

AtoM 2 TARO XSLT Conversion (atom2taro)

About

AtoM to TARO XSLT Conversion, atom2taro, facilitates XML export from the Access to Memory (AtoM) Archival Management System to a compliant XML import into Texas Archival Resources Online (TARO) system. The stylesheet converts the DTD-based 2002 EAD XML export from AtoM and conforms the data to comply with TARO best practices and schema-based XML template.

Created by Sandra Yates, McGovern Historical Center, Texas Medical Center Library, July 14, 2021. The code is maintained and updated to address development changes in AtoM and/or TARO. Updates can be found at <https://github.com/sane8s/atom2taro>. Collaboration and recommendations are welcome. More details about the archival systems at <https://txarchives.org/> and <https://www.accesstomemory.org/>.

The code is based on the dtd2schema.xsl EAD DTD to Schema XSLT Conversion Version 200701 Alpha. Editors: Stephen Yearl (Yale) and Daniel V. Pitti (Virginia). There is also modified code from ead-schema-to-dtd.xsl, developed by Woodson Research Center, Fondren Library, Rice University, that converts <c> tags to <c##> hierarchy.

Customize for your repository

You can set the variables in the stylesheet to match your repository's specifications. (Note: look for XXXX to change values)

- \$TAROID
 - Unique 5-digit TAROID identifier for each finding aid is inserted into the converted file using the Alternate Identifier field in AtoM. Modify the @label attribute to match your system settings. If you capture TAROID using another field, review the exported XML file to determine the path to value.
 - Another option is using the @type attribute
- \$repositoryUrl
 - Repository URL: Insert your repository URL
- \$taroUser
 - \$taroUser: Insert your repository agency code, [link to list]

Other values and attributes to modify for your repository

Note: approximate line locations

Repository Variables

Default repository variables divide repository corpname and subarea using the comma separator based on the repository field in AtoM. This may vary for your repository. You can use the option to insert text directly by switching comment tags.

- @repository (line 288)
- @repositorySubarea (line 289)
- @repositoryCorgname (line 290)

Agency Codes

- <eadid countrycode="US" mainagencycode="US-XXXX"> (line 296)
- <unitid repositorycode="US-XXXX"> (line 338)

A few more available options to review and modify

- <origination>, <controlaccess> Set default @source attribute or uncomment and transform both to choose for each finding aid
- empty @type="bulk" <unitdate> tag. Manually adjust after atom2taro transformation. There are no distinguishing attributes in <unitdate> in AtoM export to automatically assign bulk or inclusive.
- <unitid> Change @repositorycode to match your repository [list of repo codes]
- <language> Default English. Modify final XML for each collection. Development: needs for-each loop to list multiple languages in a collection.
- Some <archdesc> child tags are empty. Copy and modify based on templates above if you have data

How to Use

1. Download and save atom2taro.xsl and TARO-clean.xsl to an accessible file location for staff.
2. Customize atom2taro.xsl for your repository according to the options outlined above.
3. Change 'repo-' to your repository abbreviation to indicate that it is your customized version.
4. Create transformation scenarios in your XML editor (like, Oxygen)
5. Export XML file(s) from AtoM site and save file(s) to computer or accessible location for staff.
6. Open exported XML file in XML Editor (like Oxygen)

7. Run atom2taro
8. After atom2taro, review line 38. Modify if a <unitdate> tag should be bulk or inclusive dates
 - a. <unitdate label="Dates (Bulk):" type="bulk" encodinganalog="245\$g" era="ce" calendar="gregorian" normal="0000"/>
9. Do not validate the XML between transformation scenarios
10. Run TARO-clean (optional)

Finishing Up

Note: line locations refer to the final XML file, and they are approximate.

===line 2

Delete <ead> [In Oxygen, the close tag is often automatically removed]

===line 4

Review <eadid>

<eadid countrycode="US" mainagencycode="US-XXXX">urn:taro:XXXX.00000</eadid>

If no 00000, enter 5-digit taro# at end of urn:taro:XXXX._____

===line 30

Find

<archdesc level="collection" relatedencoding="ISAD(G)v2">

Replace

<archdesc level="collection" type="inventory" audience="external">

===line 47

Find

:ns0="xmlns:xlink" ns0:

Replace

:

===line 35

Review <origination> to determine appropriate @source, lcnaf or local. Delete extra <*name> tags

===line 75

Review <controlaccess> to determine appropriate @source.

===bottom of doc (optional)

Collections with uploaded finding aids

Add code at the end of the <archdesc> tag at the bottom of the document (modify URL for collection):

```
<dsc type="combined">
  <head>Detailed Description of the Collection</head>
  <p><extref xmlns:xlink="http://www.w3.org/1999/xlink" xlink:type="simple"
xlink:show="new" xlink:actuate="onRequest"
xlink:href="https://XXXX.edu/downloads/[collslug].pdf">A preliminary inventory of the collection
is available.</extref> </p>
</dsc>
```

Common validation errors and fixes

- <ead> delete extra tag and close tag.
- @normal in <unitdate> had invalid formatting, like "undated", "2000/09/09". Validate to match formatting, 0000, yyyy, yyyy/yyyy, yyyy-mm-dd/yyyy-mm-dd, yyyy-mm, etc.
- <extref> xlink namespace. Delete :ns0="xmlns:xlink" ns0:

Contact: Sandra for support.