



Converting our digital collections' legacy non-faceted subjects;

Or, how we learned to stop worrying and love the facets

Rebecca Saunders (she/her)
Cataloging and Metadata Librarian
Western Carolina University
rlsaunders@wcu.edu

Faceted Subject Access Interest Group (FSAIG)
ALA Core Interest Group Week 2024



Introducing SADC

- Southern Appalachian Digital Collections (SADC)
- Regional partnership begun by Western Carolina University (WCU) and University of North Carolina Asheville (UNCA)
- Mission to expand partnership to regional hub
- Launched April 2022

Southern Appalachian Digital Collections

Highlights

Collections

Exhibitions

▼ All

Search the collections...



Advanced search



The Southern Appalachian Digital Collections contain 27694 records.

Learn more

Explore highlights

4 / 8 Chimney Tops, 1928

Southern Appalachian Digital Collections

- Both institutions migrated well-established legacy collections from CONTENTdm to an independent CMS, Qi by Keepthinking
- New global metadata application profile (MAP) combined 25 collection-level metadata profiles and controlled vocabularies
- Combined list of 12,876 topical subjects post-migration



SADC Digital Production Team

- Digital Scholarship Librarian
- Special and Digital Collections
 - One librarian/archivist and four specialists
- Cataloging and Metadata
 - One librarian and three specialists
- Web Services and User Experience
- Systems and Discovery
- Administration
 - WCU COMET Department Head
 - UNCA Head of Special Collections and University Archives

Subject matter experts who provide feedback on usability of subjects and controlled vocabulary.

Subject experts who manage subject cataloging, vocabulary control, and vocabulary remediation.



Controlled vocabulary enhancement



Two institutions, two different levels of vocabulary control



Expanded to new vocabularies
beyond LCSH and local

Art and Architecture Thesaurus (AAT)
Thesaurus for Graphic Materials (TGM)
Homosaurus
GeoNames



Cleanup projects planned to convert natural language tags, data entry errors, and other idiosyncrasies to authorized controlled vocabularies or patterns

