

Capturing Metadata for Description of Born Digital Archival Collections

Lara Friedman-Shedlov | Digital Records Archivist | University of Minnesota

The Problem

- What kind of description do we expect for digital materials in archival finding aids?
- How much granularity is needed to be useful?
- Do we simply reproduce the file manifest in the finding aid? That seems like a lot of work!
- Often we resort to describing the entire body of digital material (possibly thousands of files) in one simple statement

Sage Awards (12th Annual), 2015-2016
Panel, 2011-2012
Display Photos, 2005
Display Photos, 2005-2007
Ending of the Sage Awards Press and Cowles Family, 2016
Digital Files, 2005-2016

Digital Files, 2005-2016

File
Performing Arts Archives Sage Awards for Dance records Digital Files, 2005-2016
Scope and Content Note
Digital files include video recordings, photographs, and administrative documents.
Dates
2005-2016
Language of Materials
From the Collection: English
Extent
78 Gigabytes (2,406 files)

- How helpful is this?
- We can link to a file manifest, but it isn't searchable along with the rest of the finding aid

Creating a Simple Tool

Goal: Write a script to capture specific information from the directory tree of the collection in a format easily reusable / repurposable in an archival finding aid

- Extract basic descriptive information from the files themselves:
 - Folder (subdirectory) titles for top two levels of the file directory
 - Size of each folder
 - Number of files in each folder
- Output in two formats:
 - Simple text that can be cut and pasted into narrative description
 - Text marked up with XML as EAD components

Download the Tool

Windows executable (created using PowerShell) freely available at github.com/laradfs/folderlist-tool

- Save the executable somewhere
- Open a command line
- Navigate to the top level of the directory you want to describe
- Run the executable

Finding Aid Display

These examples are from ArchivesSpace

Denise Berceci files

Collection Identifier: Y.USA.89



Kautz Family YMCA Archives | Denise Berceci files

Collection Overview Finding Aid View Container List

Content Description

Presentation slides, project information, data, reports and other records created by Denise Berceci in her role as program specialist for the YMCA of the USA. All files are electronic/digital. The materials reflect her role as a coordinator and synthesizer of information for Y-USA Resource Directors, who in turn consulted with local YMCAs around the country. Project charters

The files are organized into the following folders:

Contact Lists, 689.89 KB (3 files)
Healthy Living, 366.51 MB (288 files)
- Aquatics, 149.93 KB (1 files)
- AR Presentations, 16.79 MB (3 files)
- Cindy McDermott, 9.19 MB (6 files)
- Competency Guide, 2.72 MB (6 files)
- Health and Wellness Task Force, 3.02 MB (19 files)
Healthy Living, 18.26 MB (13 files)

Denise Berceci files

Collection Identifier: Y.USA.89

Kautz Family YMCA Archives | Denise Berceci files

Collection Overview Finding Aid View Container List

Contact Lists, undated and 2016

File
Dates undated and 2016
Extent 689.89 kilobytes (3 electronic files)

Healthy Living, undated and 2014-2015

File
Dates undated and 2014-2015
Extent 366.51 Megabytes (288 electronic files)

Aquatics

File
Extent 149.93 kilobytes (1 electronic files)

AR Presentations

File
Extent 16.79 Megabytes (3 electronic files)

Cindy McDermott

File
Extent 9.19 Megabytes (6 electronic files)

Denise Berceci files
Contact Lists, undated and 2016
▼ Healthy Living, undated and 20...
Aquatics
AR Presentations
Cindy McDermott
Competency Guide
Health and Wellness Task ...
Healthy Living
Membership
Membership Advisory Task...
Multidisciplinary calls
One page outlines
Presentations
SRS
Additional files
Membership by Design_CD, ci...
Overnight Camp Taskforce, 2016

Output 1: Basic Text Output to Screen

```
Contact Lists, 689.89 KB (3 files)
Healthy Living, 366.51 MB (288 files)
- Aquatics, 149.93 KB (1 files)
- AR Presentations, 16.79 MB (3 files)
- Cindy McDermott, 9.19 MB (6 files)
- Competency Guide, 2.72 MB (6 files)
- Health and Wellness Task Force, 3.02 MB (19 files)
- Healthy Living, 18.26 MB (13 files)
- Membership, 55.75 MB (82 files)
- Membership Advisory Task Force, 47.33 MB (17 files)
- Multidisciplinary calls, 58.07 MB (109 files)
- One page outlines, 12.45 MB (8 files)
- Presentations, 123.77 MB (16 files)
- SRS, 1.90 MB (3 files)
- Additional files, 17.11 MB (5 files)
Membership by Design_CD, 55.88 MB (198 files)
Overnight Camp Taskforce, 1.57 MB (23 files)
- Masters, 724.33 KB (8 files)
- Additional files, 884.79 KB (15 files)
Program overview_handouts, 419.62 KB (5 files)
Project Charters, 4.67 GB (2458 files)
- 2008, 14.40 MB (31 files)
- 2009, 63.08 MB (90 files)
```

Simple text is easy to cut & paste

Text may be pasted into a note in an archival finding aid

Output 2: EAD Mark-Up

```
<c01 level="file">
  <did>
    <untitled>Contact Lists</untitled>
    <physdesc>
      <extent altrender="materialtype spaceoccupied">689.89 kilobytes</extent>
      <extent altrender="carrier">3 electronic files</extent>
    </physdesc>
  </did>
</c01>
<c01 level="file">
  <did>
    <untitled>Healthy Living</untitled>
    <physdesc>
      <extent altrender="materialtype spaceoccupied">366.51 megabytes</extent>
      <extent altrender="carrier">288 electronic files</extent>
    </physdesc>
  </did>
  <c02 level="file">
    <did>
      <untitled>Aquatics</untitled>
      <physdesc>
        <extent altrender="materialtype spaceoccupied">149.93 kilobytes</extent>
        <extent altrender="carrier">1 electronic files</extent>
      </physdesc>
    </did>
  </c02>
</c01>
```

Top-level folders are encoded as <c01>

Filename → <untitled>
File size & number → <physdesc/extent>

Second-level folders are encoded as <c02>

Dates for top-level folders added later (see note on left)

EAD components displayed as archival objects (i.e. analogous to a folder list) in archival finding aid

Why not dates, too?

It would have been possible to pull a date from the file directory information as well (and it would be easy to alter the script to include it), but I decided not to grab that data

- Which date? Date modified or date created?
- These dates are rarely correct or useful for our collection description purposes