer on the Source Transform will assume delimited input and produce XML output.

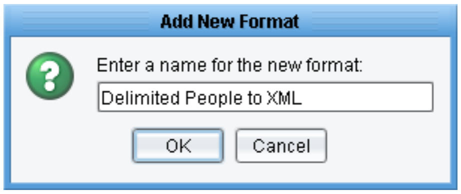
Because we wish to convert from a delimited format to XML, we'll want to add our Format on the Source Transform stage. Click that stage, then click Add Format:



Format Profile: Relay (System Format)

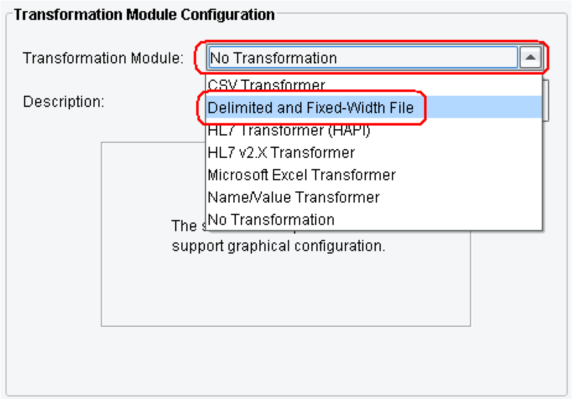
System Relay Format is not editable.

We'll name the Format based on the transform direction and the name of the structures. We'll be using a sample called “people_delimited.txt,” which should have been provided with this tutorial, so the Format will be named “Delimited People to XML”:



Enter a name for the new format: Delimited People to XML

With the Format added, the bottom half of the eiConsole screen changes to reflect the Format configuration. Once again, we're interested in the Transformation Module. Selected “Delimited and Fixed-Width File” from the “Transformation Module” drop-down:



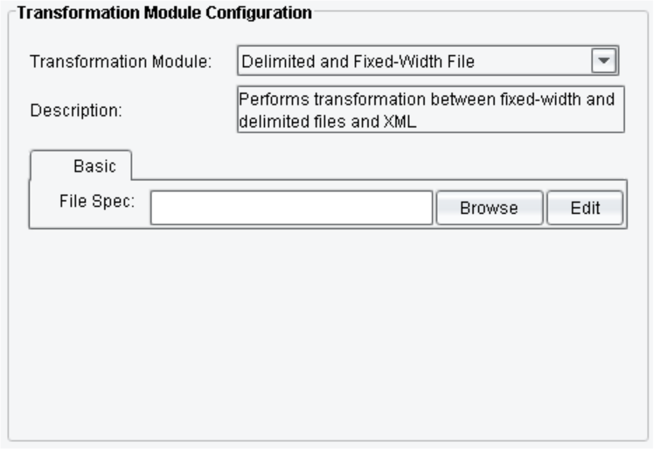
Transformation Module Configuration

Transformation Module: No Transformation

Description:

The s
support graphical configuration.

With the Delimited and Fixed-Width File Transformer selected, you should see a configuration panel like the following:



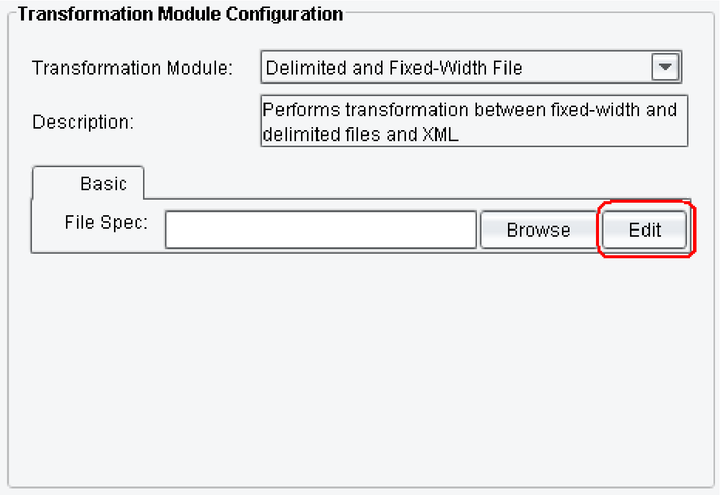
Transformation Module Configuration

Transformation Module: Delimited and Fixed-Width File

Description: Performs transformation between fixed-width and delimited files and XML

File Spec:

This particular Transformer makes use of an XML format informally called a “File Specification.” The configuration item specifies the name (and relative location) of such an XML description. If you had such an existing file, you could click “Browse” and select it, which would then copy it to the appropriate location. For this tutorial, we'll assume that we're starting fresh. Clicking the “Edit” button without any “File Specification” selected will open the “File Specification Editor;” do so:

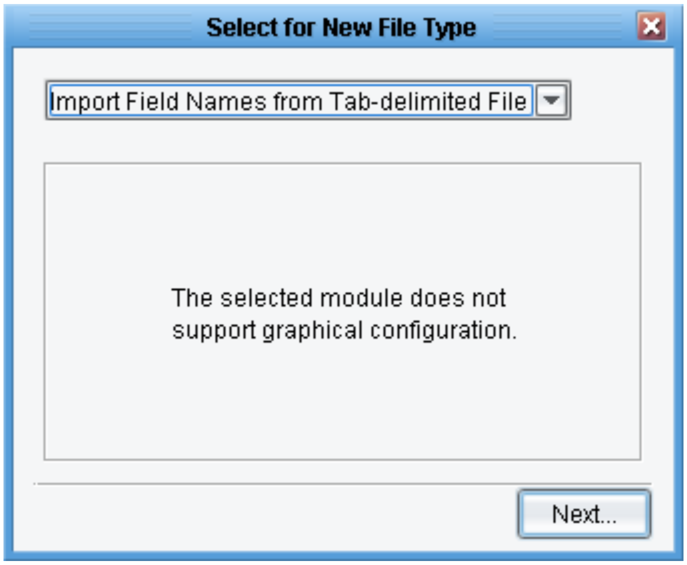


Transformation Module Configuration

Transformation Module: Delimited and Fixed-Width File

Description: Performs transformation between fixed-width and delimited files and XML

The File Specification Editor is a graphical tool used for defining the structure of a file format. When first opened, you will be presented with a dialog asking you how you wish to begin:



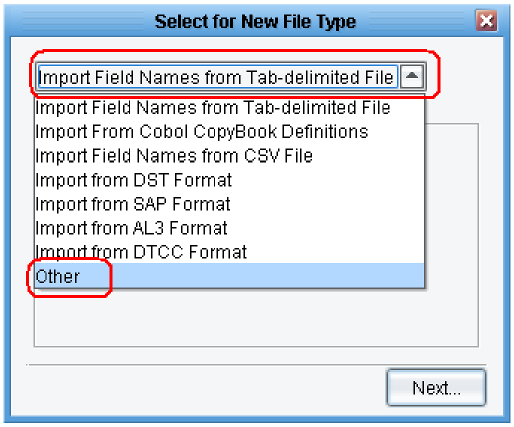
Select for New File Type

Import Field Names from Tab-delimited File

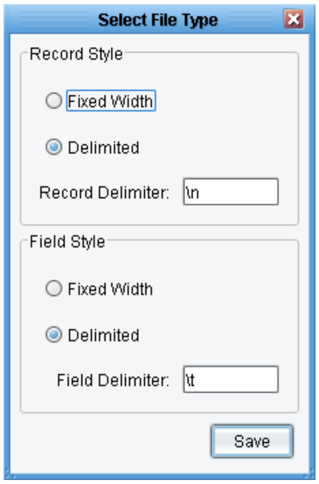
The selected module does not support graphical configuration.

The various options in this drop-down are ways to automatically build or import the structure of a File Specification. For example, you could read in definitions from a CSV file, a Cobol Copybook, or import various industry-specific formats (depending on what eiConsole version you have).

