

Effective Metadata Systems for Archives

EAD and Unified Collection Management Systems

Gregory Wiedeman

University Archivist

University at Albany, SUNY

Archival Metadata Systems

- Paper Finding Aids
- HTML Finding Aids
- EAD records
- Accession Databases
- Unified Collection Management Systems
 - Archivist's Toolkit
 - Archon
 - ArchivesSpace
 - Proprietary or Museum-focused systems

Functions of Archival Metadata Systems

- Creation
- Management
 - Editing & Modifying
 - Enable Administrative functions
- Validation
 - Error-checking
- Public Access
- Aggregation
- Future Migration

Anne J. Gilliland-Swetland, "Popularizing the Finding Aid: Exploiting EAD to Enhance Online Discovery and Retrieval in Archival Information Systems by Diverse User Groups," *Journal of Internet Cataloging* vol. 4, no. 3/4 (2001)

Local Habits Die Hard

- Flexibility of finding aids
- Archives unique, demand unique practices?
- Local practices doesn't mean bad decision-making
 - Unique practices come from unique local context
- Problem was standardizing practices at the local level

Giving Data Structure

- Paper finding aids have no structure
 - Not machine readable
 - Not able to have automated functionality
- Structured data is unitized, has predefined rules
- Necessary to develop tools

- Use of XML to give structure to hierarchical archives metadata
 - EAD Standard

Problems with EAD, a Plural Standard

- Allowance of mixed content and Unstructured Data
 - `<physdesc> 12 cubic ft.</phydesc>`
 - `<langmaterial>This collection is in <language>English</language>.</langmaterial>`
- Failure to completely differentiate storage and display
 - Complains about public accessibility
- Difficulty of validation
 - Inconsistencies in valid EAD files
- Flexibility/Permissiveness
 - Many different ways to encode the same thing
 - Obstructs aggregation, migration
 - Makes overhauling the standard difficult
 - Hinders the development of modern online access tools

Problems with EAD, a Plural Standard

- Allowance of mixed content and Unstructured Data
 - Influenced by TEI
- Failure to completely differentiate storage and display
 - Web community still implementing concept
- Difficulty of validation
 - Problem with XML as a whole
- Flexibility/Permissiveness
 - Key to the wide use of EAD

Elizabeth H. Dow, "Encoded Archival Description as a Halfway Technology," *Journal Of Archival Organization*, 7 (2009)

Still advantages to EAD

- Can be locally standardized with work
- As close to digitally universal as you can get
- Hierarchical structure matches archives metadata
- New tools and cheap processing power to take advantage of XML
- Raised the technological skills of archivists

Unified Collection Management Systems

- Manage Administrative and Descriptive Metadata
 - Specific to archives
- Graphical User Interfaces
- Actually a way to avoid EAD
 - Rely on relational databases
 - Allows for more standardization
 - Can usually export to EAD

Unified Collection Management Systems

- Archivist's Toolkit (no longer supported)
 - Focus on back-end accessioning and creation of description
- Archon (no longer supported)
 - Focus on front-end online access as well as the creation of description
- AtoM (Access to Memory)
 - Front end online access system
- ArchivesSpace
 - Goal to be complete end to end archival metadata system
 - SS code4lib article issue 9, 2010-03-22

Unified Collection Management Systems

- ArchivesSpace
 - Archivist's Toolkit backend
 - Archon public access
- Multiple systems redundant?
 - Combine resources
 - Share development effort
- Model is important, not software
 - One unified system

Unified Collection Management Systems

- Can often be stricter than EAD
- Can abrade with local practices
- May need to make minor adaptations to your local workflow
 - Many arbitrary practices
- Overall, wide use of these tools will standardize practices
 - Not always comfortable
- Often promotes better metadata creation

