Metadata Extraction

Voyager supports metadata extraction from standard XML documents using XPath queries for standard metadata specifications. You can also create your own XPath queries and map them to searchable field names within Voyager's index.

To access the Metadata Extraction, go to **Manage Voyager > Discovery > Pipeline > Metadata Extraction**.

There are three sections on the Metadata Extraction page:

* **Normalize Namespace Prefix**
* **Selector**
* **Element Selector**

 Normalize Namespace Prefix

**Testing Your Mapping**

To map the fields, configure these parameters:

**Choose the Selector**: This specifies XPath query to a specific metadata record element to be selected.

**Enter the Field Name**: This is the target field in Voyager that gets mapped to the specified metadata output.

**Confirm the Type**: This refers to the data type of the field name. For example, if field name is set to “name”, data type automatically gets set to “text.”

**Choose an Action**: Users can select from five different functions:

Set Field— Assigns value to the specified field

Append Field—Adds to/modifies pre-existing field

Set Geo—Sets a geographic bounding box based on coordinates specified in the metadata

Expand Geo—Expands geographic bounding box from previously set coordinates

Add Link—Helps point field to a URL

**Converter**

          Converter settings are optional and if the user does not specify one, Voyager, by default, assigns an appropriate converter to the field.

Bbox-- Converts bounding box values contained in the XML document

Gml\_Geometry-- Converts geometric (line, circle etc) coordinates from the XML document

Date-- If the Date field is represented as a string value in the XML document, this  converts it into a standard date format

String256-- Finds a String within the element with a maximum length of 256 characters

String512-- Finds a String within the element with a maximum length of 512 characters

StringValue-- Finds a String (of any length) within the element

 6**. Properties**

Required: Checking this box validates the field being extracted from the XML document.

Skip if Exists: If a field has been previously added, checking this box ensures that a duplicate field does not get added to your  list.

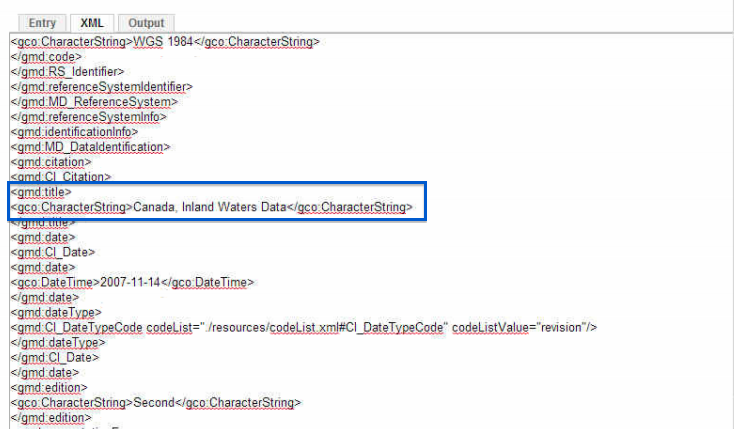
Warn On Replace: If a destination field already exists, checking this field flags the newly set field value.

**Using the XML Box**

The XML box allows you to enter in an XML document to test your XPath queries to paired elements.

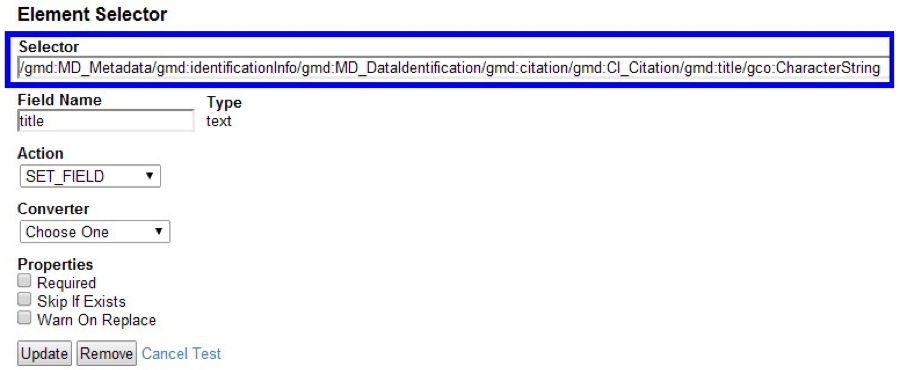
**Step 1**: Click the XML tab and paste the contents of a valid XML document here. Click Save to save the XML contents.

In this case, the element we want extracted from the XML tab is **Title**.



**Step 2**: Specify values for ***Selector***, ***Field Name*** and ***Action***.

Since we want to extract the field **Title**, we copy the XPath Query from the XML document in the Selector box.  “/gmd:MD\_Metadata/gmd:identificationInfo/gmd:MD\_DataIdentification/gmd:citation/gmd:CI\_Citation/gmd**:title/**gco:CharacterString**.**”



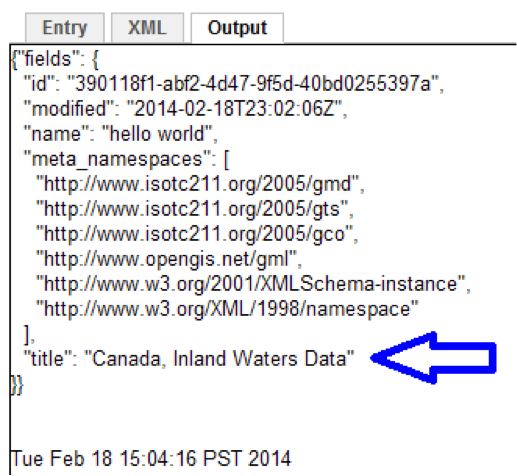
**Step 3**: Specify the corresponding **Field Name** to be mapped to the queried element. Voyager automatically detects the **Type** for the **Field Name**.

For example, here the Field Name is **Title,**whose field type is **Text**.

*Note: when selecting a field name you'll need to either select an existing field name or you can also enter a custom field name as long as it uses a prefix "meta\_", "id\_".*

**Step 4**: Click **Test. T**he extractor searches the XML document for the queried metadata element, and retrieves the value for the field**Title**. The results are presented in the **Output** tab.

In this specific example, **Canada, Inland Waters Data**, which is the value for the **Title** query. is retrieved from the XML tab and displayed in the **Output** tab. Users can use this output result to search for XML documents through Voyager's search UI.



**Step 5**: Click **Save** to add the XPath query to the list.

pic5.png

Click the **Edit**link to make changes to an existing Selector.

Use the up or down **Arrows** help change the order of a Selector.

Select the **[X]** to delete an existing Selector.