If you wish to build a File Specification from scratch, you'll want to select "Other":


Select for New File Type
Import Field Names from Tab-delimited File
Import Field Names from Tab-delimited File
Import From Cobol CopyBook Definitions
Import Field Names from CSV File
Import from DST Format
Import from SAP Format
Import from AL3 Format
Import from DTCC Format
Other
Next...

Click "Next" and you'll be presented with this dialog:



Record Style

○ Fixed Width

● Delimited

Record Delimiter: \n

Field Style

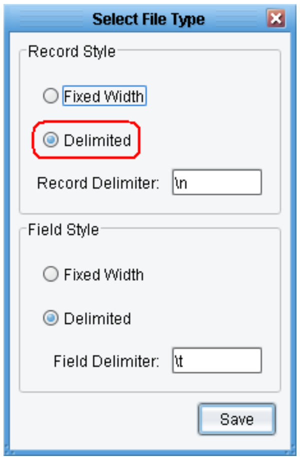
○ Fixed Width

● Delimited

Field Delimiter: \t

This dialog asks you to describe the style of Records and Fields. A File Specification is made up of Records with additional Records and Field as children. In most formats, Records are delimited, meaning that some character or set of characters separates each Record. In our “people_delimited.txt” sample, each line in the file represents a “Person,” which is a type of Record.

Select “Delimited” under “Record Style”:



Record Style

○ Fixed Width

● Delimited

Record Delimiter: \n

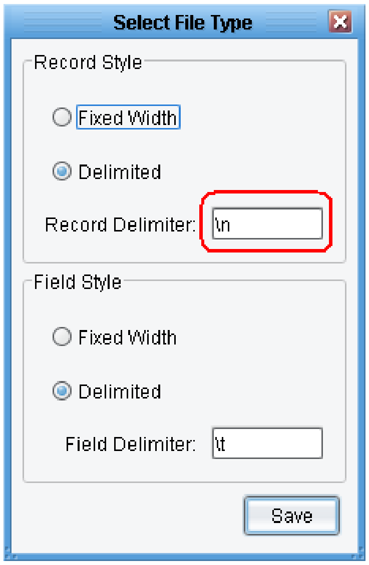
Field Style

○ Fixed Width

● Delimited

Field Delimiter: \t

The next item is “Record Delimiter,” which specifies what separates each Record. The default value, “\n,” represents a “line feed.” The convention used for escape characters is that used for Java (and most languages). Most formats use either a single line feed or a “carriage return” (\r) and a line feed (\r\n). Our sample file uses only a line feed, so set the Record Delimiter to “\n”:



Record Style

○ Fixed Width

● Delimited

Record Delimiter: \n

Field Style

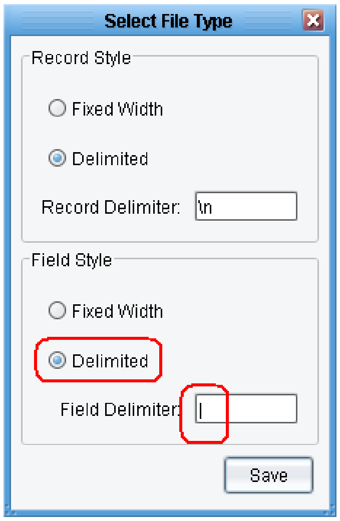
○ Fixed Width

● Delimited

Field Delimiter: \t

Next we'll need to define the style of Fields. In our sample file,each Field within a Record is separated by a pipe (“|”) character. Delimited files typically use tabs (“\t”), commas (“,”), pipes, and other such characters. Fields may also be “Fixed Width,” meaning that each field has a specific length. For example, the “First Name” field might be exactly 10 characters long.

Because our sample file is delimited by pipes, select “Delimited” and set the “Field Delimiter” to “|”:



Select File Type

Record Style

○ Fixed Width

● Delimited

Record Delimiter: \n

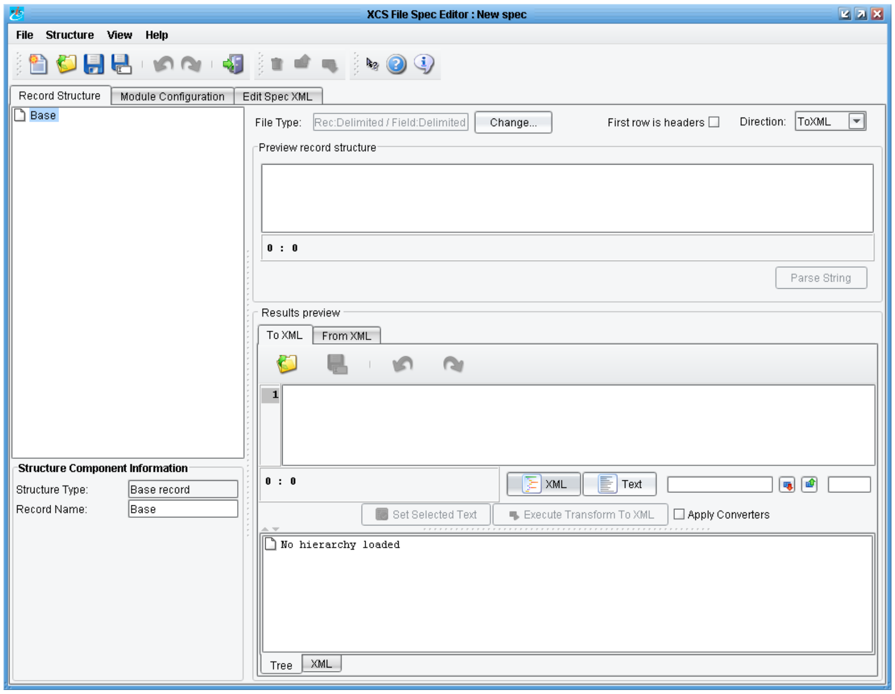
Field Style

○ Fixed Width

● Delimited

Field Delimiter: |

Finally, click the “Save” button to open the main File Specification Editor window:



XCS File Spec Editor : New spec

File Type: Rec:Delimited / Field:Delimited

First row is headers ☐

Direction: ToXML

Structure Type: Base record

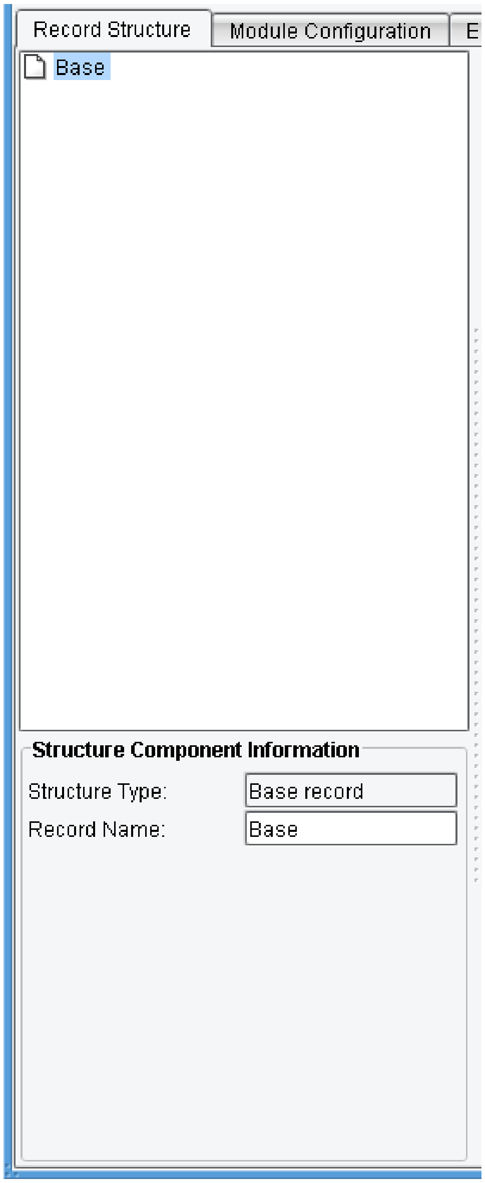
Record Name: Base

☐ Apply Converters

No hierarchy loaded

There are a lot of panels and areas for the File Specification Editor, but we'll review each of these in turn to explain them.