The Persistence of EAD

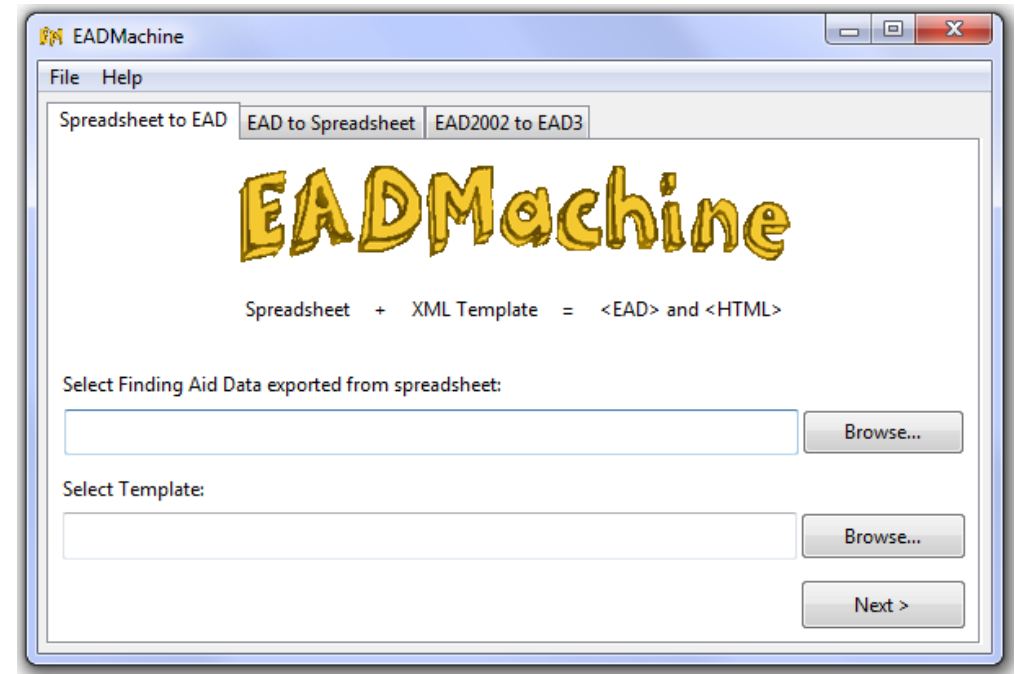
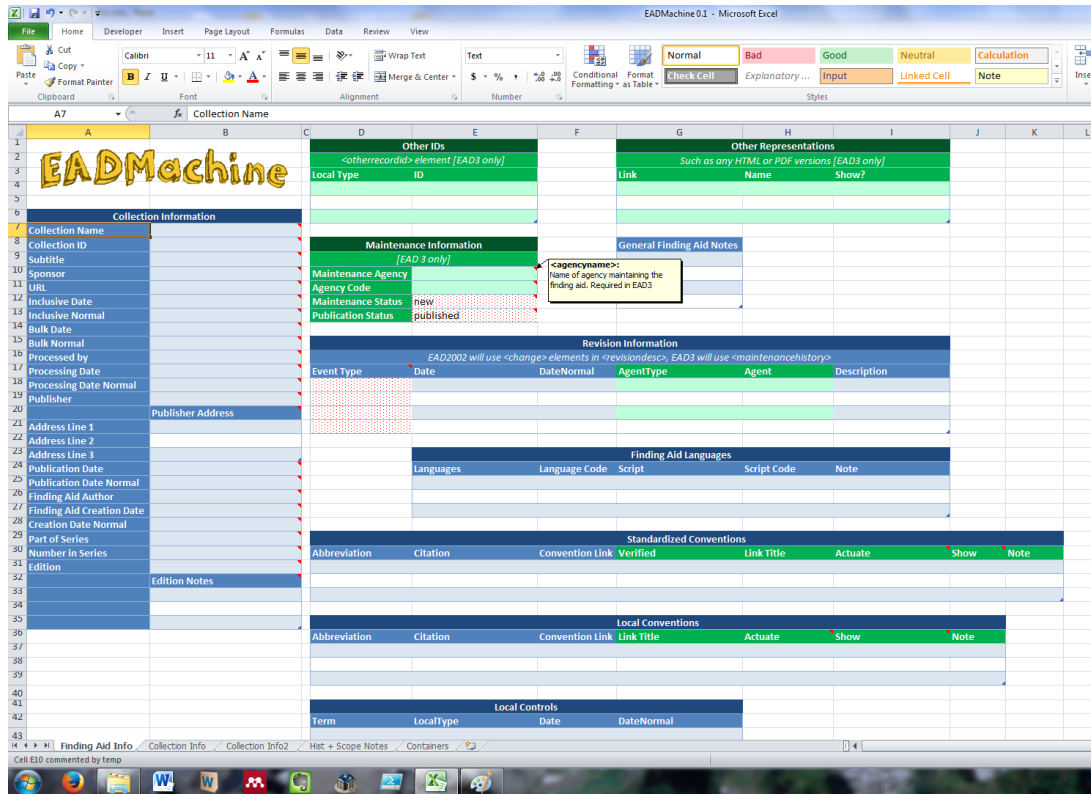
- Effective as Preservation Standard
- Yet EAD is still at the center of complex workflows
 - Automated, not manual
- EAD as a Metadata Dump
 - Unifying systems
 - Single Access point

