Best Practices for Data Entry and Optimizing Input for PACSCL Finding Aids Site

Hidden Collections Processing Project

Holly Mengel
3/1/2012
Table of Contents

INTRODUCTION.................................................................................................................................................. 3

COLLECTION LEVEL RECORD IN AT ............................................................................................................ 3

FINDING AID NOTES........................................................................................................................................ 4

ACCESS POINTS, OR NAME AND SUBJECT AUTHORITIES........................................................................ 5

PACSL Stylesheet Quirks: ................................................................................................................................. 6

  Quirk #1: Adding a Repository Code in Resource Identifier is Required ................................................. 6
  Quirk #2: Punctuation automatically added at end of certain fields ......................................................... 7
  Quirk #3: Pre-coordinated Subject Authorities ....................................................................................... 9
  Quirk #4: Authorities and their role in faceted searching: .................................................................. 10
  Quirk #5: Abstract Note—A Necessity! .................................................................................................. 13
  Quirk #6: Numbering Series .................................................................................................................. 14
    1. You can add the series number into the “Component Unique Identifier” field ............................ 14
    2. Or, you can add the series number into the “Title” field ............................................................... 15
  Quirk #7: Oversized Materials ............................................................................................................... 16

PACSL Style Sheet BUGS and (when possible) how to get around them: .................................................. 17

  BUG #1: Multiple Instances for one entry ............................................................................................... 17
  BUG #2: Instances consisting of three container types ....................................................................... 18
  BUG #3: Large pdfs do not work .......................................................................................................... 19
  Bug #4: Multiple notes at the folder level do not display: .................................................................. 19
INTRODUCTION

Creating standardized DACS compliant finding aids that format correctly in the PACSCL Finding Aids Site is not hard, but there are a few tricks discovered by the PACSCL/CLIR team during the course of the project. The project also identified a few bugs with the stylesheet. The following guide will help guarantee that your finding aids display properly and that all data formats correctly.

COLLECTION LEVEL RECORD IN AT

The following information is required to create a collection level DACS compliant finding aid:

- Collection title (example: John Smith papers)

  For full information on DACS compliant collection titles, please see Describing Archives: A Content Standard, 2.3. The type of record (papers, collection, records, photographs, etc.) always follows the creator and is never capitalized.

  A few examples of DACS compliant titles include:

  John Smith Heating Company records
  John Smith diaries
  John Smith collection on Abraham Lincoln
  John Smith civil rights collection
  Alfred Smith collection of John Smith materials
  John Smith family papers
  Smith family papers

- Collection date (example: 1850-1875; or 1850; or 1850-1875, 1888)

- Extent of the collection (example: 14 linear feet)

- Name of creator (example: Smith, John, 1825-1878)

- Conditions governing access (example: see page 4)

- Conditions governing use (example: see page 4)

- Preferred citation (example: see page 4)

- Immediate source of acquisition, if known (example: see page 4)

- Language of materials (example: English)

- Access points (or name & subject authorities) (example: see page 5)
FINDING AID NOTES

These notes should be standard to all other finding aids at a repository. For repositories that did not have pre-existing note templates, the PACSCL/CLIR “Hidden Collections” Processing Project used:

Conditions governing access:

This collection is open for research use.

OR

[Series/Boxes] of this collection are open for research use. However, [series/box/folder] is restricted until [date]. Please contact [name of repository] for additional information.

Conditions governing use:

Copyright restrictions may apply. Please contact the [name of repository] with requests for copying and for authorization to publish, quote or reproduce the material.

Preferred citation:

[Description and date of item], [Box/folder number], [Collection number], [Collection name, collection dates], [Repository].

Immediate Source of Acquisition:

Sometimes this information is unknown and therefore, this is not a required note.

Typically:

Gift of John Smith, 1954
Purchased, September 4, 1855
Bequest of John Smith, 1954

Related Archival Materials: This note simply identifies other collections within your repository or elsewhere that are directly related to the collection. The related collections may be created by the same person, family or business, or may concentrate on the same topic (such as the Sesquicentennial in Philadelphia). Sometimes there are no related archival collections and therefore, this is not a required note.

For example:
Haverford College: John Smith collection, 1855-1910, Mss. 205;
Swarthmore College: James Smith papers, 1840-1888, Collection 1450.
ACCESS POINTS, OR NAME AND SUBJECT AUTHORITIES

Identifying access points for a collection can be difficult. For each access point you wish to add, make sure there is enough information in the collection to justify a name or subject heading. In other words, if a researcher came to use this collection based on the access points you provided, would there be enough on that topic to have made the trip worthwhile?

