The University of Kansas Libraries has moved from an archival finding aid creation system that lacked authority control to ArchivesSpace, an open-source archives information management system, which has some limited authority control capabilities. This presentation will discuss the benefits and challenges associated with evaluating and updating the close to 5,000 subject headings generated from our legacy finding aid (i.e. EAD) records, including the use of an Authorities Unit staff member more familiar with MARC cataloging.

Brief History of EAD at KU

- Started use Encoded Archival Description (EAD) in 2004/2005
- Retrospective conversion project
  - Scanned paper finding aids (over 2,500)
  - Inserted into template created in NoteTab Pro
  - Work done primarily by students
- XML file created in NoteTab Pro
- Converted to HTML (2004-2008) for website
- Currently use XTF (2009-)

We'll start with a brief history of EAD processing at the University of Kansas. We began encoding finding aids in EAD starting in late 2004 and developed two workflows, one for newly processed collections and the other was a project to convert the existing paper finding aids into EAD.

For new collections, staff & students used the EAD Cookbook in NoteTab Pro to create xml files that were then converted to HTML and published on our website. We implemented XTF in 2009 which allowed us to skip the convert to HTML step (but no time was actually saved since we also started assigning handles and using version control software)

The retrospective project involved scanning the paper finding aids and saving as a Word document. Using a macro in NoteTab Pro to step through each section, students copy/pasted information from the Word document to build an EAD record.

Brief History of EAD at KU

UNIVERSITY OF KANSAS LIBRARIES

KENNETH SPENCER RESEARCH LIBRARY

Guide to the George Walter Rayhill Collection

Papers

COLLECTION SUMMARY

Creator:	Rayhill, George Walter, 1848-1930
Title:	Papers
Dates:	[185-]-1952
Quantity:	1/2 linear foot
Call Number:	RH MS 82
Call Number:	RH MS-P 82
Repository:	University of Kansas
	Kenneth Spencer Research Library
	Kansas Collection
	1450 Poplar Lane, Lawrence, KS 66045-7616
	Phone: (785) 864 - 4334
	Fax: (785) 864-5803
	URL: http://spencer.lib.ku.edu
	pg May 1975
	mg 2004

<click> As you can imagine, this project resulted in a lot of bad data due to poor OCR scans, lack of standardization, students pasting information in the wrong fields or not pasting information at all, etc. In this example, the title is very generic and the quantity uses the fraction ½ instead of .5
Currently on our website we have 3,536 finding aids and more than 2,000 of these are from the recon project.

Authority Control?

- Newly processed materials -- Used LC subject and name headings set up and/or verified by a cataloger
  - EADs are converted to MARC for addition to OCLC
- Recon Project – Paste whatever was in the finding aid
- Similar to card catalog
  - Headings valid at the time of entry
  - Text strings, not links – increase in potential typos
  - No cross references
- EAD records filled out by students and staff not familiar with MARC encoding

KU LIBRARIES
The University of Kansas

4

An additional part of the EAD workflow for new materials was to create a MARC record for the collection. Instead of training Processing staff and students in the intricacies of AACR2, MARC and LCSH, a monograph cataloger was given the responsibility of converting the EAD record into a MARC record and adding to OCLC. The cataloger set up the headings suggested by the processors following LCSH, AACR2 (and now RDA) rules. For the retrospective project, the students just pasted whatever was in the paper finding aid.

So, the heading may have been valid at the time the finding aid was created (if the typist didn't introduce any typographical errors) but no ongoing authority maintenance was ever performed on them.

