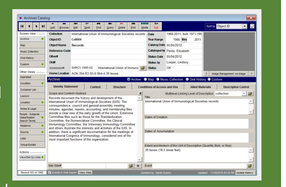
Implementing an EAD-XML compliant workflow at UMBC

In 2015, librarians at the University of Maryland, Baltimore County (UMBC) began work to implement an archival description workflow compliant with Encoded Archival Description (EAD). Existing collection descriptions in PastPerfect, formatted according to Describing Archives: A Content Standard (SAA, 2013), are exported as PastPerfect-XML and then transformed into both EAD-XML and MARC-XML using two XSLT stylesheets.

Why are you using PastPerfect?
PastPerfect is used by several Library departments to track and share contacts, mailing lists, and gift/accession information. Using a separate system for archival description would introduce a silo system just for archives.



What is EAD-XML? Encoded Archival Description (EAD) is a standard format of XML that is used to create finding aids. XML is a mark-up langauge that is used for a document machine-readable - this means that a cstructured data or to make omputer can parse out different sections and identify the sections by the XML tags.

Why is EAD important? Creating finding aids using EAD allows archives and libraries to share their files and make them available in shared catalogs like ArchiveGrid. It also means that the finding aids are somewhat uniform, although how the finding aid is displayed to the user can vary quite a lot.

Single level and multi level description
A single level record includes only one level of description, usually the collection level; this is what we produce in PastPerfect. All archival collections at UMBC have a collection level record and we make these available through PastPerfect Online as an EAD-XML finding aid and as a MARC record.

Processed collections will have a multi level record. This aligns with what we think of in a finding aid - description of the collection as a whole, but also series descriptions and a container list (box or folder level description). Multi level records are only available at UMBC as an EAD-XML online finding aids.

UMBC AN HONORS UNIVERSITY IN MARYLAND

NATIONAL ARCHIVES — NATIONAL HISTORICAL PUBLICATIONS & RECORDS COMMISSION

Lindsey Loeper
lindseyloeper@umbc.edu
@lindseyloeper

Emily Somach
emilysomach@gmail.com

Vicki Sipe
sipe@umbc.edu

Eric Sonnenberg
esonnen@umbc.edu

Github: /UMBC-Library/EAD-XML
EAD finding aids: http://bit.ly/umbcead

Description standardization: At the start of the project, Special Collections had a working Description Manual available for the creation of accession and collection level records for archival materials. Over the course of the project, we have continually updated this manual to reflect new practices and findings. We also developed a data dictionary for each field of data; we ensured that required formatting, content rules, examples, and the location of the data in PastPerfect, EAD, and MARC were defined for each field. UMBC follows DACS/*Describing Archives: A Content Standard* for description of archival collections so, in addition to the dictionary, we created crosswalks from DACS to PastPerfect to EAD and MARC and added them to the Description Manual.

Currently, the standardization of our data ensures compliance with DACS and that the data will work with the export, transformation, and display processes used in later stages.

Provance and acquisition information

PastPerfect field	Guidelines/Rules	Description level
Accruals <accruals>	DACS: 5.1 Custodial History (Added Value) & 5.2 Immediate Source of Acquisition (Added Value)	Collection - Required
	Includes source name, date of acquisition, and information about how the materials were received. Confidential information is not included because this is a public record. This field corresponds to the Provenance and Acquisition Information found in the collection's finding aid (if there is one); you can copy and paste that information into this field. Item and accessional level provenance available in the accession record and should not be duplicated in the item and accession level descriptive records.	Accession - Not in use
PastPerfect screen location		Item - Not in use
Archive > Structure		Include in export?
Description		PastPerfect Online, EAD
Provides the source (provenance) of the materials, as well as the context of the acquisition.		MARC field
Example	EAD tag	n/a
The collection was donated by Susan and Philip Wagner on January 23, 1997 and transferred from Philip Wagner's home in Ridgewood, Maryland.	<archdesc><acqinfo> </acqinfo>	

