Converting our digital collections' legacy non-faceted subjects;

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

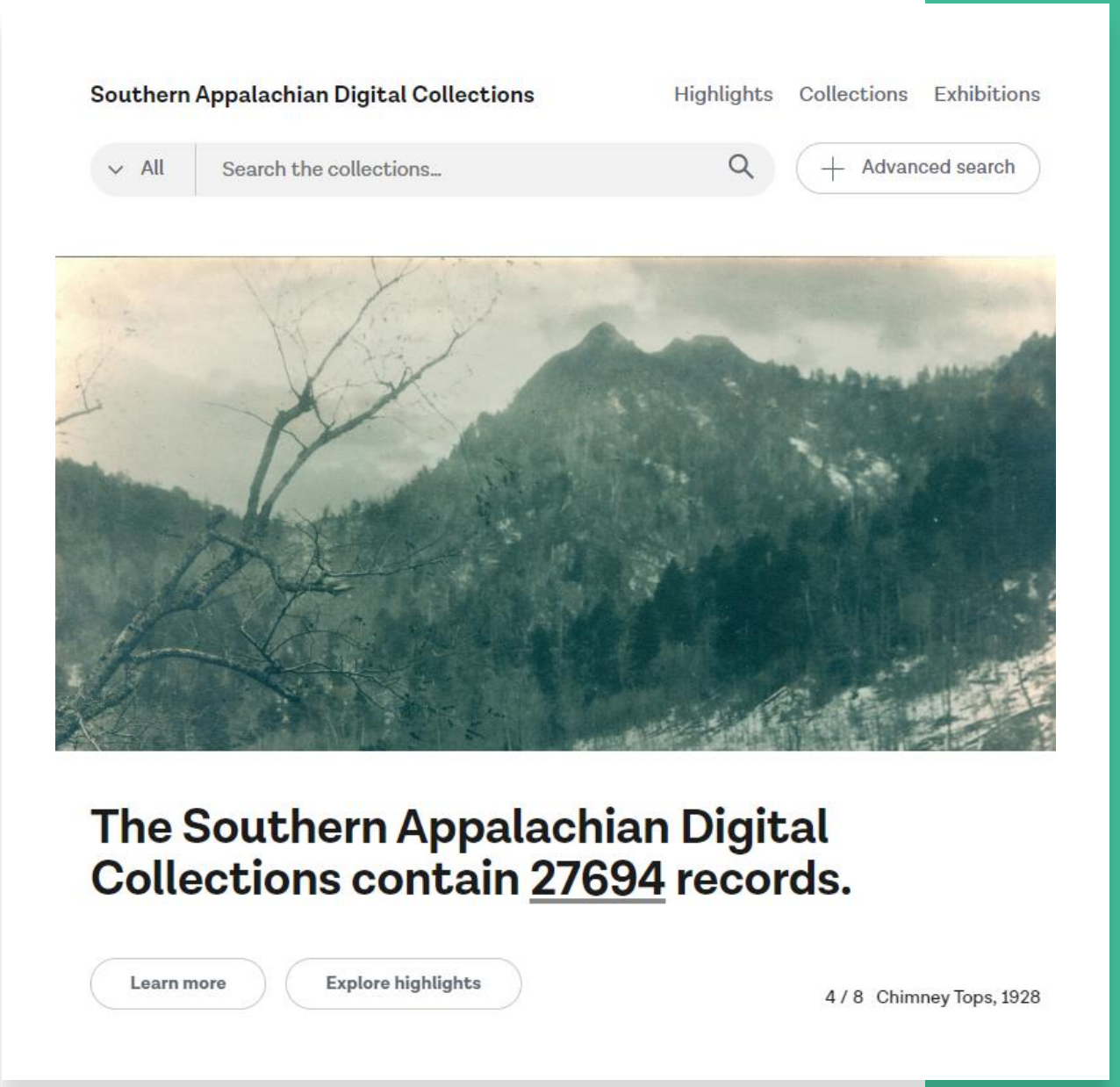
Rebecca Saunders (she/her)
Cataloging and Metadata Librarian
Western Carolina University
rlsaunders@wcu.edu