On the left side of the screen is the "Record Structure" panel:

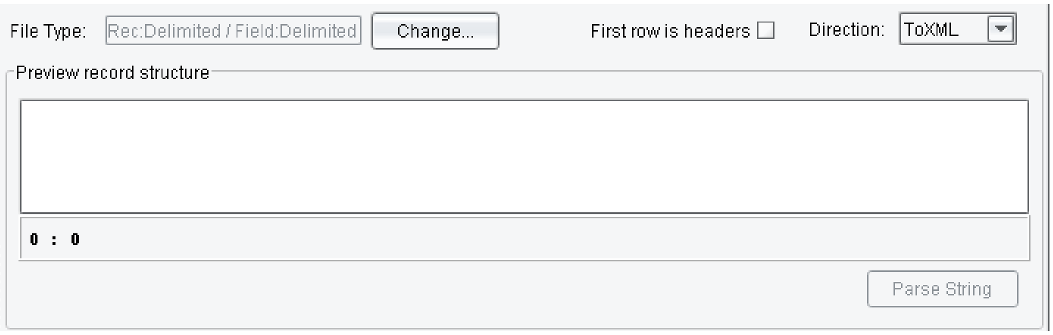


Structure Type: Base record

Record Name: Base

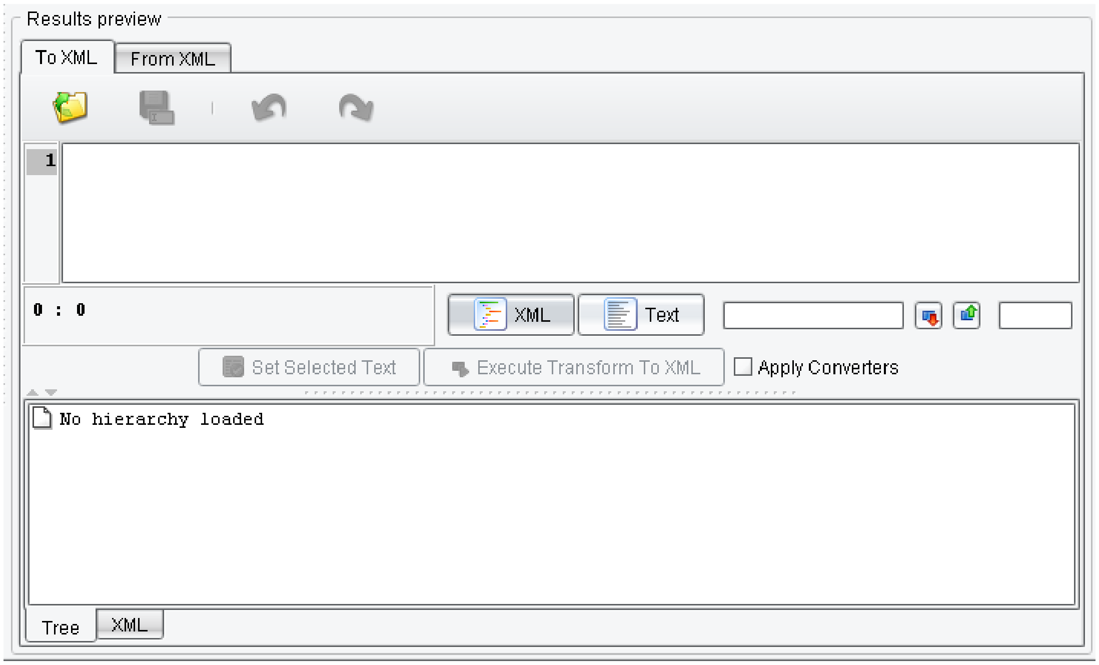
This panel contains the structure (so far, it's empty) of the File Specification, defined as a hierarchy (a tree) of Records, sub-Records, and Fields. The bottom half of the panel is used to provide information and configurations for the selected Record or Field.

The top-right panel is used to show currently loaded sample data with highlighting for a selected Record:

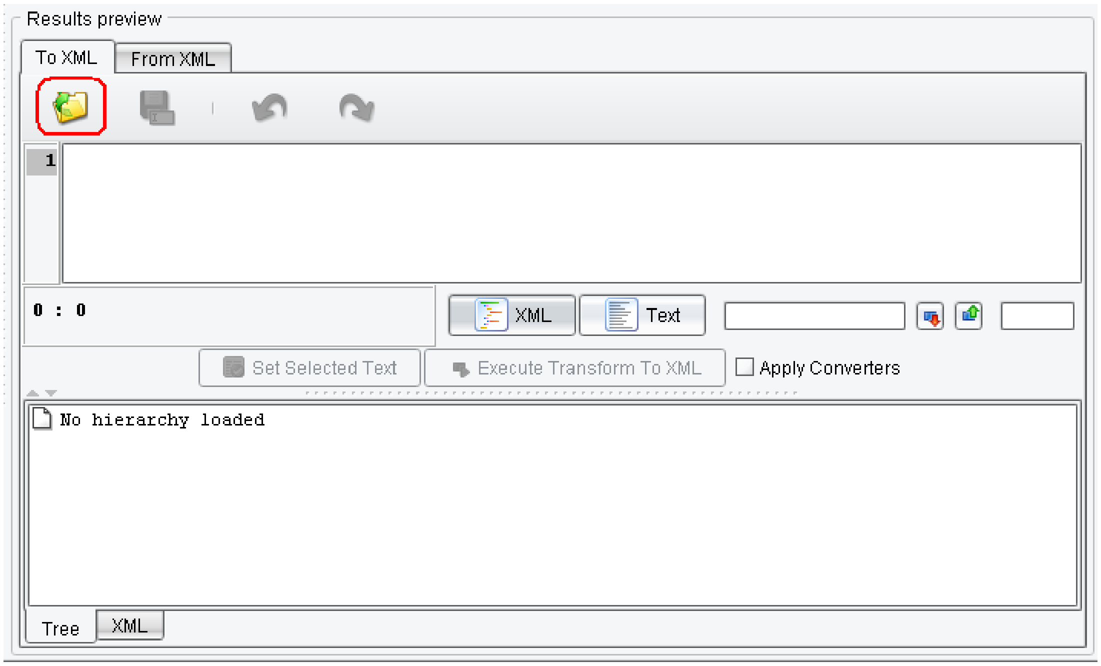


Until we have a structure defined and a sample file loaded, this will be empty and fairly uninteresting.