Developing Better Descriptive Metadata at UAlbany

- No Unified Collection Management Tool
- Finishing full implementation of EAD
- Inconsistent metadata

- Manage metadata before migration

Using EADMachine to Create Consistent EAD

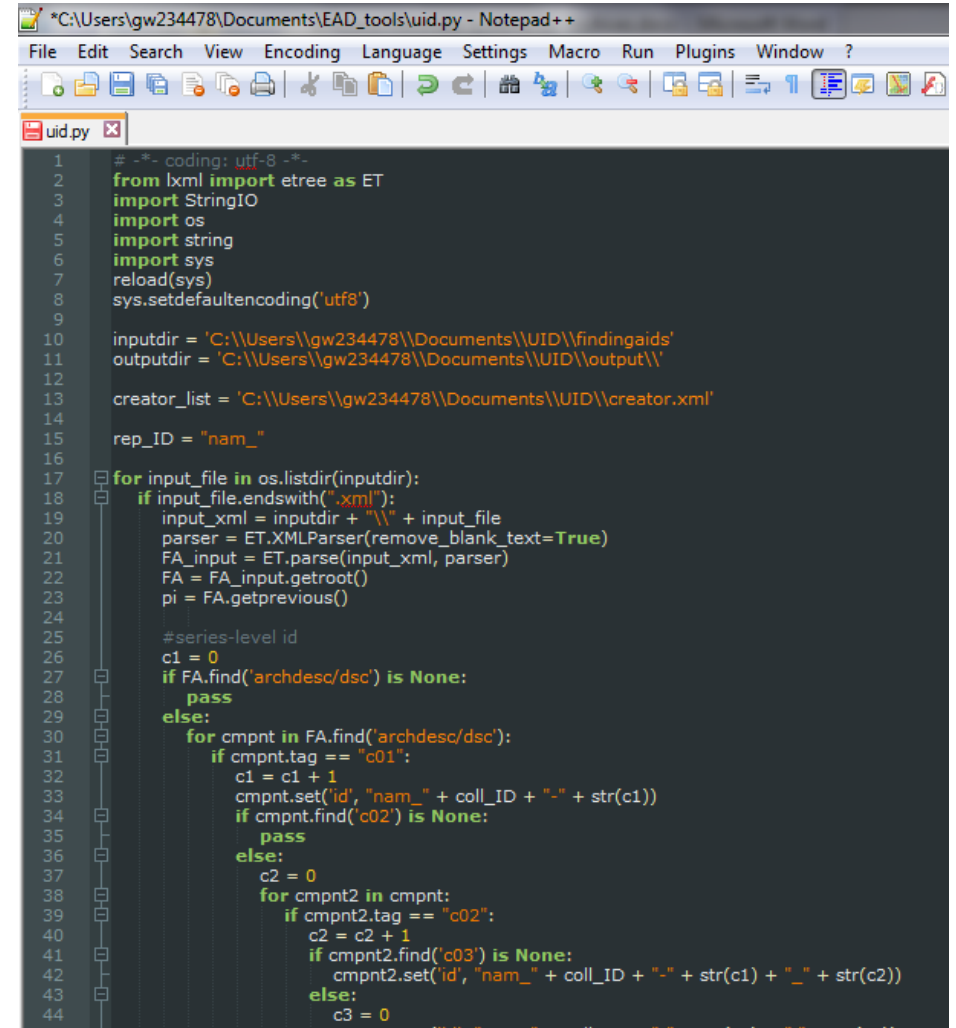


<http://github.com/gwiedeman/EADMachine>

Python Scripts to Manage EAD

- Unique ID System
- Remove special characters
- Standardize encoding
- Remove mixed content when possible
- Automate changes for digital objects
- Rule based-validation
- Separate public access system

http://github.com/gwiedeman/EAD_Tools



```
*C:\Users\gw234478\Documents\EAD_tools\uid.py - Notepad++
File Edit Search View Encoding Language Settings Macro Run Plugins Window ?
uid.py
1  #-*- coding: utf-8 -*-
2  from lxml import etree as ET
3  import StringIO
4  import os
5  import string
6  import sys
7  reload(sys)
8  sys.setdefaultencoding('utf8')
9
10 inputdir = 'C:\\Users\\gw234478\\Documents\\UID\\findingaids'
11 outputdir = 'C:\\Users\\gw234478\\Documents\\UID\\output\\'
12
13 creator_list = 'C:\\Users\\gw234478\\Documents\\UID\\creator.xml'
14
15 rep_ID = "nam_"
16
17 for input_file in os.listdir(inputdir):
18     if input_file.endswith(".xml"):
19         input_xml = inputdir + "\\ " + input_file
20         parser = ET.XMLParser(remove_blank_text=True)
21         FA_input = ET.parse(input_xml, parser)
22         FA = FA_input.getroot()
23         pi = FA.getprevious()
24
25         #series-level id
26         c1 = 0
27         if FA.find('archdesc/dsc') is None:
28             pass
29         else:
30             for cmpnt in FA.find('archdesc/dsc'):
31                 if cmpnt.tag == "c01":