Arrangement

PastPerfect field	Guidelines/Rules	Description level
System of Arrangement <system>	DACS: 3.2 System of Arrangement (Added Value)	Collection - Required
	All processed collections must have this field completed to the Series level (do not include sub-series). Include series number, series title, and inclusive dates. Unprocessed or accession level records should document any arrangement present when accessioned and/or if any arrangement work was completed when accessioned. May be used for item level records if applicable. Format as a paragraph, not a list. May also be expanded in the finding aid to include sub-series, or to display as a list, if desired.	Accession - Required
PastPerfect screen location		Item - Optional
Archive > Structure		Include in export?
Description		PastPerfect Online, EAD, MARC
Outlines how the collection or materials being described are currently organized.	Assign series with Roman numerals and subseries with capital letters. Example: Series I. Series I.A. Series I.B. Series II. Series II.A.	MARC field
Example	EAD tag	351 $a
Organized into three series: Series I. Personal, 1839-1994. Series II. Sun Papers, 1927-1995. Series III. Boordy Vineyard, 1933-1995.	<archdesc><arrangement encodinganalog="351$a"> Organized into three series: Series I. Personal, 1839-1994. Series II. Sun Papers, 1927-1995. Series III. Boordy Vineyard, 1933-1995. </arrangement>	

Sample field description

DACS to PastPerfect to EAD3

DACS	PastPerfect	EAD
1. Level of Description	Multilevel Linking Level of Description <level>	<archdesc> @level
2.3 Title (Required)	Collection <collection>	
2.3 Title (Required)	Title <title>	<titleproper>, <unititle>
2.4 Date (Required)	Date <date>	<unitdate>
2.4 Date (Required)	Year Range <earlydate>	<fromdate>
2.4 Date (Required)	Year Range thru <latedate>	<todate>
2.5 Extent (Required)	Extent and Medium of the Unit of Description (Quantity, Bulk, or Size) <extent>	<physdesc>; <physdescset>
2.6 Name of Creator(s) (Required, If Known)	Name of Creator <creator>	<origination>
2.6 Name of Creator(s) (Required, If Known)	Other Creators <author>	<controlaccess>
2.7 Administrative/Biographical History (Optimum)	Administrative/Biographical History <admin>	<bioghist>
3.1 Scope and Content (Required)	Scope and Content/Abstract <descrip>	<abstract>
3.2 System of Arrangement (Added Value)	System of Arrangement <system>	<arrangement>
4.1 Conditions Governing Access (Required)	Conditions Governing Access <access>	<accessrestrict>
4.4 Conditions Governing Reproductions and Use (Added Value)	Conditions Governing Reproduction/Copyright <rights>	<userestrict>
4.5 Languages and Scripts of the Material (Required)	Language/Script of Material <language>	<language> @langcode
4.6 Finding Aids (Added Value)	Finding Aids <findaid>	<otherfindaid>
5.2 Immediate Source of Acquisition (Added Value)	Accruals <accruals>	<acqinfo>
6.3 Related Archival Materials (Added Value)	Associated Material in Other Repositories <associate>	<separatedmaterial>
7.1.5 Notes (Added Value)	Citation <cdf222>	<prefercite>
7.1.6 Processing information	Archival history <custodial>	<processinfo>
8.1.5 Archivist and date (Added Value)	Archivist Name <archivist>	<author>

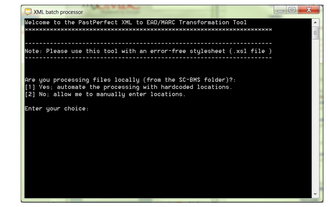
EAD3 crosswalk

Data export: After cleaning up collection level records in PastPerfect and ensuring they adhered to our Description Manual rules, we exported the data using PastPerfect's export function. The result of this export was a single file with all of the record data stored as PastPerfect's own unique flavor of XML.

XSLT and Batch Processor Development: To transform the PPXML into EAD-XML and MARC-XML, we developed two XSLT stylesheets along with a batch processor. We first transformed the PPXML file into an EAD-XML file and a MARC-XML file using the stylesheets and batch processor. We left the resulting MARC-XML file as a single batch file but used the batch processor to split the EAD-XML file into separate files – one for each collection level record.

- PPXML to EAD-XML using XSLT and batch processor
- PPXML to MARC-XML using XSLT, batch processor, and MarcEdit

Subject Access QC and Manual Post-processing: The Catalog Librarian helped us to create guidelines for structuring and formatting subject headings and other access points in the records. Subjects and creator names are formatted to conform with LC guidelines, and the associated punctuation can be used to aid the XSLT in assigning some of the EAD elements, for example <persname> or <corpname>. The subjects mostly map to a default encoding for both EAD and MARC. Access points require the most extensive hands-on review and updating after the initial export.