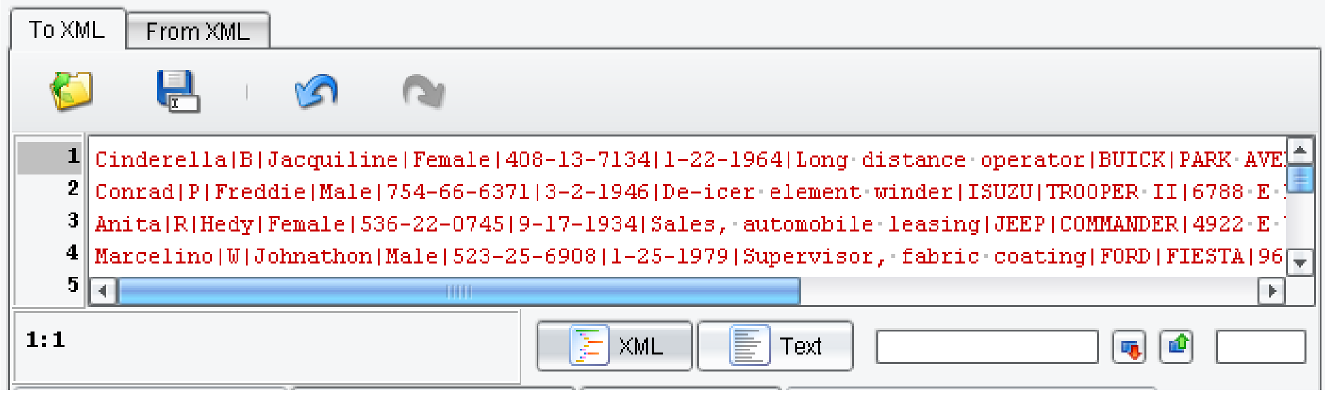
Finally, the “Results Preview” panel:



This panel is used to load and / or edit a sample file, transform it, and see the results. We'll start here. Click the “Load” icon and select “people_delimited.txt”:

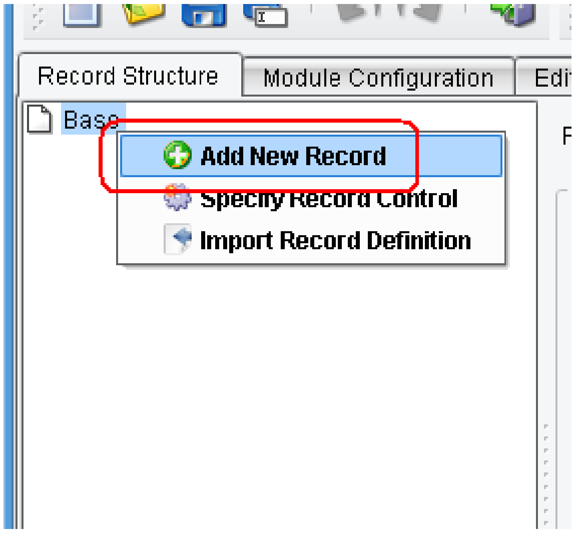


With the sample file loaded, the sample file panel should resemble the following:

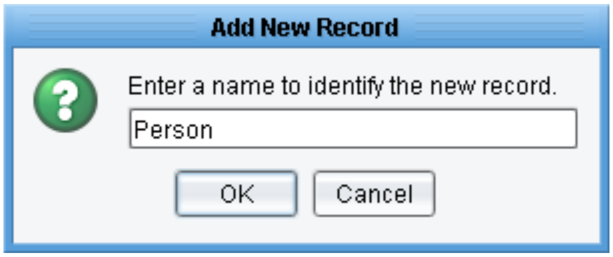


```
1 Cinderella|B|Jacquiline|Female|408-13-7134|1-22-1964|Long distance operator|BUICK|PARK AVE
2 Conrad|P|Freddie|Male|754-66-6371|3-2-1946|De-icer element winder|ISUZU|TROOPER II|6788 E
3 Anita|R|Hedy|Female|536-22-0745|9-17-1934|Sales, automobile leasing|JEEP|COMMANDER|4922 E
4 Marcelino|W|Johnathon|Male|523-25-6908|1-25-1979|Supervisor, fabric coating|FORD|FIESTA|96
5
1:1
```

We'll now need to define our structure. We can start by defining a Record to represent each "Person" in our sample file. Right-click on the root node (labeled "Base") and select "Add New Record":



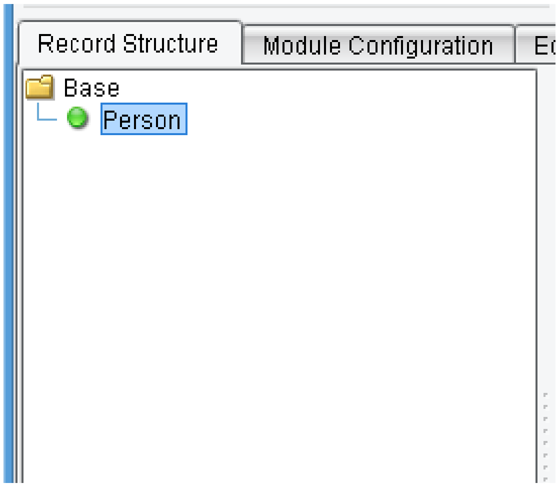
You will next be presented with a dialog where you are expected to name the new Record. Since this Record will represent a Person, we'll name it accordingly:



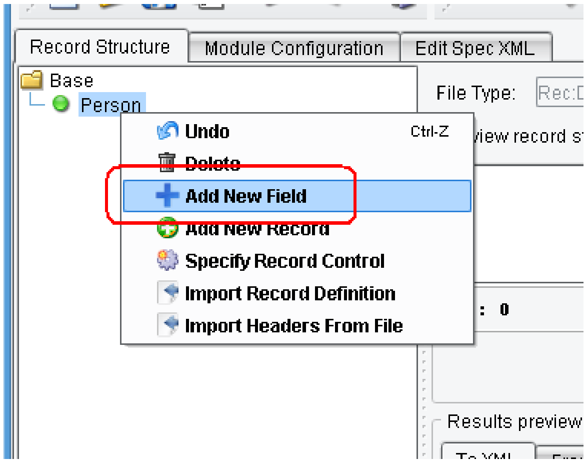
Enter a name to identify the new record.

Person

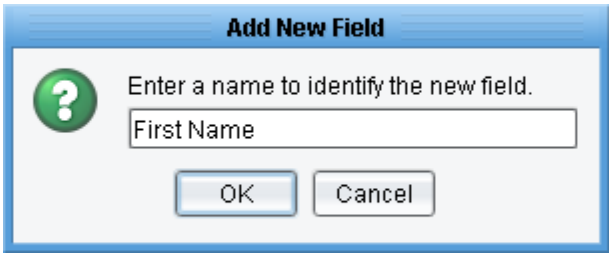
Click the “OK” button and you should see the “Person” record shown in the structure underneath “Base”:



We'll now wish to start adding Fields to the Person Record. To do so, right-click on Person and select "Add New Field":



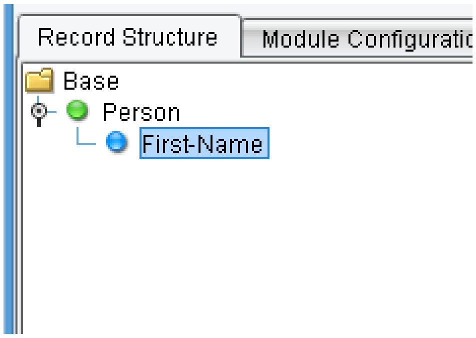
Once again, a dialog will be raised asking you to name the Field. Our first Field is the Person's first name, so we'll name the Field "First Name":



Enter a name to identify the new field.

First Name

Click the “OK” button and you should see the Field displayed under Person:

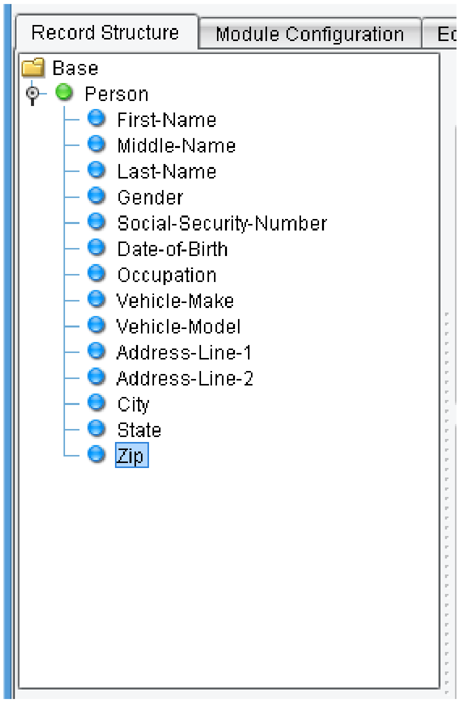


For defining the remaining Fields, you could continue to right-click on “Person” and select “Add New Field” or, with the “First-Name” field selecting, simply hit “Enter / Return” on your keyboard. The next field is “Middle Name.”