32                     c1 = c1 + 1
33                     cmpnt.set('id', "nam_" + coll_ID + "-" + str(c1))
34                     if cmpnt.find('c02') is None:
35                         pass
36                     else:
37                         c2 = 0
38                         for cmpnt2 in cmpnt:
39                             if cmpnt2.tag == "c02":
40                                 c2 = c2 + 1
41                                 if cmpnt2.find('c03') is None:
42                                     cmpnt2.set('id', "nam_" + coll_ID + "-" + str(c1) + "_" + str(c2))
43                                 else:
44                                     c3 = 0
45                                     for cmpnt3 in cmpnt2:
46                                         if cmpnt3.tag == "c03":
47                                             c3 = c3 + 1
48                                             cmpnt3.set('id', "nam_" + coll_ID + "-" + str(c1) + "_" + str(c2) + "_" + str(c3))
```


Modular Metadata System Design

- Multiple systems
- Move to ArchivesSpace for accession system
- Would need effort to migrate anyway

- Maintain separate public access system
- Possible continue to use EADM Machine for now

Consistency, Consistency, Consistency

- Inconsistent application of tools undermines their use
 - Doesn't matter how good the tool is
- Need to focus on use of tools, than functionality of tools
- Consistency of Metadata, not tools
 - Do we really need one, single unified tool?
 - Benefits of shared resources?

Effective Metadata Systems for Archives

EAD and Unified Collection Management Systems

Gregory Wiedeman

University Archivist

University at Albany, SUNY