**SADC
vocabulary
remediation
project series**
2022-2023

Phase 1

Topical subjects



Phase 2

Local controlled vocabularies



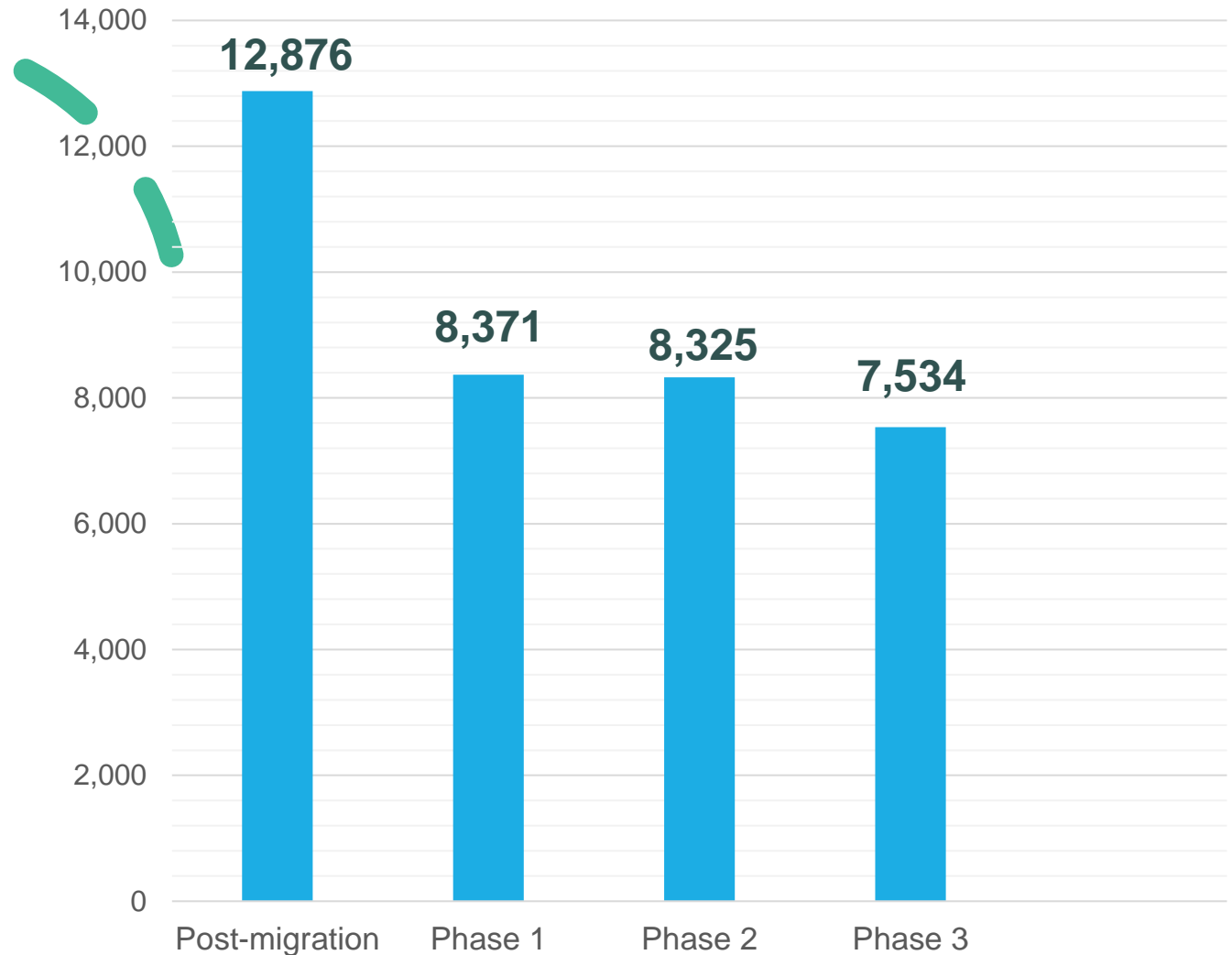
Phase 3

Personal and corporate names

Vocabulary remediation projects

- Decreasing size of legacy subject list while introducing new terms.
- Reestablishing vocabulary control.
- Developing a more usable vocabulary.

Number of legacy topical subjects



Heavily subdivided, non-faceted subjects

- Long-standing policy was to pre-coordinate subjects using subdivisions with as much specificity as possible.
- Subject metadata specialists originally held commitment to subdivided, non-faceted vocabularies.
- Value in rich information invested in pre-coordinated headings and in disambiguation provided by subdivisions.

Examples SADC non-faceted subjects with subdivisions

African American churches – North Carolina – Cullowhee

Camp sites, facilities, etc. – Great Smoky Mountains (N.C. and Tenn.) – Maps

Farm life – North Carolina – Haywood County

Microbreweries – North Carolina – Jackson County

Paper industry – Environmental aspects – North Carolina – Canton

Steam locomotives – North Carolina – Swain County

Waterfalls – North Carolina – Swannanoa River, North Fork

Non-faceted subjects create silos

Having several terms represent the same topic obstructs collocation and inhibits discoverability through pre-selected subject filters and subject index searching.

| Base term | Number of subdivided terms | Total object relationships | Filter term relationships % of total |
|-----------------|----------------------------|----------------------------|--------------------------------------|
| Logging | 17 | 184 | 56% |
| Paper industry | 20 | 183 | 21% |
| Railroad trains | 7 | 109 | 63% |
| Waterfalls | 11 | 254 | 26% |
| Wood-carving | 5 | 777 | 42% |

Pre-selected subject filters lose functionality

With a non-faceted vocabulary, subject filters link to a relatively low percentage of the total number of digital objects associated with a topical subject.