Continue adding Fields for each of the following:

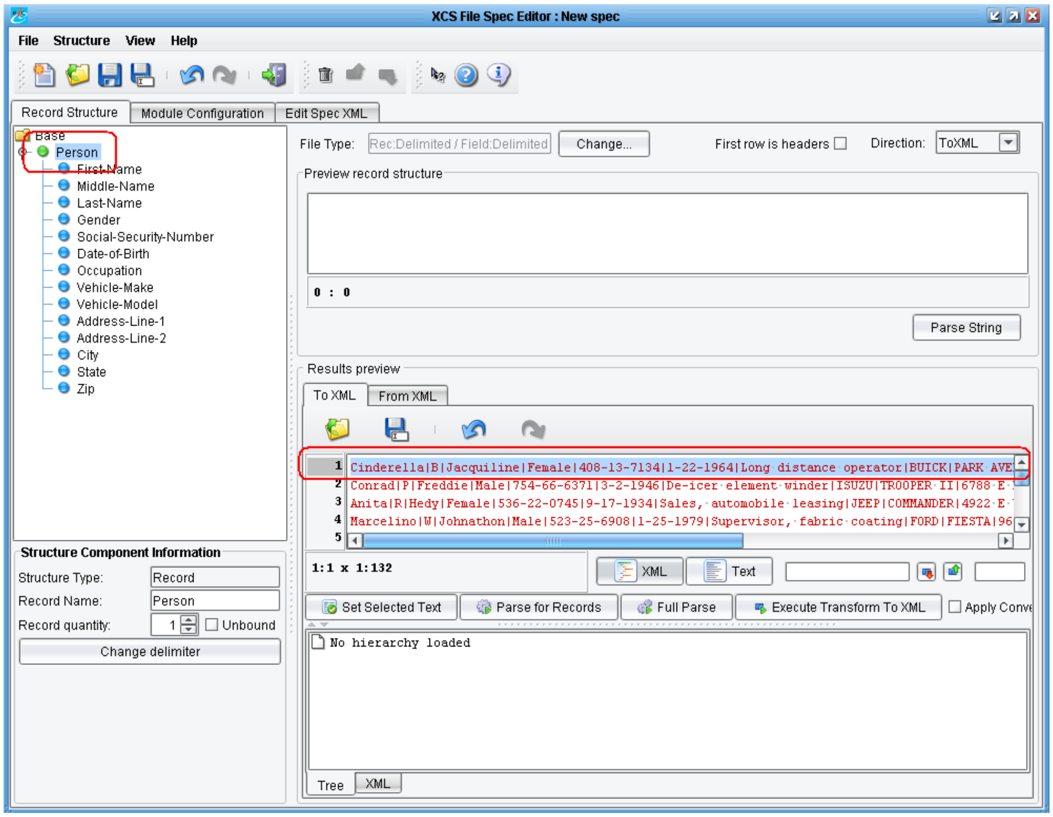
- First Name
- Middle Name
- Last Name
- Gender
- Social Security Number
- Date of Birth
- Occupation
- Vehicle Make
- Vehicle Model
- Address Line 1
- Address Line 2
- City
- State
- Zip

When you are finished, the structure should look like this:



- Base
  - Person
    - First-Name
    - Middle-Name
    - Last-Name
    - Gender
    - Social-Security-Number
    - Date-of-Birth
    - Occupation
    - Vehicle-Make
    - Vehicle-Model
    - Address-Line-1
    - Address-Line-2
    - City
    - State
    - Zip

Our structure is now fully defined. Next, select the Person Record and then highlight a single line of the sample file (this might be easier if you click at the beginning of that line, hold Shift, and hit End):



Structure Type: Record
Record Name: Person
Record quantity: 1 □ Unbound

File Type: Rec:Delimited / Field:Delimited
First row is headers □
Direction: ToXML

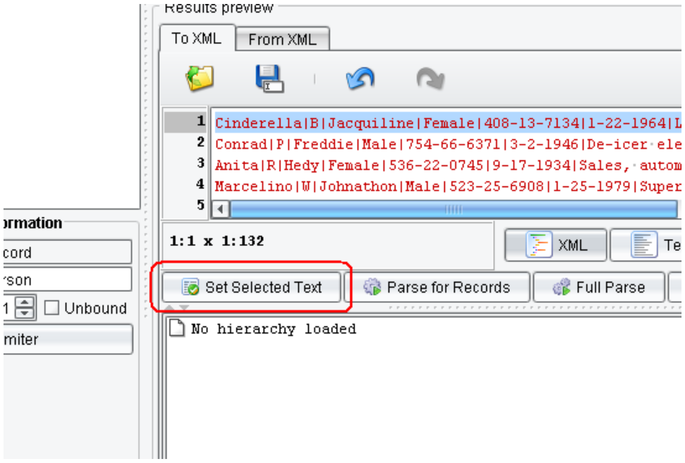
```
0 : 0
```

```
1 Cinderella|B|Jacquiline|Female|408-13-7134|1-22-1964|Long distance operator|BUICK|PARK AVE
2 Conrad|P|Freddie|Male|754-66-6371|3-2-1946|De-icer element winder|ISUZU|TROOPER II|6788 E
3 Anita|R|Hedy|Female|536-22-0745|9-17-1934|Sales, automobile leasing|JEEP|COMMANDER|4922 E
4 Marcelino|W|Johnathon|Male|523-25-6908|1-25-1979|Supervisor, fabric coating|FORD|FIESTA|96
5
```

```
1:1 x 1:132
```

```
No hierarchy loaded
```

Next, click “Set Selected Text”:



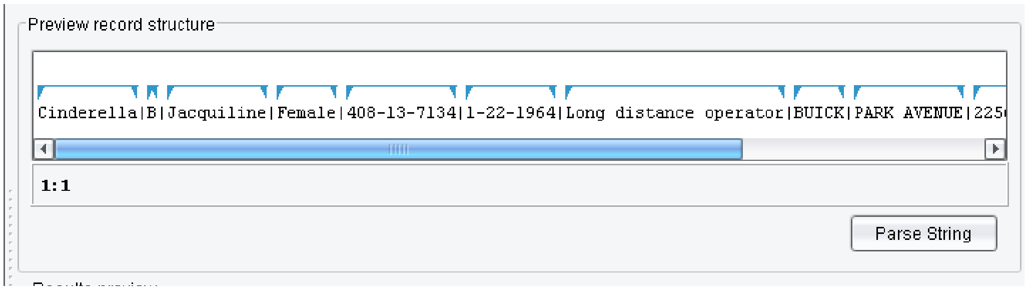
cord
son
1 □ Unbound

```
1 Cinderella|B|Jacquiline|Female|408-13-7134|1-22-1964|L
2 Conrad|P|Freddie|Male|754-66-6371|3-2-1946|De-icer ele
3 Anita|R|Hedy|Female|536-22-0745|9-17-1934|Sales, autom
4 Marcelino|W|Johnathon|Male|523-25-6908|1-25-1979|Super
5
```

```
1:1 x 1:132
```