Authority Control?

```
<controlaccess id="a12">
    <head>INDEX TERMS</head>
    <controlaccess>
        <head>Organizations:</head>
        <corpname role="subject" encodinganalog="710" source="local">Mother's League (Lawrence, Kan.)</corpname>

    </controlaccess>

    <controlaccess>
        <head>Subjects:</head>
        <subject encodinganalog="650" source="local">Lawrence (Kan.) Social life and customs.</subject>
        <subject encodinganalog="650" source="local">Women. Kansas. Lawrence. Societies and clubs.</subject>
        <subject encodinganalog="650" source="local">Child rearing. Kansas. Lawrence. Twentieth Century.</subject>
    </controlaccess>

</controlaccess>
```

<click> Since in both workflows, the records are created by people unfamiliar with MARC, encoding errors were introduced that didn't make a noticeable difference in how the records display on our website but they do make a significant difference in a more complex system. For instance in this example, though technically a subject, the heading Lawrence (Kan.)—Social life and customs is actually a place name subdivided by a topical heading. This was encoded as a 650 rather than a 651.
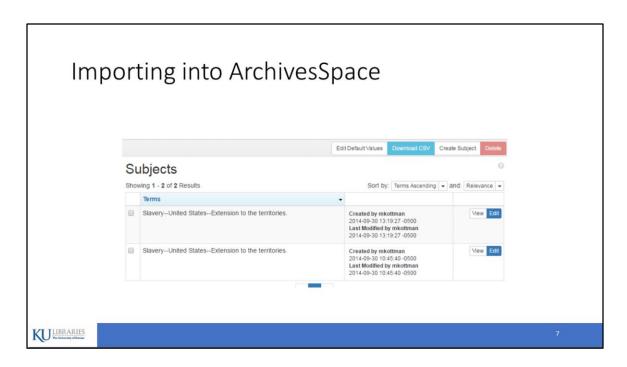
Importing into ArchivesSpace

- Imported EADs over multiple versions of software
  - Early versions did not recognize when duplicate headings were created
  - Later versions recognized duplicates
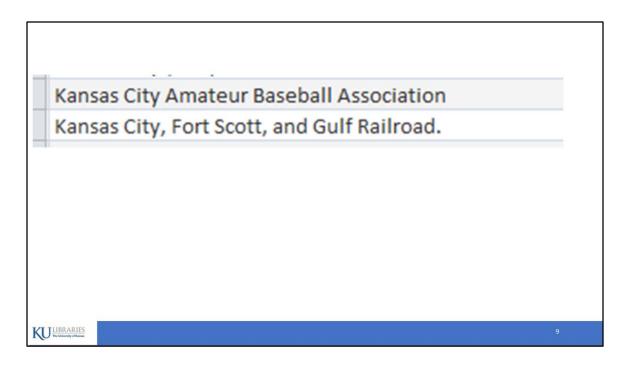- Duplicates assumed if source, type and term the same

6

Like most institutions that have implemented ArchivesSpace, we had a lot of data to cleanup before we could import our legacy records into ArchivesSpace. We worked on this cleanup in batches so the importing of our EADs was spread out over several versions of ArchivesSpace. The earliest versions of Aspace created new agent and subject records regardless of if there was a matching record already in the system.
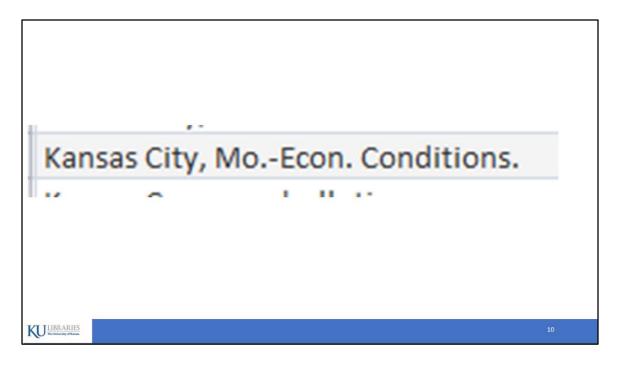
Importing into ArchivesSpace

Subjects

Showing 1 - 2 of 2 Results

Terms			
☐ Slavery--United States--Extension to the territories.	Created by mkottman 2014-09-30 13:19:27 -0500 Last Modified by mkottman 2014-09-30 13:19:27 -0500	View Edit
☐ Slavery--United States--Extension to the territories.	Created by mkottman 2014-09-30 10:45:40 -0500 Last Modified by mkottman 2014-09-30 10:45:40 -0500	View Edit

<click> Fortunately, Aspace had added duplicate checking before we imported the bulk of our records.