To provide a DACS compliant “Single-level Minimum” description (collections level record), you must identify at least a collection creator (can be a personal or corporate entity), but you can (and are encouraged to) identify additional access points including:

- Names
  - Creator
  - Person
  - Corporate Body
  - Family
- Subject headings
  - Topical Term
  - Geographic Name
  - Genre/Form

For names and subjects, the PACSCL/CLIR “Hidden Collections” Project uses the Library of Congress Authorities at: [http://authorities.loc.gov/](http://authorities.loc.gov/) (search under names or subjects, as appropriate). The project opted to use only authorized terms from Library of Congress. If the name was not authorized, the project followed DACS formatting rules for the creation of names in the Archivists’ Toolkit and ran the unauthorized names through the University of Pennsylvania’s NACO Funnel, thereby authorizing the names. For genre terms, the PACSCL/CLIR “Hidden Collections” Project used the Art and Architecture Thesaurus at: [http://www.getty.edu/research/tools/vocabularies/aat/](http://www.getty.edu/research/tools/vocabularies/aat/) Again, the project opted to use only authorized terms from the Art and Architecture Thesaurus.

While the Library of Congress and Art and Architecture Thesaurus provide the basics for authority control, it may be necessary to create a central standardized list within your repository:

For example: Native Americans, American Indians, and Indians of North America are ALL authorized terms in the Library of Congress Authorities. If this is a common topic within your collections, you may wish to choose one of those terms and use it throughout your collections.

See the Archivists’ Toolkit manual for instructions on adding and linking name and subject authorities.

See pages for Stylesheet Quirks relating to name and subject authorities on pages 9-13!
PACSL Stylesheet Quirks:

There are 6 quirks unique to the PACSCL Finding Aids Site style sheet to be addressed:

1. Why adding a repository code is required
2. Where punctuation is automatically added at the end of certain fields
3. Why pre-coordinated subject authorities are broken up (and how to work with it)
4. Why careful authority work is essential when facetted searching is used
5. Why an abstract in the note field is required at the collection level
6. How to number series
7. How to handle oversized materials

Quirk #1: Adding a Repository Code in Resource Identifier is Required

The PACSCL Finding Aids Site uses a harvester to locate finding aids from all participating repositories to be displayed on the PACSCL Finding Aids Site. The harvester looks in web folders shared with the University of Pennsylvania, locating each repository’s unique identifier, which helps distinguish the finding aids from one another. To ensure that your institution’s identifier is unique, supply a repository code as the prefix for every resource you create. In the example below, the Independence Seaport Museum has selected the acronym ISM as its unique identifier:

Why this is so important: Archivists frequently use the same type of accessioning/naming scheme for collections, and as a result, it is likely that multiple repositories will have the same resource identifier. For example, it is entirely likely that many repositories will have an accession number: 1966.17. If all repositories choose to use the accession number for the Resource Identifier, the harvester will find more than one file with that same resource identifier, and it will assume they are the same finding aid, and overwrite the first file.

When the finding aid exports, this code will NOT display, so repositories who are not keen on displaying this information should not worry. The screen shot pictured on the next page shows how resource identifier will be displayed.

The PACSCL/CLIR project used the Repository’s acronym (ISM for Independence Seaport Museum; RML for Rosenbach Museum and Library, etc.); however, the University of Pennsylvania uses the MARC list of repository codes (for example, PU-L for Biddle Law Library). You can locate your code here:

http://www.loc.gov/marc/organizations/
**Quirk # 2: Punctuation automatically added at end of certain fields:**

The PACSCL Finding Aids Site stylesheet provides most of the punctuation and other stylistic formatting for the finding aids. Therefore, to avoid duplication of punctuation, you are advised to follow these procedures for data entry.

**Resource Identifier:**

Periods are automatically added between ISM, 1966 and 17 resulting in final formatting: ISM.1966.17 (for how this looks in the PACSCL Finding Aids Site, see page 6)

**Unit title and dates:**

The PACSCL Finding Aids Site stylesheet automatically adds a comma between the title and the date (regardless of where in the collection) and a period after the date. Therefore, this will display:

If there is no date, the stylesheet will NOT add a comma after the title, but will add a period (see blue arrow).
Extent:
Formatting is automatic for information in the container summary field on the collection-level data entry screen:

After formatting, it will display:

---

**Example of a good finding aid**

This is a finding aid. It is a description of archival material held at the scoti. Unless otherwise noted, the materials described below are physically available in our reading room, and not digitally available through the web.

**Summary Information**

Repository: scoti
Creator: Bruckner, Johann Jakob, 1696-1778
Title: Example of a good finding aid
Date (inclusive): 2000-2012
Call Number: Example
Extent: 10 Linear feet (5 cartons)
Language: English
Language of Materials note: The bulk of this material is in German, however, there are 14 items in French.
Abstract: This is a short description of what the collection is about and what is in the collection. Can be only 2 sentences.
Quirk #3: Pre-coordinated Subject Authorities:
To take advantage of the faceted browsing feature, the PACSCL Finding Aids Site is configured to break apart any pre-coordinated subject headings in a finding aid. Although the subject heading displays as entered (for example, “Correspondence—John Wilkes Booth”) on the actual finding aid, it will appear as two facets on the site: “Correspondence” and “John Wilkes Booth.” Therefore, to ensure that these facets are not misleading and appear in the correct category (i.e. Genre/Format for “Correspondence” and Names for “John Wilkes Booth,”) it is recommended to break up pre-coordinated subject headings in AT prior to export whenever possible.