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

Core
LEADERSHIP
INFRASTRUCTURE
FUTURES

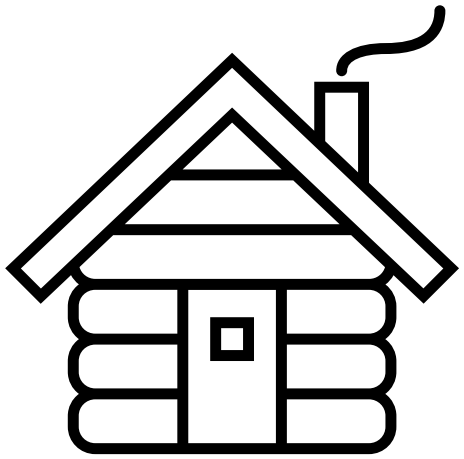
Introducing SADC

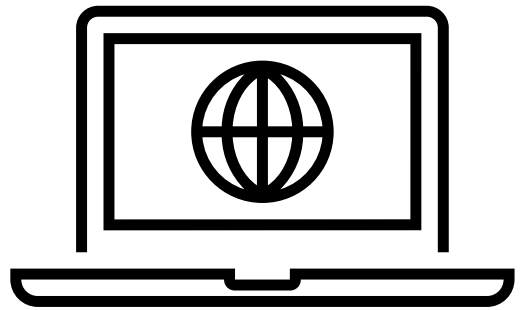
- Southern Appalachian Digital Collections (SADC)
- Regional partnership begun by Western Carolina University (WCU) and University of North Carolina Asheville (UNCA)
- Mission to expand partnership to regional hub
- Launched April 2022



Southern Appalachian Digital Collections

- 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

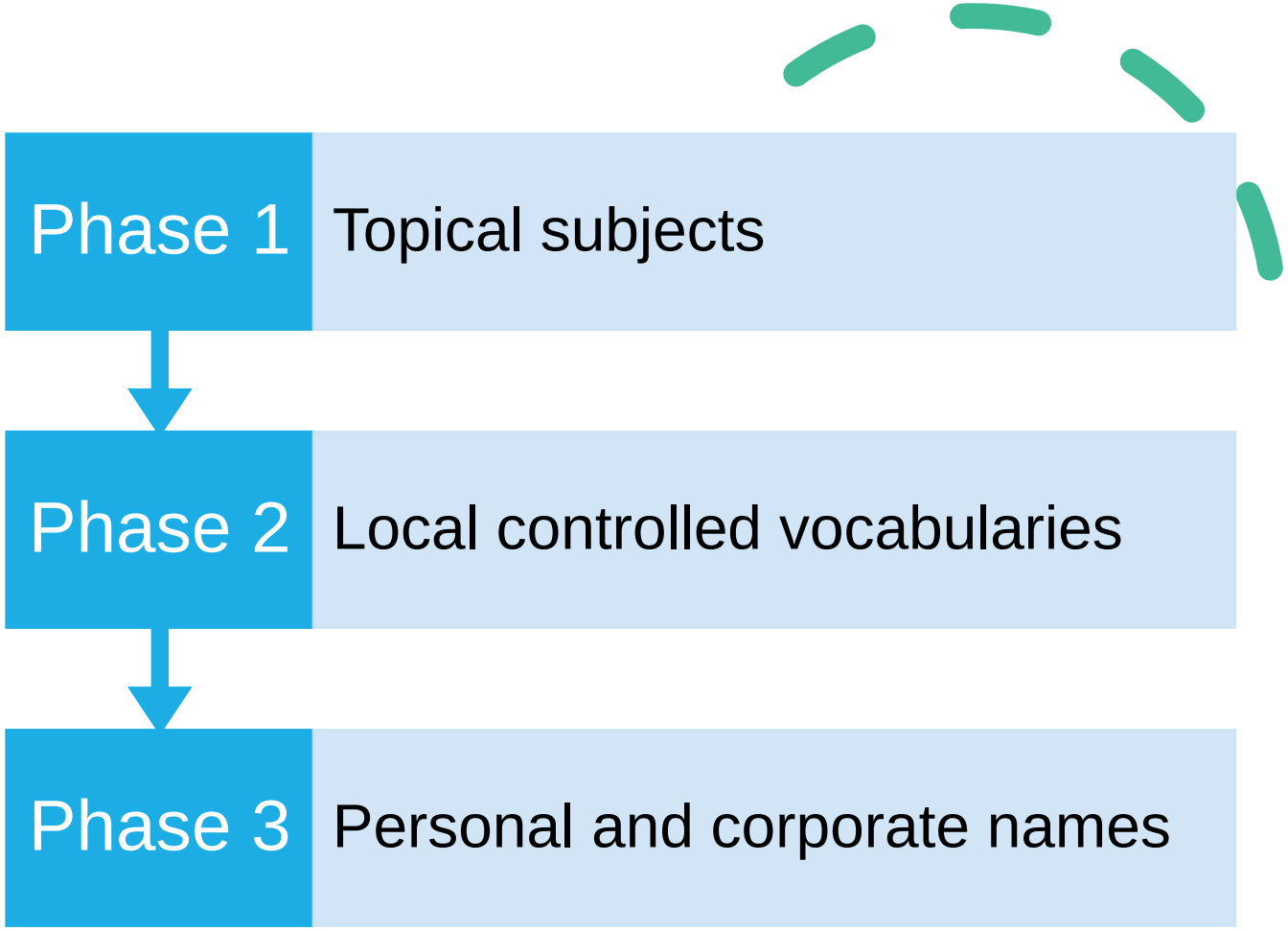
Expanded to new vocabularies beyond LCSH and local