Bad data!

- Agents as Subjects (600/610s as 650s)
- Abbreviations of subdivision:
- Obsolete headings
- Punctuation

Kansas City (Mo.)
Kansas City (Mo.) - Buildings, structures, etc.
Kansas City (Mo.) -- Description and travel
Kansas City (Mo.)--Buildings, structures, etc.
Kansas City (Mo.)--History--Sources.
Kansas City (Mo.)--History.
Kansas City (Mo.)--History.
Kansas City (Mo.)--Pictorial works.
Kansas City (Mo.)--Social life and customs--Sources.
Kansas City Amateur Baseball Association
Kansas City, Fort Scott, and Gulf Railroad.
Kansas City, Kansas. Bernard Corrigan House.
Kansas City, Kansas. Huron Cemetery.
Kansas City, Kansas. J.A. Hoel House.
Kansas City, Kaw Valley & Western Railway--Records and correspondence.
Kansas City, Missouri
Kansas City, Missouri--History.
Kansas City, Missouri--Industries.
Kansas City, Missouri. Boley Building.
Kansas City, Missouri. Dean Brothers Realty Company building.
Kansas City, Missouri. Louis Curtiss office building.
Kansas City, Mo.--Social policy.
Kansas City, Mo.--Statistics.
Kansas City, Mo.-Econ. Conditions.

KU LIBRARIES
The University of Kansas

8

I used Microsoft Access with an ODBC connection to our ArchivesSpace backend to get a list of the 4,825 headings in ArchivesSpace designated as subject headings. As those of you in the front row can see, this screenshot of the results for Kansas City amply demonstrates why our records could benefit from a thorough review and regular authority maintenance.

Kansas City Amateur Baseball Association

Kansas City, Fort Scott, and Gulf Railroad.

<click> For example, these headings are corporate bodies which in ASpace should be agent records.
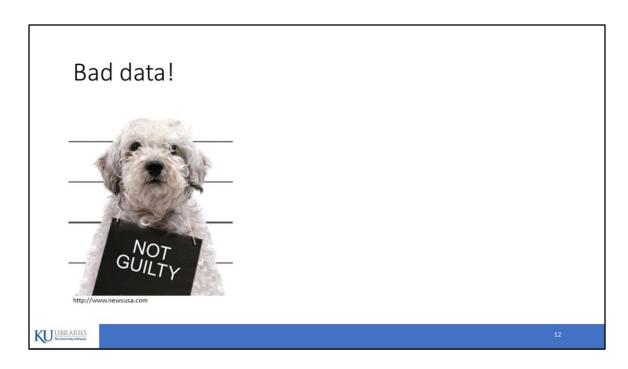
<click> We also had a lot of headings that abbreviated the subdivisions like you'd find on old catalog cards. There are also obsolete subdivisions like "Description – Views" or Industries used after place.

Kansas City, Missouri--History.
Kansas City, Missouri--Industries.
Kansas City, Missouri. Boley Building.
Kansas City, Missouri. Dean Brothers Realty Company building.
Kansas City, Missouri. Louis Curtiss office building.
Kansas City, Mo.--Social policy.
Kansas City, Mo.--Statistics.

KU LIBRARIES
The University of Kansas

<click> Of course, the most glaring issue is the lack of standardization. In this example we have Missouri spelled out or abbreviated when it should be abbreviated and enclosed in parenthesis.
Processing 4,825 headings would be very time consuming for someone who is not experienced with authority maintenance or familiar with the conceptual changes to subject heading over time. Fortunately, since the authority maintenance folks report to me too, I was easily able to assign this project to our subject heading processor.