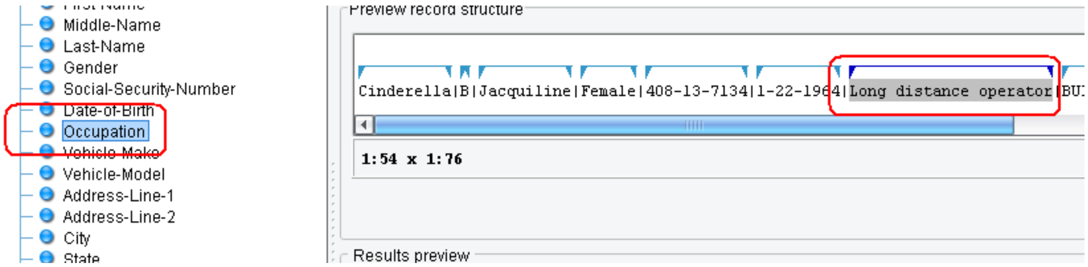
```
No hierarchy loaded
```

You'll notice the top panel now has various fields shown, as well as blue arrows:



```
Cinderella|B|Jacquiline|Female|408-13-7134|1-22-1964|Long distance operator|BUICK|PARK AVENUE|225
1:1
```

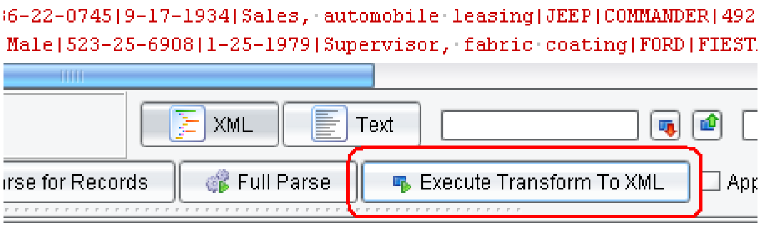
If you click on the various Fields in the structure, the appropriate field will be highlighted in this panel. For example, if we click Occupation, we'll see "Long distance operator" highlighted:



```
Cinderella|B|Jacquiline|Female|408-13-7134|1-22-1964|Long distance operator|BU
1:54 x 1:76
```

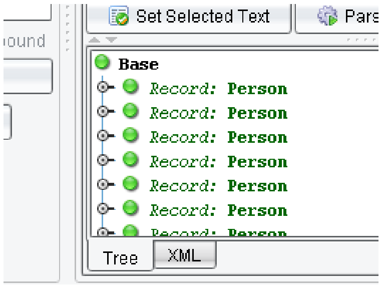
This panel allows you to visually confirm that your structure matches your sample (or vice-versa). While this is less important in delimited files, it can be an enormous time saver when dealing with fixed-width formats.

Next, we'll want to test our definition. Click the "Execute Transform to XML" button:



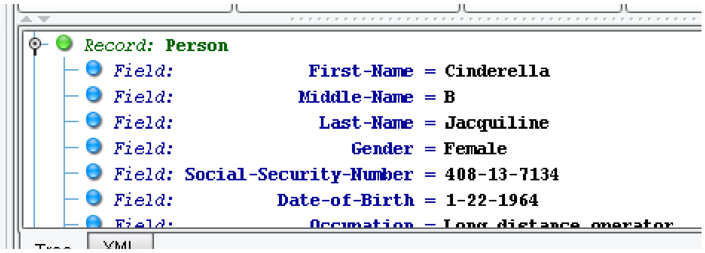
```
6-22-0745|9-17-1934|Sales, automobile leasing|JEEP|COMMANDER|492
Male|523-25-6908|1-25-1979|Supervisor, fabric coating|FORD|FIEST
```

You should see various Person elements appear in the results panel:



```
Base
  Record: Person
  Record: Person
  Record: Person
  Record: Person
  Record: Person
  Record: Person
  Record: Person
```

If you click to the left of these elements, it will expand them to show their respective child elements and values. For example, the first Person element should resemble this:

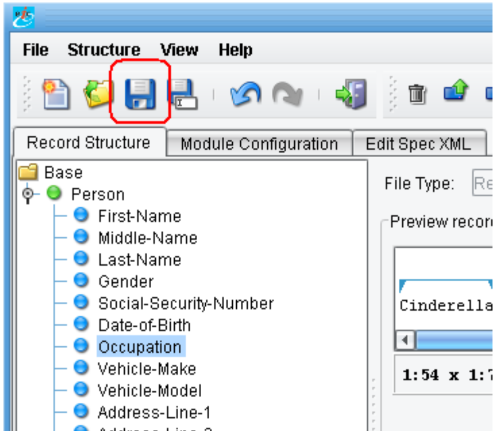


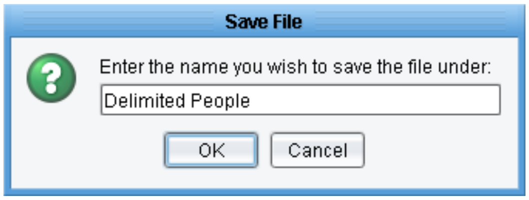
```
Record: Person
  Field:             First-Name = Cinderella
  Field:            Middle-Name = B
  Field:              Last-Name = Jacquiline
  Field:                 Gender = Female
  Field: Social-Security-Number = 408-13-7134
  Field:          Date-of-Birth = 1-22-1964
  Field:             Occupation = Long distance operator
```

If you click the “XML” tab, you can also view how this would look in the resulting XML.

Our File Specification is now fully defined. One interesting aspect to note is that this specification will convert the “people delimited” format to XML or the resulting XML format back to the delimited format without any additional configuration or changes; File Specifications are inherently bi-directional.

Save your specification with the “Save” button, providing any name you feel appropriate (we'll use “Delimited People”):

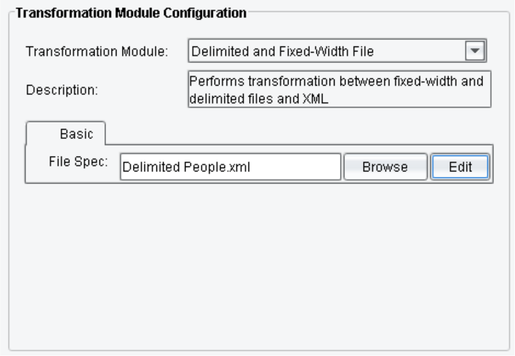



Save File

Enter the name you wish to save the file under:

Delimited People

Finally, close the Editor to return to the eiConsole. Our Format configuration should now look like this:


Transformation Module Configuration

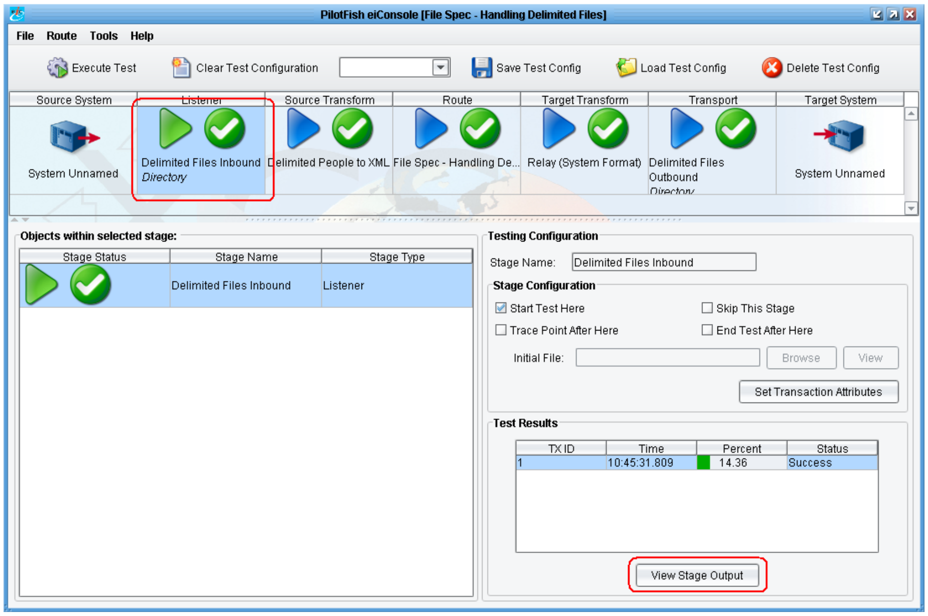
Transformation Module: Delimited and Fixed-Width File