Percentage of total number of objects associated with topic across several subdivided subjects

| | Format | Source Collection | <u>Subject - Topic</u> | Type |
|-----|--------|--|--|------|
| | | | Great Smoky Mountains National Park (N.C. and Tenn.) (152) | |
| 44% | → | Hunting (38) | | |
| 38% | → | Landscape photography (10) | | |
| 56% | → | Logging (103) | | |
| | | Maps (84) | | |
| 29% | → | Mines and mineral resources (8) | | |
| | | North Carolina -- Maps (18) | | |
| 21% | → | Paper industry (38) | | |
| | | Postcards (255) | | |
| 28% | → | Pottery (135) | | |
| 63% | → | Railroad trains (69) | | |
| | | Rural electrification -- North Carolina, Western (3) | | |
| 4% | → | School integration -- Southern States (2) | | |
| 9% | → | Segregation -- North Carolina, Western (5) | | |
| 31% | → | Slavery (5) | | |

Subject index searching is limited

Similarly, subdivided subjects limit users' serendipitous discovery through subject searching starting from an object of interest.

WCU - Picturing Appalachia

Carolina Wood Turning Company in snow



Percentage of total number of objects associated with topic across several subdivided subjects

| | Subject - Topic |
|-----|--|
| 71% | → <u>Industrial buildings</u> |
| 78% | → <u>Lumber -- North Carolina, Western</u> |
| 34% | → <u>Lumber trade -- North Carolina, Western</u> |
| 46% | → <u>Mills and mill-work -- North Carolina, Western</u> |
| 39% | → <u>Mountains -- North Carolina, Western</u> |
| 8% | → <u>Railroad tracks -- North Carolina -- Swain County</u> |
| 12% | → <u>Roads -- North Carolina, Western</u> |
| 73% | → <u>Snow</u> |

Transitioning to faceted subjects



Began considering transition to faceted subjects when exploring discovery through search filters on website.



Pre-coordinated subjects can inhibit discovery in a system that relies heavily on search filters and limits subject browsability.



Inciting case: “African Americans” subject filter returning too few results led to Phase 4 trial.

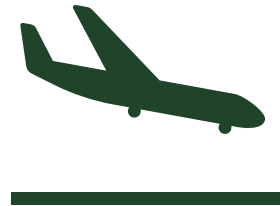
Initiate Phase 4 of vocabulary remediation

- Phase 4 began in late summer 2023 and is currently in progress.
- Concerns about potential loss of rich, nuanced information invested in subject subdivisions.
- Goal is to retain information in existing pre-coordinated subjects while enabling faceted searching.