Bad data!

NOT GUILTY

http://www.newsusa.com

<click> Since ASpace is a web-based system, all that was necessary was to set up an account and provide some training. The training covered not only how to use ASpace, but also a mapping of ASpace/EAD terminology to the MARC terms and fields the processor was more familiar with. For example, once it was explained that the Resource Record in ASpace could be considered the bib record, and the Subject Record was like a subject authority file record, the training went much quicker.

The Fix is In!!

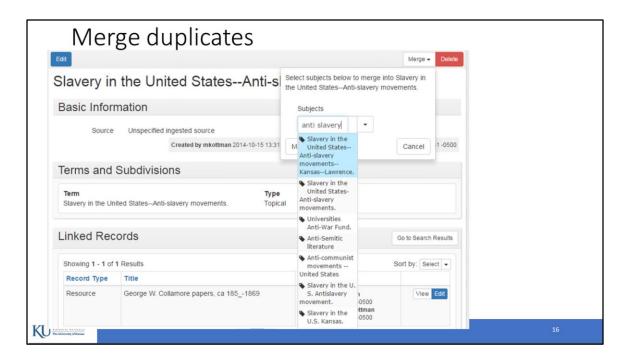| |
|---|
| Slaveholders -- Missouri -- Jackson County. |
| Slaveholders -- Missouri -- Monroe County. |
| Slavery -- United States |
| Slavery in the U. S. Antislavery movement. |
| Slavery in the U. S. Extension to the territories. |
| Slavery in the U. S. Fugitive slaves. |
| Slavery in the U.S. Extension to territories. |
| Slavery in the U.S. Extension to the territories. |
| Slavery in the U.S. Kansas. |
| Slavery in the U.S. Missouri. |

The review and editing process was very similar to how we verified headings in our ILS. We have all of LC's Name and Subject authority records in our Voyager catalog so the processor searched our catalog to determine the appropriate form of the heading and updated the ASpace subject record accordingly.

<click> In this example the heading being fixed is: Slavery in the U.S. Antislavery movement. Thanks to cross-references, the search reveals that Slavery in the U.S. should be Slavery—United States.
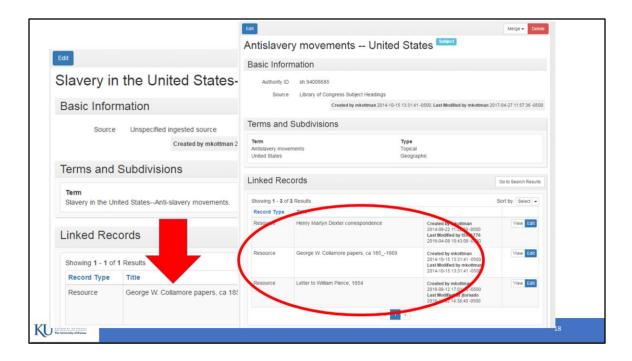
The Fix is In!!

Leader	00728cz a2200229n 4500
005:	20040318220713.0
008	831116 i l _ a n a n n b a b l _____ l a _ a n a _▉ _ _

Tag	I1	I2	Subfield Data
010			‡a sh 85123315
035			‡a (DLC)sh 85123315
040			‡a DLC ‡c DLC ‡d DLC
150			‡a Antislavery movements
450			‡a Abolitionism
450			‡a Anti-slavery movements
450			‡w nne ‡a Slavery ‡x Anti-slavery movements
550			‡w g ‡a Human rights movements
667			‡a Source for change: W.L. 11 (1994)
670			‡a Amer. Heritage dict. ‡b (antislavery; abolitionism)
670			‡a Random House ‡b (antislavery)
670			‡a Web. 3 ‡b (antislavery)
670			‡a Dict. Amer. hist. ‡b (antislavery movement)
670			‡a Hennepin ‡b (anti-slavery movements)

KU LIBRARIES
The University of Kansas

15

<click> Slavery—Antislavery movements is a cross-reference under Antislavery movements which can be subdivided geographically so the correct heading would be Antislavery movements—United States.