Batch processor, built-in house

```
<!-- ==========================  Start: If selector is 'creator'  ========================== -->
<!-- ==========================================================================================-->
<xsl:when test="contains($selector, 'creator')">
  <xsl:choose>
    <xsl:when test="contains($first, ', ')">
    <!-- If string contains a comma, treat it as a persname -->
    <!-- ========================================================================= -->
      <persname relator="creator" encodinganalog="100" source="lc">
        <part>
          <xsl:value-of select="replace($first, ', ', '&lt;/part&gt;&lt;part&gt;')" disable-output-escaping="yes"/>
        </part>
      </persname>
    </xsl:when>
    <xsl:otherwise>
    <!-- If string does not contain a comma, treat as a corpname -->
    <!-- ========================================================================= -->
      <corpname relator="creator" encodinganalog="110" source="lc">
        <part><xsl:value-of select="$first"/></part>
      </corpname>
    </xsl:otherwise>
  </xsl:choose>
</xsl:when>
```

Section from XSLT PPXML-to-EADXML.
This section shows how the subjects are processed.

```
</control>
<archdesc level="collection">
    <did>
        <unittitle encodinganalog="245$a">American Society for Cell Biology records</unittitle>
        <unitid countrycode="us" repositorycode="mdubc">Coll008</unitid>
        <origination>
            <corpname relator="creator" encodinganalog="110" source="lc">
                <part>American Society for Cell Biology</part>
            </corpname>
        </origination>
        <unitdate unitdatetype="inclusive" encodinganalog="245$f">1959-1995</unitdate>
        <unitdatestructured unitdatetype="inclusive">
            <daterange>
                <fromdate>1959</fromdate>
                <todate>1995</todate>
            </daterange>
        </unitdatestructured>
        <physdesc encodinganalog="300$a$f">75 boxes (83 linear feet)</physdesc>
        <physdescset>
            <physdescstructured coverage="whole" physdescstructuredtype="carrier">
                <quantity>75</quantity>
                <unittype>boxes</unittype>
            </physdescstructured>
            <physdescstructured coverage="whole" physdescstructuredtype="spaceoccupied">
                <quantity>83</quantity>
                <unittype>linear feet</unittype>
            </physdescstructured>
        </physdescset>
        <abstract encodinganalog="520$a">Files documenting the origins of the American Society for Cell Biology (ASCB),
        <langmaterial>
            <language langcode="eng">eng</language>
        </langmaterial>
        <repository>
            <corpname>
                <part>University of Maryland, Baltimore County</part>
            </corpname>
            <address>
                <addressline>Albin O. Kuhn Library and Gallery</addressline>
                <addressline>UMBC</addressline>
                <addressline>1000 Hilltop Circle</addressline>
                <addressline>Baltimore, MD 21250</addressline>
                <addressline>410-455-2353</addressline>
                <addressline>speccoll@umbc.edu</addressline>
            </address>
        </repository>
    </did>
    <scopecontent encodinganalog="520">
```