Description: Performs transformation between fixed-width and delimited files and XML

Basic

File Spec: Delimited People.xml

With that, we've now completely configured our Route, as well. Switch to Testing Mode, place the sample file (“people_delimited.txt”) into the input directory, and let the test run all the way through. Select the Listener stage and view the stage output to see the format in it's “before” stage:



Objects within selected stage:

Stage Status	Stage Name	Stage Type
▶ ✓	Delimited Files Inbound	Listener

Testing Configuration

Stage Name: Delimited Files Inbound

Stage Configuration

☑ Start Test Here ☐ Skip This Stage

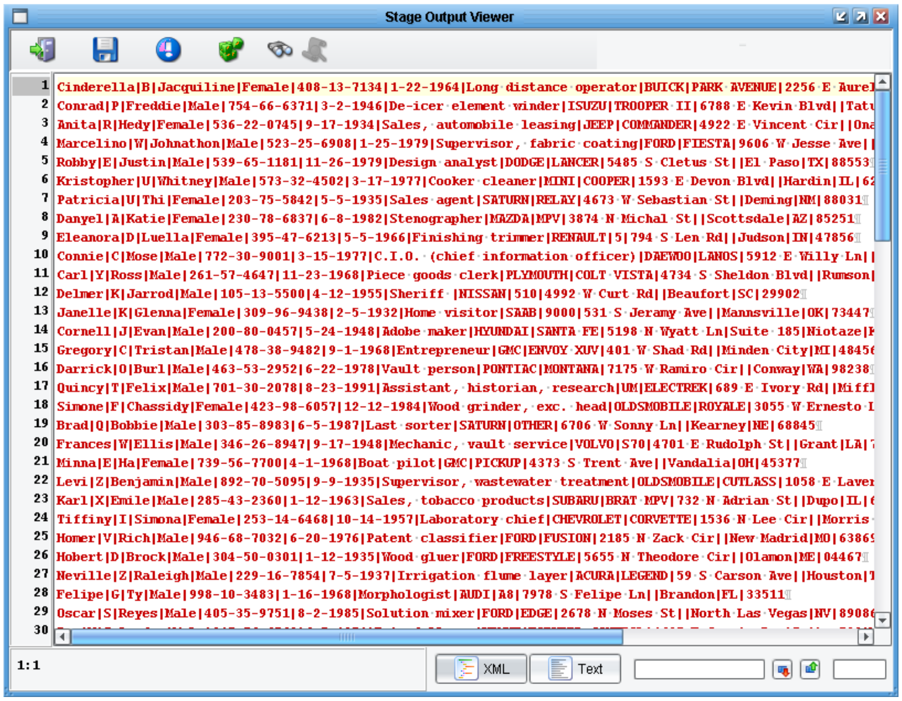
☐ Trace Point After Here ☐ End Test After Here

Initial File:

Test Results

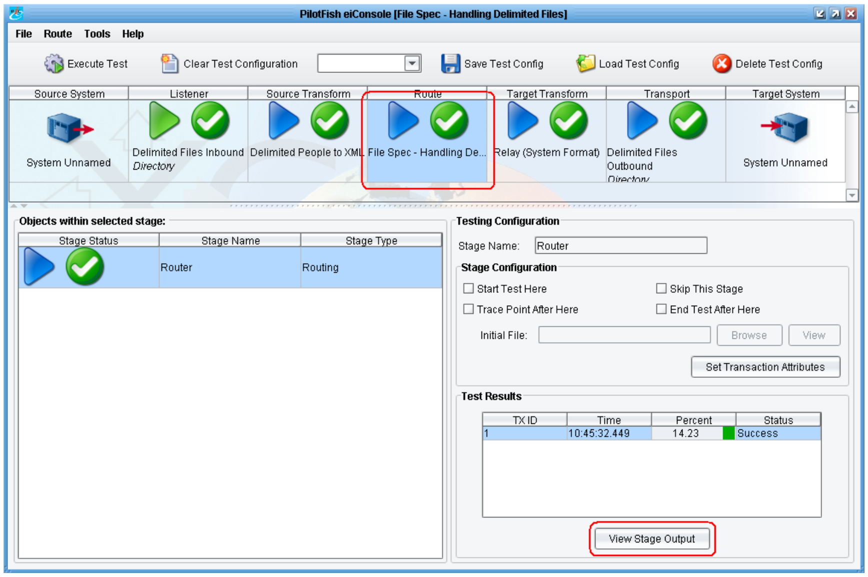
TX ID	Time	Percent	Status
1	10:45:31.809	🟩 14.36	Success

As you can see in the Stage Output Viewer, the file picked up matches our expected, delimited format:



```
Cinderella|B|Jacquiline|Female|408-13-7134|1-22-1964|Long distance operator|BUICK|PARK AVENUE|2256 E Aurel
Conrad|P|Freddie|Male|754-66-6371|3-2-1946|De-icer element winder|ISUZU|TROOPER II|6788 E Kevin Blvd||Tatu
Anita|R|Hedy|Female|536-22-0745|9-17-1934|Sales, automobile leasing|JEEP|COMMANDER|4922 E Vincent Cir||Ona
Marcelino|W|Johnathon|Male|523-25-6908|1-25-1979|Supervisor, fabric coating|FORD|FIESTA|9606 W Jesse Ave||
Robby|E|Justin|Male|539-65-1181|11-26-1979|Design analyst|DODGE|LANCER|5485 S Cletus St||El Paso|TX|88553
Kristopher|U|Whitney|Male|573-32-4502|3-17-1977|Cooker cleaner|MINI|COOPER|1593 E Devon Blvd||Hardin|IL|62
Patricia|U|Thi|Female|203-75-5842|5-5-1935|Sales agent|SATURN|RELAY|4673 W Sebastian St||Deming|NM|88031
Danyel|A|Katie|Female|230-78-6837|6-8-1982|Stenographer|MAZDA|MPV|3874 N Michal St||Scottsdale|AZ|85251
Eleanora|D|Luella|Female|395-47-6213|5-5-1966|Finishing trimmer|RENAULT|5|794 S Len Rd||Judson|IN|47856
Connie|C|Mose|Male|772-30-9001|3-15-1977|C.I.O. (chief information officer)|DAEWOO|LANOS|5912 E Willy Ln||
Carl|Y|Ross|Male|261-57-4647|11-23-1968|Piece goods clerk|PLYMOUTH|COLT VISTA|4734 S Sheldon Blvd||Rumson|
Delmer|K|Jarrod|Male|105-13-5500|4-12-1955|Sheriff |NISSAN|510|4992 W Curt Rd||Beaufort|SC|29902
Janelle|K|Glenna|Female|309-96-9438|2-5-1932|Home visitor|SAAB|9000|531 S Jeramy Ave||Mannsville|OK|73447
Cornell|J|Evan|Male|200-80-0457|5-24-1948|Adobe maker|HYUNDAI|SANTA FE|5198 N Wyatt Ln|Suite 185|Niotaze|K
Gregory|C|Tristan|Male|478-38-9482|9-1-1968|Entrepreneur|GMC|ENVOY XUV|401 W Shad Rd||Minden City|MI|4845
Darrick|O|Burl|Male|463-53-2952|6-22-1978|Vault person|PONTIAC|MONTANA|7175 W Ramiro Cir||Conway|WA|98238
Quincy|T|Felix|Male|701-30-2078|8-23-1991|Assistant, historian, research|UM|ELECTREK|689 E Ivory Rd||Miffl
Simone|F|Chassidy|Female|423-98-6057|12-12-1984|Wood grinder, exc. head|OLDSMOBILE|ROYALE|3055 W Ernesto L
Brad|Q|Bobbie|Male|303-85-8983|6-5-1987|Last sorter|SATURN|OTHER|6706 W Sonny Ln||Kearney|NE|68845
Frances|W|Ellis|Male|346-26-8947|9-17-1948|Mechanic, vault service|VOLVO|S70|4701 E Rudolph St||Grant|LA|7
Minna|E|Ha|Female|739-56-7700|4-1-1968|Boat pilot|GMC|PICKUP|4373 S Trent Ave||Vandalia|OH|45377
Levi|Z|Benjamin|Male|892-70-5095|9-9-1935|Supervisor, wastewater treatment|OLDSMOBILE|CUTLASS|1058 E Laver
Karl|X|Emile|Male|285-43-2360|1-12-1963|Sales, tobacco products|SUBARU|BRAT MPV|732 N Adrian St||Dupo|IL|6
Tiffiny|I|Simona|Female|253-14-6468|10-14-1957|Laboratory chief|CHEVROLET|CORVETTE|1536 N Lee Cir||Morris
Homer|V|Rich|Male|946-68-7032|6-20-1976|Patent classifier|FORD|FUSION|2185 N Zack Cir||New Madrid|MO|6386
Hobert|D|Brock|Male|304-50-0301|1-12-1935|Wood gluer|FORD|FREESTYLE|5655 N Theodore Cir||Olamon|ME|04467
Neville|Z|Raleigh|Male|229-16-7854|7-5-1937|Irrigation flume layer|ACURA|LEGEND|59 S Carson Ave||Houston|T
Felipe|G|Ty|Male|998-10-3483|1-16-1968|Morphologist|AUDI|A8|7978 S Felipe Ln||Brandon|FL|33511
Oscar|S|Reyes|Male|405-35-9751|8-2-1985|Solution mixer|FORD|EDGE|2678 N Moses St||North Las Vegas|NV|8908
```

