

IMPORTING METADATA FROM XML FILES TO M-FILES



VERSION 1.2

M-Files supports ingesting metadata for new documents from XML files in the Connections to External File Sources feature.

This feature is often used to implement integration with multi-function devices and network scanners. In a typical configuration, user scans files to a network folder and M-Files Server imports those files as new documents to the vault. Many network scanners support insertion of metadata. Metadata is often saved to XML files by the scanner and M-Files can apply these property values as metadata for newly created documents.

1. XML in a nutshell

XML stands for Extensive Mark-up Language. Although, HTML is a mark-up language too, XML is not a replacement for it because it was created to **describe data** while HTML is a good language to **display data**. It is important to know that XML itself was created to structure and store information. Please note that **XML tags are case sensitive**.

In short, XML consist of five different elements: declaration, root element, child elements, attributes and name spaces.

In this example we have a scanned image named *testfile.pdf* and an XML file named *testfile.xml* having the following content:

```
<?xml version="1.0" encoding="UTF-8"?>
<document>
  <name>Test File Name</name>
  <category>Proposal</category>
  <keywords>
    <![CDATA[Test/Scan [><] XML]]>
  </keywords>
</document>
```

2. XPath in a nutshell

XPath is a W3C standard syntax for defining parts of an XML document. It uses expressions to navigate in XML documents and has a library of standard functions. In M-Files, XML files are parsed with Microsoft XML parser, so the user needs to input only correct paths inside the file to retrieve desired data.

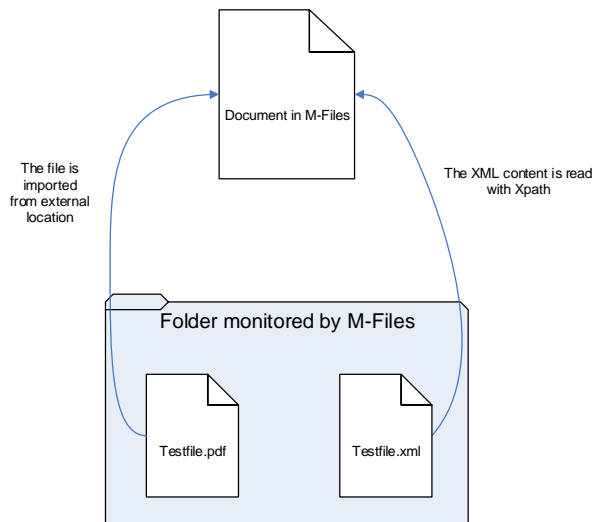
With about 100 standard functions of XPath, the user can easily find data from an XML documents. Moreover, calculations and operators can be used.

The data of XML files is selected with nodes. In M-Files the user needs to input only the path expression to retrieve data, i.e. to fetch "Test File Name" out of the testfile.xml the path is document/name. In more complex XML files, operators and attribute become useful. For more information, refer to for example <http://www.w3schools.com/xpath/>.

3. How to import files with Metadata to M-Files

Documents and their Meta data content is imported into M-Files using “Connections to External File Sources” feature. For further information of this feature, please refer to M-Files User’s Guide pp 133 - 136.

The picture below describes the functionality in more detail. The document file and the Meta Data file must have the name so that M-Files can distinguish between Meta Data and File Data (e.g. Testfile.pdf and Testfile.xml).



Picture 1 The files are imported to M-Files using the “Connections to external locations” feature

1. Create a new connection to an external file source in M-Files Server Administrator.
2. Name the connection.
3. Define the path to be monitored (UNC-paths are allowed too).
4. If your local system account does not have an access to the monitored path, press “Set account” button to define another user name to access the folder.
5. Select “Import files” in Action section and check “Delete source file after importing” (optional).
6. Uncheck “Preserve folder structure”.
7. When importing scanned images, the default delay (15 minutes) between checks may be too long.
8. In Meta Data tab, check “Read values from an XML file” and “Delete the XML file after use” (optional).
9. Select *Class* and press *Edit* from *Properties* windows, fill in the form as follows and press *OK*:

The screenshot shows the 'Define Property' dialog box with the following settings:

- Property: Class
- ☐ Use a fixed value:
- ☐ Read from an XML file produced by HP DSS
Specify the name of the prompt you have defined in HP Digital Sending Software (DSS):
- ☒ Read from an XML file
Specify the XPath expression that determines the appropriate node in the XML document:
`/document/category`
- ☐ Use an OCR value source
Define... (not defined)
- Conversion to value list item:
 - ☐ Use the value read as the ID of the item
 - ☒ Use the value read as the name of the item
 - ☐ Add a new item to the list if a matching item is not found

Buttons: OK, Cancel

10. Add new property and select *Keywords* from the *Property* drop-down menu. Choose *Read from an XML file* and type in the textbox *document/keywords*. Finally press *OK*.

11. Add new property and select *Name or title* from *Class* drop-down menu. Choose *Read from an XML file* and type in the textbox *document/name*. Finally press *OK*.
12. Finally press *OK* to finalize the new connection to external location.

NOTE! If you cannot retrieve metadata from an XML document as supposed, have a look at Windows Event Viewer of the M-Files Server computer as XPath syntax errors are stored in the application log there.

4. Registry settings

You can change the behavior of XML import by applying registry settings described in this chapter.

4.1 Changing the Locale ID for Imported Values

When importing properties from XML, you must take regional settings for numbers into account. This is because the date 11/9/2015 is interpreted as November 9, 2015 when using the U.S. Locale Settings and as 11 September 2015 when using the UK Locale Settings. Similarly, most European regional settings use comma as a decimal symbol and space as a thousand separator (digit grouping symbol). When using the U.S. Locale settings, a dot is used as a decimal symbol and comma as a thousand separator.

You can specify the Locale ID by adding the following registry value on M-Files Server.

Key location	HKEY_LOCAL_COMPUTER\Software\Motive\M-Files\<version>\Server\MFServer\ExternalFileMonitor	
Value name	DateLocaleLCID	
Value type	DWORD	
Description	Locale ID	
Default value	0409	Locale ID number. English- United States is the default.
Valid values	Any Locale ID. See full list at https://msdn.microsoft.com/en-us/goglobal/bb964664.aspx . Examples:	
	0409	English – United States
	0809	English – United Kingdom
	040b	Finnish

5. Applies to

- M-Files 10.0.3911.100 and later M-Files 10.0 builds
- M-Files 10.1.3915.55 and later M-Files 10.1 builds
- M-Files 102.3920.101 and later M-Files 10.2 builds
- M-Files 11.0.4300.29 and later M-Files 2015 builds
- M-Files 2015.1 and later

6. Change History

DATE	VERSION	DESCRIPTION
2014-06-24	1.0	Initial published version
2015-11-17	1.1	Chapters 4, 5 and 6 added
2016-09-13	1.2	Clarified chapter 4.1