Archivists' Toolkit Inventory Import Tools: Instructions

The Metropolitan Museum of Art Archives, August 2011

Overview

The following instructions describe a suite of simple tools designed to import collection inventories into the Archivists' Toolkit. This system, in use at The Metropolitan Museum of Art since 2009, primarily responds to the long-established procedure by which record-creators oversee the preparation of folder- or box- level inventories prior to the transfer of physical records to the Archives. The import tools enable Archives staff to transform these inventories into resource records within the Toolkit without the hassle of rekeying. Once in the Toolkit, inventories are immediately searchable alongside finding aids and other resources, and can be manipulated and revised at a later date when the collections undergo traditional processing. The same system has supported the import of the Metropolitan's backlog of inventories.

The import tools work around the limitations of the Archivist's Toolkit's client-based system, enabling staff outside the Archives—without access to the Toolkit—to create useful content. Because there is virtually no lag-time between the physical transfer of records and the accessibility of inventories to Archives staff members, intellectual control and service is improved. The import tools utilize Excel and Notepad, which most staff members and interns are comfortable using. No knowledge of EAD or XML is required.

For further information about the import tools, email Adrianna Del Collo, Archivist, The Metropolitan Museum of Art Archives, at adrianna.delcollo@metmuseum.org.

Preparing the Inventory Spreadsheet

Legacy and newly-created inventories come to us in a variety of formats: Microsoft Word or other text files, spreadsheets, or analog, typewritten formats for which no electronic files exist. In order to use the import tools, each inventory must be consolidated into one simple spreadsheet column of alpha-numeric characters. All the text in the column will import into the Toolkit's Title field and be given a File (aka, folder) level designation. Since the objective is to import useful, searchable entries for administrative use, it is permissible and encouraged to import as much information into this field as is practicable (e.g., box and folder numbers or dates). At a later stage, the information can be shifted to the appropriate Toolkit fields.

A number of measures may need to be taken to prepare the spreadsheet for the import process: in the case of analog inventories, scanning and OCR conversion may be required; formatting may need to be added or removed; and separate columns in spreadsheets may need to be joined into one. It would be impossible to account for all possible scenarios, but there are a few tools that are generally useful for preparation of the spreadsheet; the "Find and Replace" feature of Microsoft Word, Excel, and Notepad, including the special characters options in Word; copying and pasting into notepad (to get rid of hard-to-detect formatting); and concatenating columns in Excel. Once the inventory has been edited into a single column of an Excel spreadsheet, ensure that it is sorted in the desired order as this is the order in which it will appear in the Toolkit after the import.

EAD encoding does not recognize certain symbols or diacritical marks as text, and the presence of these symbols in the inventory can cause errors in the XML file. The most common symbols used in folder names that can create import errors are the ampersand, which often appears in the authorized form of corporate names; non-straight quotation marks and apostrophes; and certain kinds of dashes or hyphens, which may, for example, appear in date spans. If such symbols are present, conduct a "find and replace" action to replace the problematic symbols (copied and pasted directly from the inventory) with the corresponding HTML number codes. For example, ampersands (**&**) are replaced with **&**, quotation marks ("") are replaced with **"**, apostrophes (') are replaced with **'**, and dashes or hyphens (—) are replaced with **-**. Always replace ampersands *first* since the codes themselves include ampersands as part of their construction. For additional HTML number codes, see: http://www.ascii.cl/htmlcodes.htm. Many items listed here do not actually cause problems for EAD import, so only use code when necessary. The internet browser will identify the general locations of problematic symbols when the completed EAD-encoded document is opened later in the process, and further edits can take place at that time.

Preparing the Encoded Inventory

Once the inventory has been cleaned up per the above instructions, it can be inserted into the inventory tag template for automatic EAD-encoding by following the steps below:

- Open the file "InventoryTagTemplate.xls" and save as a meaningful name on your desktop.
- Open the folder inventory for the collection at hand, click on the heading of the column that contains the folder names and copy.
- Click on the heading of column "I" in the template, and paste the names into the column.
- Click on numeral 1 in column "H", hold down the shift key and click on numeral 2. Place the cursor on the lower right corner of the box that surrounds the numbers. The pointer will turn into a black plus sign. Click and drag the corner of the box down the column H, even with the last row containing folder names. This action will have "painted" consecutive numbers down column "H"
- Click on the first row of column "A", hold down the shift key and click on the first row of column "G". Place the cursor on the lower right of the box. Again, the cursor will turn into a black plus sign. Drag the corner of the box to the last row containing folder names, painting entries for each column.
- The template is complete. Save the spreadsheet.

Preparing the XML file

Now that each line of the inventory has been properly encoded, the complete EAD document/XML file—including EAD-header information and place-holder collection information—must be created by following the steps below:

- Open the file "EADTemplate.txt" and save as a meaningful name on your desktop.
- In the completed Excel template, click on the heading of column "A", hold down the shift key and click on column "G" and copy.
- Highlight "**[PASTE FOLDER TAGS HERE]**" and paste.
- Save the completed text document and close out of the file.