Merge duplicates

Since there were so many duplicate headings, we added a step to check for duplicates/fuzzy matches of each heading. This was done using the Merge function in ASpace. With a heading record open, clicking on the merge button opens a search box that provides a

<click>type-ahead result list so that you can quickly see if there are multiple duplicates for your search. In this example, searching for anti slavery finds two additional subject headings that need to be merged into the already open record.

<click>Merging records will automatically re-link the finding aid records using the merged subject headings to the remaining corrected subject heading record, so that you don't have to do it manually. So in this illustration the original heading Slavery in the U.S. Antislavery movement had one resource record attached. After the merge the corrected heading Antislavery movements—United States has three records attached.

Additional cleanup

- Change heading to authorized form per LCSH
  - Change Source field to Library of Congress Heading
- Add 010 field to Authority ID field if SAF record exactly matches
- Parse heading into appropriate term and subdivision fields

KU LIBRARIES
The University of Kansas

In addition to changing the heading to the authorized LCSH form, we changed the Source field to Library of Congress Subject Heading, added the Library of Congress Control Number to the Authority ID field when there was an LC subject heading record and put each element of the heading in separate fields with the appropriate term type indicated.

<click> Here's what the record looked like before… unspecified ingested source indicates it was created when the EAD record was imported and the heading is a single text string

<click> This is the record after it was updated with source filled out and the subdivision in its own field.

Challenges: Though you can set up an ODBC connection to ASpace, it's not how the developers envisioned it being used so there isn't any documentation or an ER Diagram for the table relationships which made creating a query challenging. Another challenge was searching. There isn't a left anchored search so searching single word terms like "Kansas" in our database resulted in over 14,000 hits to page through. To determine what was meant by some of the more creative headings, the resource record needed to be checked to see if there were any clues. This is when we ran into issues with the differences in the number and type of headings we'd add to a MARC record vs. the finding aids. It was challenging to explain why a processor in the past added 30 subject headings to a single letter while a 50-box collection just has a single subject heading.

As for benefits, I finally got our IT department to set up a read-only connection to the ASpace backend to run my queries. Our MARC cataloger is happy that she doesn't have to look for and fix fuzzy matches when adding new headings. Our ongoing project to review and update all the finding aids that were in ASpace prior to our upgrade to version 1.5 is no longer bogged down by having to update subject headings. And if needed, we could easily add the URL prefix to the Authority IDs for linked data compatibility.

Next steps

- Finish off the subject heading cleanup
  - Delete "unused" subject headings (978 subject headings no longer linked to a resource record)
  - Resolve outstanding conflicts/questions (57 records)
    - Great Britain (chiefly faculty-members at various universities) also US, Canada, Ireland, Italy; Lawrence—Architecture (1017 [1015] Tennessee)
  - Remove improperly coded personal names/corporate bodies from subject file – add to agent file (670 records)
- Create a report of agent records to review/update/merge
- Complete version 1.5 review project and re-publish EADs and/or start using Aspace public interface

And finally, what are our next steps? To finish off the subject heading cleanup we need to delete the 978 headings that are no longer linked to a resource record (3,847 remaining), have the manuscript processors figure out what is going on in the records linked to the 57 headings we couldn't figure out, and the most time consuming project will be deleting the 670 subject headings for agents and adding them to the agent file instead. We'll likely wait until we complete a similar project on the agent records. Once that is done and we've completed our project to review the pre-version 1.5 records (began Sept. 2016), we'll export new EADs and publish to our website or start using ArchivesSpace's public interface (or maybe both)

Contact information

Miloche Kottman: mkottman@ku.edu
Head of Cataloging & Archival Processing
University of Kansas Libraries

ArchivesSpace
Website: http://archivesspace.org/
Github: https://github.com/archivesspace

KU LIBRARIES
The University of Kansas