Sample completed EAD3 record

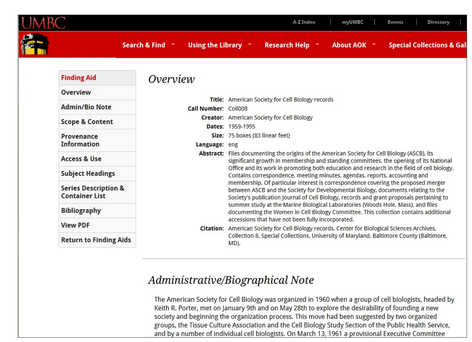
```
000    20090611143810.0
008    071220k1960194xx                    eng d
035    |a ocn184955181
040    |a MUB |c MUB |e dacs |d MUB
049    |a MUBB
1102   |a American Society for Cell Biology.
24510  |a American Society for Cell Biology records, |f 1960-1994.
300    |a 66,000 |f items |a (83.0 |f linear feet)
351    |a Organized into five series: National Office, Meetings, Publications, Public
       Affairs, and Committees.
545    |a Organized in 1960 by members of the Tissue Culture Association, the Cell
       Biology Study Section of the Public Health Service, and other cell biologists,
       and led by Keith Porter; the American Society for Cell Biology (ASCB) was
       incorporated in the state of New York on July 31, 1961. It opened a National
       Office in 1978 in Bethesda, Maryland where it still maintains its headquarters.
520    |a The correspondence, meeting minutes, agendas, reports, accounting and
       membership files provide a clear view of the early development of the
       organization including its partnership with Rockefeller Institute in the
       publication of the Journal of Cell Biology, as well as the proposed merger
       between ASCB and the Society for Developmental Biology (SDB). The growth
       and expansion that led ASCB to open a National Office and begin publishing the
       American Society for Cell Biology Newsletter in 1978 and the journal Cell
       Regulation in 1987 is also clearly outlined. Of particular interest is the large
       volume of correspondence connected with the 1978 by-laws changes. The
       decision not to hold meetings in states that had not ratified the Equal Rights
       Amendment provoked a flurry of letter writing both for and against the
       decision. In addition, there are annual meeting and committee files such as
       those of the Program Committee, Public Policy Committee, the Education
       Committee, the Minorities Affairs Committee, and the Women in Cell Biology
       Committee. Among these files are extensive documents relating to a variety of
       educational programs and student scholarships. Of note is the summer program
       at the Marine Biology Laboratory in Woods Hole, Mass.
524    |a Records of the American Society for Cell Biology. Center for Biological
       Sciences Archives, Special Collections. Albin O. Kuhn Library & Gallery,
       University of Maryland, Baltimore County.
61020  |a American Society for Cell Biology |v Archives.
61020  |a Society for Developmental Biology.
63000  |a Journal of cell biology.
650 0  |a Cellular control mechanisms.
60010  |a Porter, Keith.
650 0  |a Cytology |x Societies, etc.
650 0  |a Women scientists |z United States.
85642  |3 Finding aid |u http://ask.lib.umbc.edu/speccll/ASCBsub/faASCB.php
8528   |b BC-BC |c SP85A |h SARCV
950    |a Ongoing |c 20080418 |h 1500
950    |a Ongoing |c 20071220 |h 1619
```

Sample completed MARC record

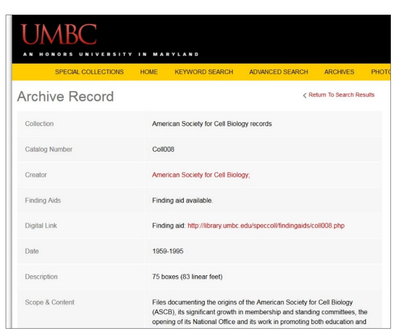
Web Display: We developed an XSLT to transform the EAD-XML files into PHP/HTML. This allows us to display each finding aid as a CSS styled webpage on the Library website. We then created PDF versions of from the PHP for printing and/or offline viewing.

- EAD-XML -> XSLT -> PHP -> PDF

PHP allows us to create a directory page of EAD-XML files which is then available for search engines to crawl or people to download for distribution or research



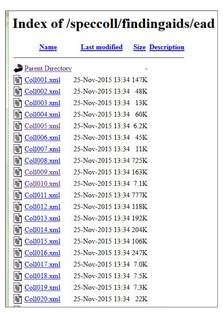
Sample finding aid (PHP)



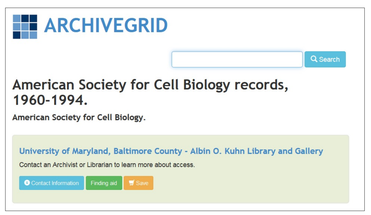
Sample PastPerfect Online record

PastPerfect: The collection level records created in PastPerfect are made accessible through PastPerfect Online: http://umbc.pastperfectonline.com. Links to the EAD-XML finding aids and any related digital collections are included in the record. All records in PastPerfect Online are included in search engine results.

The **MARC records** are added to UMBC's ILS (Aleph) for access through the USMAI library catalog. UMBC uses EBSCO EDS for the catalog discovery layer, and the collection level records are also included in this search. Finally, they are included in OCLC's WorldCat and ArchiveGrid union catalogs.



Directory page for EAD-XML files



Sample from OCLC's ArchiveGrid catalog

Sample from OCLC's Worldcat catalog