- Find the text document on your desktop and "save as" and XML file or change the file extension from .txt to .xml. (Note that EAD is a form of XML. Therefore, the file type is XML, but it is referred to in the Toolkit as EAD)
- To validate that the document will successfully import into the Toolkit, open the file in your Web browser (either by opening the browser and selecting "open" from the file menu and selecting the XML file; or, in most cases, double-clicking the XML file will automatically open the browser and display the file)
- Scroll through the document to make sure the complete file displays, ending with the final **</ead>** tag. If there is a problem with the coding or the text entries, the browser will indicate the line or the general area containing the error, and the rest of the document will not display. The most common type of error would be a symbol that cannot be read by the browser. If this is the case, change the .xml file back to .txt and find and replace the symbol with its proper HTML number code, using the following website as a guide http://www.ascii.cl/htmlcodes.htm. Change the file back to an .xml file and test again until the entire document successfully displays in the browser.

Importing the XML file into the Toolkit

The final step in this process is to import the XML file into the Toolkit. The Toolkit user must be set at an access class of 5 in order to import the file.

- Open the Archivists' Toolkit.
- Go to the Import menu and select "Import EAD."
- On the right, select the appropriate repository.
- Find the XML document, select it; then select the appropriate Toolkit Repository and click "Import."
- The folder names will read out as the document imports.
- When the import is complete, double click on resources to see "New Import" listed. Open the resource record and edit the collection information as needed (all of the fields required for the creation of a resource in the Toolkit are set to the default values of: Level=Collection; Title=New Import; Date expression=2009-2010; Language=English; Resource Identifier=NewImport.1; Extent number=100 Linear feet). Note that the resource identifier must be unique, and it is recommended that this field be edited in the Toolkit immediately following import.

Appendix: Template Details

Details about the two templates required for the inventory import process follows. If the templates are unavailable, they can be recreated using the information below.

1. InventoryTagTemplate.xls

The inventory tag template is a Microsoft Excel document that automates EAD encoding of each folder listed in the inventory. Following is a screen-shot of the template, and a key that identifies how each row is entered and provides descriptions of the characteristics the purpose of these entries.

	A	B	C	D	E	F	G	H	I
1	<dsc>	<c id="ref1" level="file">	<did>	<unittitle>[FileName]</unittitle>	</did>	</c>	</dsc>	1	[FileName]
2								2	
3									

Column A. <dsc>
      Identical in each row.
Column B. ="<c id="""ref"&H1&""" level="""file""">"
      Displays as: **<c id="ref1" level="file>**
      This tag identifies the element as a file (folder) and places the entry within a sequence (in this example it is first (ref1)); and draws on the sequential number in column H and inserts it into the text of the tag.
Column C. <did>
      Identical in each row.
Column D. ="<unittitle>"&I1&"</unittitle>"
      Displays as: **<unittitle>[FileName]<unittitle>**
      This tag identifies the name of the file. The spreadsheet draws on the file name that has been pasted in column I and inserts it into the identifying tags.
Column E. </did>
      Identical in each row.
Column F. </c>
      Identical in each row.
Column G. </dsc>
      Identical in each row.
Column H. Sequential numbers in each row, only two numbers need be entered: 1, 2 . . .
      These numbers are "painted down" and correspond to the file name entries that are pasted in as the EAD template is being prepared. This is the only column in the template in which more than one row is populated.
Column I. [FileName]
      File names are pasted in this column from the inventory spreadsheet.

2. EADTemplate.txt

The EAD template is a text file containing EAD code with basic placeholder entries (for those fields required by the Archivists' Toolkit). This template supports collection- and folder- level import only. The tagged folder-level data from the inventory tag template is copied and pasted over the "[Paste Folder Tags Here]" line. After the tagged inventory is pasted into the text document, it is "saved as" an xml file for import.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ead xsi:schemaLocation="urn:isbn:1-931666-22-9 http://www.loc.gov/ead/ead.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:ns2="http://www.w3.org/1999/xlink" xmlns="urn:isbn:1-931666-22-9">
   <eadheader repositoryencoding="iso15511" countryencoding="iso3166-1"
dateencoding="iso8601" langencoding="iso639-2b">
      <eadid></eadid>
      <filedesc>
         <titlestmt>
            <titleproper>
               <num>NewImport.1</num>
            </titleproper>
         </titlestmt>
         <publicationstmt>
            <publisher>Archivists' Toolkit</publisher>
         </publicationstmt>
      </filedesc>
      <profiledesc>
         <creation>This finding aid was produced using the Archivists' Toolkit
            <date>2009-12-23T12:26-0500</date>
         </creation>
      </profiledesc>
   </eadheader>
   <archdesc level="collection">
      <did>
         <unittitle>New Import</unittitle>
         <unitid>NewImport.1</unitid>
         <repository>
            <corpname>Archivists' Toolkit</corpname>
         </repository>
         <langmaterial>
            <language langcode="eng"/>
         </langmaterial>
         <physdesc>
            <extent>100.0 Linear feet</extent>
         </physdesc>
         <unitdate>2009-2010</unitdate>
      </did>
```

[PASTE FOLDER TAGS HERE]

```
</archdesc>
</ead>
```