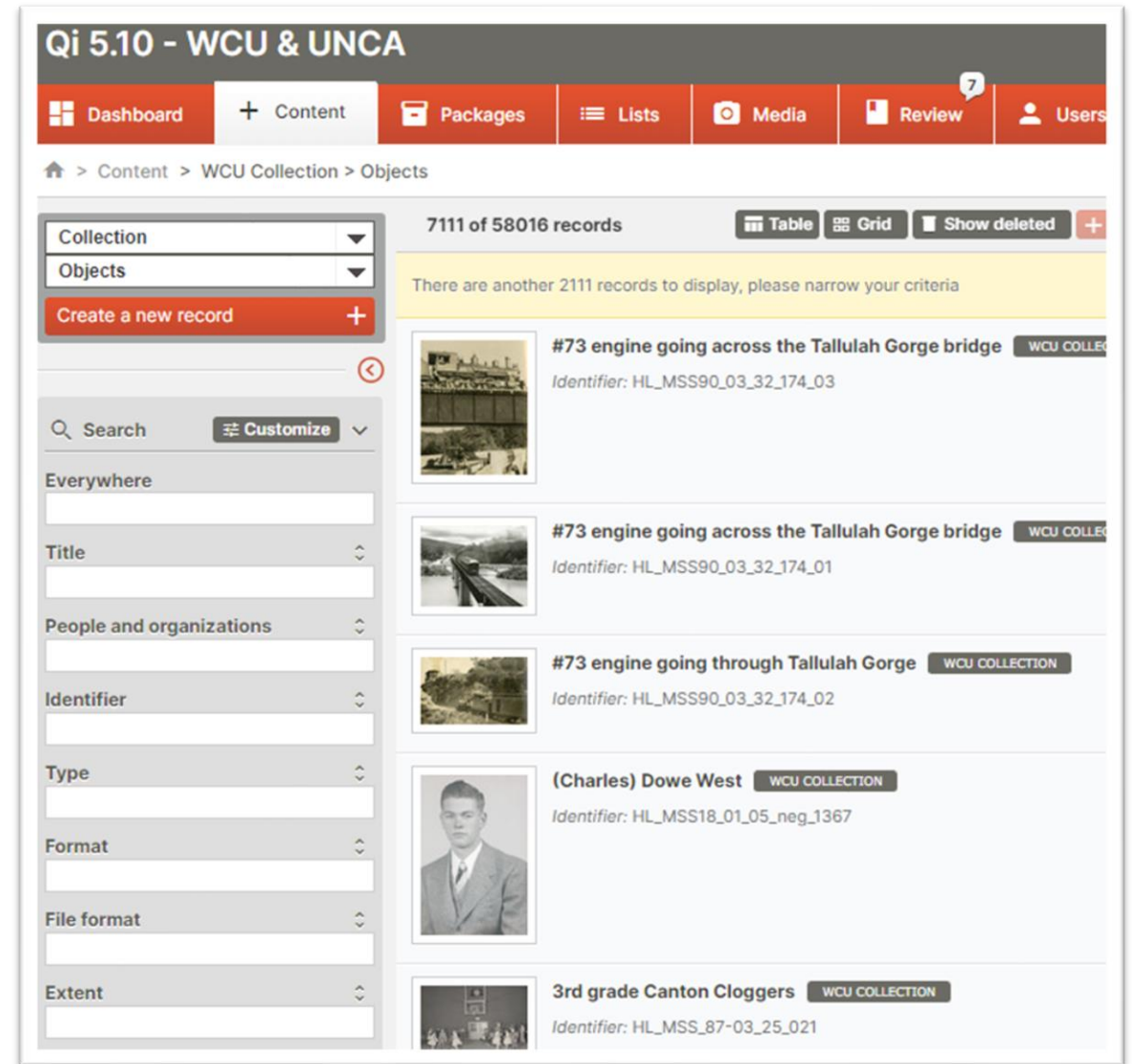
Phase 4: Conversion to faceted subjects

- Solution: Add information in subject subdivisions as values in other metadata elements in records linked to subdivided subjects before converting individual terms to a faceted vocabulary.
- “Final destinations”:
 - Geographic subdivisions are added to Location element
 - Form subdivisions are added to Format element
 - Requires translation to Art and Architecture Thesaurus (AAT)
 - Dropping chronological subdivisions
- Current focus is on converting geographic, form, and chronological subdivisions to faceted vocabularies. More nuanced approach for assessment and conversion of topical subjects will be another phase.



System tools in Qi

- SADC's content management system (CMS) is Qi, developed by Keepthinking.
- Designed for facile record editing within system.
- Tools used for subject conversion:
 - Boolean searching in CMS
 - Batch-add relationships between objects and vocabulary terms
 - Reduction tool: merge like terms and relationships to objects
 - "List" records: locally-developed and maintained records for each controlled vocabularies



Phase 4 Conversion Procedure in Qi

1. Identify and record subdivided subjects for conversion to faceted vocabulary
2. Search subject in Qi to add subdivision information to appropriate element
3. Rename or merge subject as appropriate

1. Identify and record subdivided subjects for conversion to faceted vocabulary

- Select pre-designated alphabetical range in project worksheet.
- Comb through the A-Z topical subject list to add terms with geographic, form, and chronological subdivisions to worksheet.
- Record actions for each term in project worksheet moving forward.

The image displays a project worksheet with a list of subdivided subjects on the left and a corresponding alphabetical range in a project worksheet on the right. The left list includes:

- ☐ Interracial marriage -- North Carolina, Western
- ☐ Intramural sports
- ☐ Intrusions (Geology)
- ☐ Invasive plants -- North Carolina
- ☐ Investments
- ☐ Invitation cards
- ☐ Invitations
- ☐ Iobst, Richard W. (Richard William), 1934-2010 -- Correspondence
- ☐ Iraq War, 2003-2011 -- Personal narratives
- ☐ Iraq War, 2003-2011 -- Personal narratives, American
- ☐ Irises (Plants)
- ☐ Irish Americans -- Ethnic identity
- ☐ Iron -- Metabolism
- ☐ Iron and steel bridges -- Georgia
- ☐ Iron and steel bridges -- North Carolina
- ☐ Iron and steel bridges -- South Carolina
- ☐ Iron and steel bridges -- Tennessee -- Sevier County
- ☐ Iron implements -- Appalachian Region, Southern
- ☐ Ironworks -- North Carolina -- Jackson County