Next, select the Routing stage (since it occurs after our transformation) and view its output:



Objects within selected stage:

Stage Status	Stage Name	Stage Type
✓	Router	Routing

Testing Configuration

Stage Name: Router

Stage Configuration

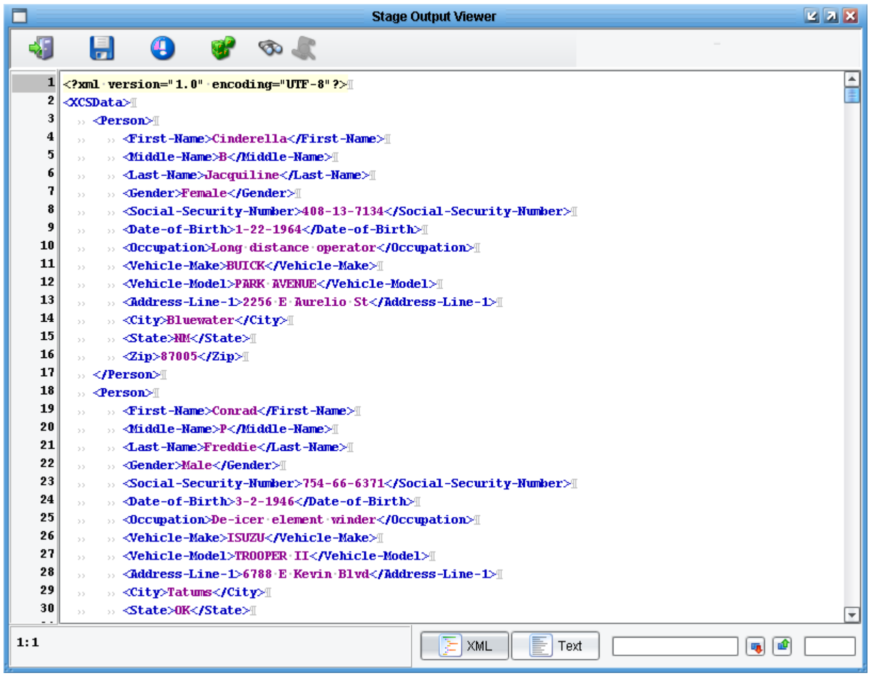
☐ Start Test Here ☐ Skip This Stage
☐ Trace Point After Here ☐ End Test After Here

Initial File:

Test Results

TX ID	Time	Percent	Status
1	10:45:32.449	14.23	🟩 Success

You should see a nice, well-structured XML document:



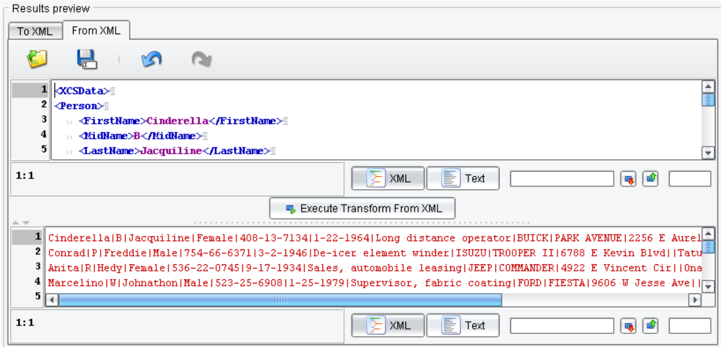
```
<?xml version="1.0" encoding="UTF-8"?>
<XCSData>
    <Person>
        <First-Name>Cinderella</First-Name>
        <Middle-Name>B</Middle-Name>
        <Last-Name>Jacquiline</Last-Name>
        <Gender>Female</Gender>
        <Social-Security-Number>408-13-7134</Social-Security-Number>
        <Date-of-Birth>1-22-1964</Date-of-Birth>
        <Occupation>Long distance operator</Occupation>
        <Vehicle-Make>BUICK</Vehicle-Make>
        <Vehicle-Model>PARK AVENUE</Vehicle-Model>
        <Address-Line-1>2256 E Aurelio St</Address-Line-1>
        <City>Bluewater</City>
        <State>NM</State>
        <Zip>87005</Zip>
    </Person>
    <Person>
        <First-Name>Conrad</First-Name>
        <Middle-Name>P</Middle-Name>
        <Last-Name>Freddie</Last-Name>
        <Gender>Male</Gender>
        <Social-Security-Number>754-66-6371</Social-Security-Number>
        <Date-of-Birth>3-2-1946</Date-of-Birth>
        <Occupation>De-icer element winder</Occupation>
        <Vehicle-Make>ISUZU</Vehicle-Make>
        <Vehicle-Model>TROOPER II</Vehicle-Model>
        <Address-Line-1>6788 E Kevin Blvd</Address-Line-1>
        <City>Tatums</City>
        <State>OK</State>
```

1:1

You can compare the format before-and-after if you like, but that's pretty much all that's necessary to freely convert between delimited and XML formats.

*** BONUS ***

Use the file specification created above to transform from the generated XML back into a flat file:



```
<XCSData>
<Person>
    <FirstName>Cinderella</FirstName>
    <MidName>B</MidName>
    <LastName>Jacquiline</LastName>
```

```
Cinderella|B|Jacquiline|Female|408-13-7134|1-22-1964|Long distance operator|BUICK|PARK AVENUE|2256 E Aurel
Conrad|P|Freddie|Male|754-66-6371|3-2-1946|De-icer element winder|ISUZU|TROOPER II|6788 E Kevin Blvd||Tatu
Anita|R|Hedy|Female|536-22-0745|9-17-1934|Sales, automobile leasing|JEEP|COMMANDER|4922 E Vincent Cir||Ona
Marcelino|W|Johnathon|Male|523-25-6908|1-25-1979|Supervisor, fabric coating|FORD|FIESTA|9606 W Jesse Ave||
```