On some occasions, there is no alternative, but the display is less than ideal. For example, the authorized term for breast cancer is “Breast—Cancer.” Because of the breaking apart of pre-coordinated subject authorities, “Breast” displays as a stand-alone subject authority which is quite misleading.
Quirk #4: Authorities and their role in faceted searching:

Because the name and subject headings have a dual purpose (secondary being the faceted searching), it is particularly important that data entry is precise: adding a period or an extra space between words or capitalizing words that should not be capitalized will create an extra subject heading, resulting in the same heading being displayed multiple times. As a result of this, a researcher, clicking on one of these terms, will not locate all collections in the system having to do with the term.

The Stylesheet will also separate your authorities into types: Creator; [All] Names; Subject: Corporate Names; Subject: Person; Subject: Topic; Subject: Place; and Form/Genre. This means that you must be careful to add your names and subject with the appropriate designation or “Type” in the Archivists’ Toolkit:
Names:

American Red Cross Example
The American Red Cross has been entered into multiple repositories’ Archivists’ Toolkits in three different ways, which means it displays in three different places: A researcher, clicking on any one of these three terms, will not locate all collections in the system having to do with the American Red Cross. In fact, they get only 1 or only 4 of the collections with a facet for the American Red Cross.

This is CORRECT! American Red Cross has been designated as a corporate name and displays within the facet titled: “Subject: Corporate Name.” Four collections have been correctly designated.
This is INCORRECT!

While it is entered as American Red Cross (an authorized heading), it has been designated as a Topical term (650) and is therefore displayed within the facet “Subject: Topic” and only 1 collection is identified.

This is INCORRECT!

Here it is entered as Red Cross (NOT an authorized heading), and it has been designated as a Topical term (650) and is therefore displayed within the facet “Subject: Topic” and only 1 collection is identified.
Genre/Form; Geographical Name; and Topical Term

Quirk #5: Abstract Note—A Necessity!

The PACSCL Finding Aids Site search results page includes an abstract or a small blurb allowing a researcher to learn the most basic information about the collection. This information is pulled from the “Abstract” note at the collection level in the Archivists’ Toolkit. If you do not create an abstract, there will be no further information regarding the collection on the search result page.
Quirk #6: Numbering Series

You can number series in two ways.

1. You can add the series number into the “Component Unique Identifier” field

A period will automatically be added after the “I” in the finding aid.

NOTE: If you select this method, you must select “Number component levels” after clicking “Export EAD” in the Archivists’ Toolkit and before saving into your web folder. Clicking this box ensures that this data is exported.

This method displays:
2. Or, you can add the series number into the “Title” field

This is the method that the PACSCL/CLIR “Hidden Collections” Project used. For finding aids with complicated hierarchy, this method displays most clearly during data entry:
**Quirk #7: Oversized Materials**

Using “Oversize” in the instance results in the finding aid formatting like this:

“Oversized” should not be used in the instance. Instead, place “[oversized]” in the “container summary” field in The Archivists’ Toolkit:

Your title will now be formatted on the Finding Aid Site as follows, and will not disrupt the box and folder display:
PACSL Style Sheet BUGS and how to get around them:

There are 4 bugs in the PACSL Stylesheet:
1. Multiple Instances for one entry display incorrectly
2. Instances consisting of three container types display incorrectly
3. Large pdfs do not work.
4. Multiple notes at the file level do not display.

**BUG #1: Multiple Instances for one entry:**
Archivists’ Toolkit allows you to add multiple instances for one entry (or folder title); however, the PACSL Stylesheet actually changes the order of the instances and, as a result, makes the data incorrect!

Data entered into AT:

Data displayed on the Finding Aid Site (notice, as displayed: Box 1 contains folders 1-2 and Box 2 contains folders 4-5):

To get around this bug, you need to add two entries that are identical with separate instances:

- To John Smith, 1956, Box 1 Folder 4-5
- To John Smith, 1956, Box 2 Folder 1-2
**BUG #2: Instances consisting of three container types:**

Archivists’ Toolkit allows you to add up to three container types; however, the PACSCL Stylesheet does not display properly.

Data entered into AT:

![Image of Data Entry Interface]

This displays:

![Image of Displayed Data]

To get around this, we have combined the first two container types:

![Image of Combined Container Types]

Which displays:

![Image of Displayed Combined Data]
**BUG #3: Large pdfs do not work**

This is what you will see when you click on a pdf for a large finding aid (small finding aids from the same repository do not have the same problem, indicating that the server times out before the entire pdf can be generated). There is currently no work-around for this problem; however, the University of Pennsylvania has been informed of the problem.

**Bug #4: Multiple notes at the folder level do not display:**

AT allows you to add notes and names and subject authorities to any level of the finding aid, however, the stylesheet is not built to accommodate such detailed cataloging. The example below demonstrates what happens when multiple notes are included at the file level:

To get around this, it is recommended to avoid including more than one note at the file level. Include all necessary information into a single more general note at the collection or series levels! The PACSCL/CLIR project generally did not add notes to the folder level.