The right side shows a project worksheet with a blue header "Inf-lvy" and a list of topics. The topics are:

- 1 Inf-lvy
- 2 Qi topical subject
- 3 Infants -- Development -- Sex differences
- 4 Infants -- Portraits
- 5 Insects -- Great Smoky Mountains (N.C. and Tenn.)
- 6 Intergenerational relations -- North Carolina -- Jackson County
- 7 Interior decoration -- Georgia -- Atlanta
- 8 Internal migrants -- Washington
- 9 Interracial marriage -- North Carolina, Western
- 10 Invasive plants -- North Carolina
- 11 Iobst, Richard W. (Richard William), 1934-2010 -- Correspondence
- 12 Iraq War, 2003-2011 -- Personal narratives
- 13 Iraq War, 2003-2011 -- Personal narratives, American
- 14 Iron and steel bridges -- Georgia
- 15 Iron and steel bridges -- North Carolina
- 16 Iron and steel bridges -- South Carolina
- 17 Iron and steel bridges -- Tennessee -- Sevier County
- 18 Iron implements -- Appalachian Region, Southern
- 19 Ironworks -- North Carolina -- Jackson County
- 20 Ivy (N.C. : Township) -- Maps

Blue arrows indicate the mapping from the left list to the right worksheet, showing how the subdivided subjects are recorded in the project worksheet.

2. Search subject in Qi to add subdivision information to appropriate field

- Subject index search for term.
- Boolean search in Qi excludes objects that already have subdivision information in relevant element.
- Once all the records that don't have subdivision values in relevant fields are retrieved, batch add the value to Format or Location field.

The screenshot displays a digital collection management interface. On the left, a sidebar contains search filters: Extent, Date from, Date to, Date text, Description, Subject - Topic, Time period, Location, Repository, Digital collection, and Source collection. The 'Subject - Topic' filter is highlighted with a blue box and contains the text '"Hotels -- North Carolina -- Asheville"'. The main area shows '29 of 65934 records' with a list of objects. A modal window is open for batch updating, showing a table with columns for Type, Location, Record, and Notes. The 'Location' column is highlighted with a blue box and contains the value 'Asheville (N.C.)'. A blue arrow points from the '29 of 65934 records' text to the 'Add' button in the modal. Below the modal, a list of objects is visible, including 'Asheville, N.C., The Kenilworth Inn', 'Berkeley Hotel, Asheville, N.C.', and 'Biltmore Avenue One of two Views (other C278-4) of Biltmore Avenue'.

Content > WCU Collection > Objects

29 of 65934 records

Identifier: ball1530

29 of 65934 records (29 records selected)

Table Grid Show Relationship Package Reports Import

Type

Location

Record

→ Asheville (N.C.)

Notes

0 / 65535

Cancel Add +

Asheville, N.C., The Kenilworth Inn WCU COLLECTION

Identifier: HL_RP_1767_A

Berkeley Hotel, Asheville, N.C. WCU COLLECTION

Identifier: HL_RP_1494_A

Biltmore Avenue One of two Views (other C278-4) of Biltmore Avenue UNICA COLLECTION

Identifier: B367-4

3. Rename or merge subject as appropriate

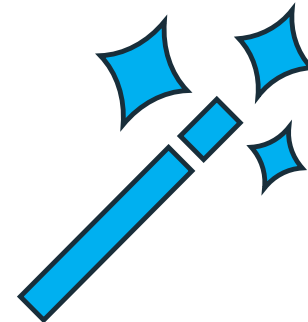
Before:

- ☐ Bankers
- ☐ Banks and banking
- ☐ Banners
- ☐ Baptism
- ☒ Baptist church buildings
- ☒ Baptist church buildings -- Georgia -- Alto
- ☒ Baptist church buildings -- Georgia -- Clarkesville
- ☒ Baptist church buildings -- Georgia -- Cornelia
- ☒ Baptist church buildings -- Georgia -- White County
- ☒ Baptist church buildings -- North Carolina -- Asheville
- ☒ Baptist church buildings -- North Carolina -- Canton
- ☒ Baptist church buildings -- North Carolina -- Haywood County
- ☒ Baptist church buildings -- North Carolina -- Jackson County
- ☒ Baptist church buildings -- North Carolina -- Swain County
- ☐ Baptisteries
- ☐ Baptists -- Appalachian Region
- ☐ Baptists -- Missions
- ☐ Baptists -- North Carolina -- Cullowhee
- ☐ Baptists -- North Carolina, Western
- ☐ Barbecues (Fireplaces)
- ☐ Barbecuing

After:

- ☐ Bankers
- ☐ Banks and banking
- ☐ Banners
- ☐ Baptism
- ☒ Baptist church buildings
- ☐ Baptist church buildings -- Georgia -- Alto
- ☐ Baptist church buildings -- Georgia -- Clarkesville
- ☐ Baptist church buildings -- Georgia -- Cornelia
- ☐ Baptist church buildings -- Georgia -- White County
- ☐ Baptist church buildings -- North Carolina -- Asheville
- ☐ Baptist church buildings -- North Carolina -- Canton
- ☐ Baptist church buildings -- North Carolina -- Haywood County
- ☐ Baptist church buildings -- North Carolina -- Jackson County
- ☐ Baptist church buildings -- North Carolina -- Swain County
- ☐ Baptisteries
- ☐ Baptists -- Appalachian Region
- ☐ Baptists -- Missions
- ☐ Baptists -- North Carolina -- Cullowhee
- ☐ Baptists -- North Carolina, Western
- ☐ Barbecues (Fireplaces)
- ☐ Barbecuing

- If no correlating term exists, then change the label of subject list record to a faceted term.
- If a correlating faceted term does exist, then merge terms using the using Qi term reduction tool.



Before and after merging terms.

Conversion to faceted subjects: General principles

1. Set the scope of the project
2. Migrate pre-coordinated subject values to other elements in metadata records
 - Investigate your system's capabilities for exporting record data by subject and batch importing changes to records.
3. Adjust your non-faceted vocabulary term to a faceted vocabulary term by removing subdivisions
 - Ensure that you add non-faceted subjects to records that had subdivided terms when combining subjects with multiple subdivisions.



Improvements to discovery

- Effective faceted searching with multiple filters
- Positive feedback from Special and Digital Collections
- Plans for faceted search guide to provide users direction on post-coordinating subjects

Search interface showing faceted search results for "Dwellings".

Search bar: Search the collections... (All) (Advanced search)

Facets:


- Digital Collection
- Creator
- Time Period
- Location
- Format
- Source Collection
- Subject - Topic
- Type

Location results:

- Appalachian Region, Southern (9)
- Asheville (N.C.) (114)
- Blount County (Tenn.) (1)
- Buncombe County (N.C.) (55)
- Cherokee County (N.C.) (1)
- Clay County (N.C.) (53)
- Graham County (N.C.) (1)
- Great Smoky Mountains National Park (N.C. and Tenn.) (4)
- Haywood County (N.C.) (144)
- Henderson County (N.C.) (11)
- Macon County (N.C.) (19)
- Madison County (N.C.) (1)
- McDowell County (N.C.) (1)
- Mitchell County (N.C.) (1)
- Polk County (N.C.) (2)
- Qualla Boundary (2)
- Rutherford County (N.C.) (1)
- Swain County (N.C.) (141)
- Transylvania County (N.C.) (24)
- Waynesville (N.C.) (11)
- Yancey County (N.C.) (1)

1316 results for: Dwellings X

Sort:



“Dwellings” formerly spanned 22 subdivided subjects.

Thank you!

Contact:

Rebecca Saunders (she/her)

rlsaunders@wcu.edu