- 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

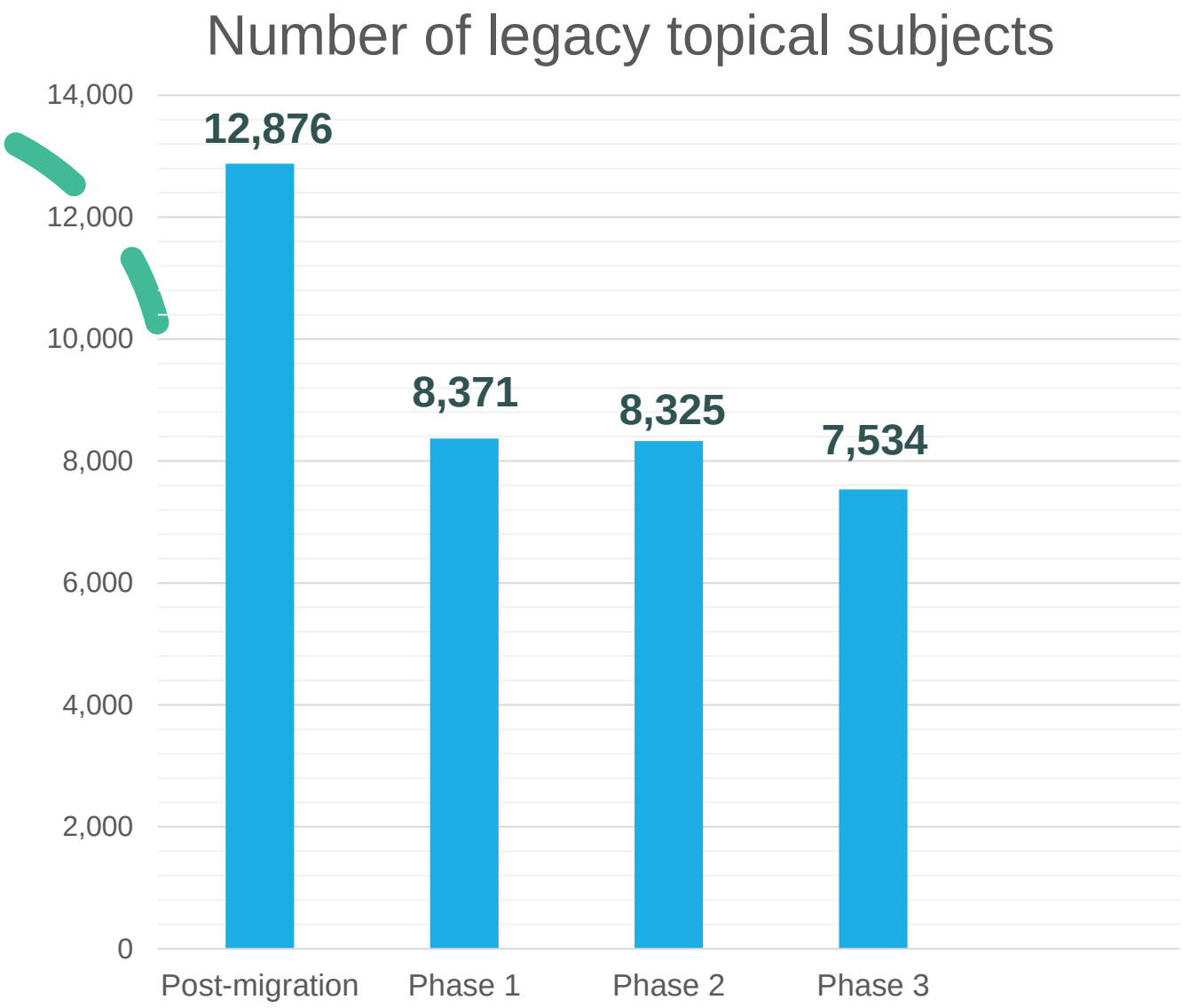
SADC vocabulary remediation project series

2022-2023



Vocabulary remediation projects

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



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

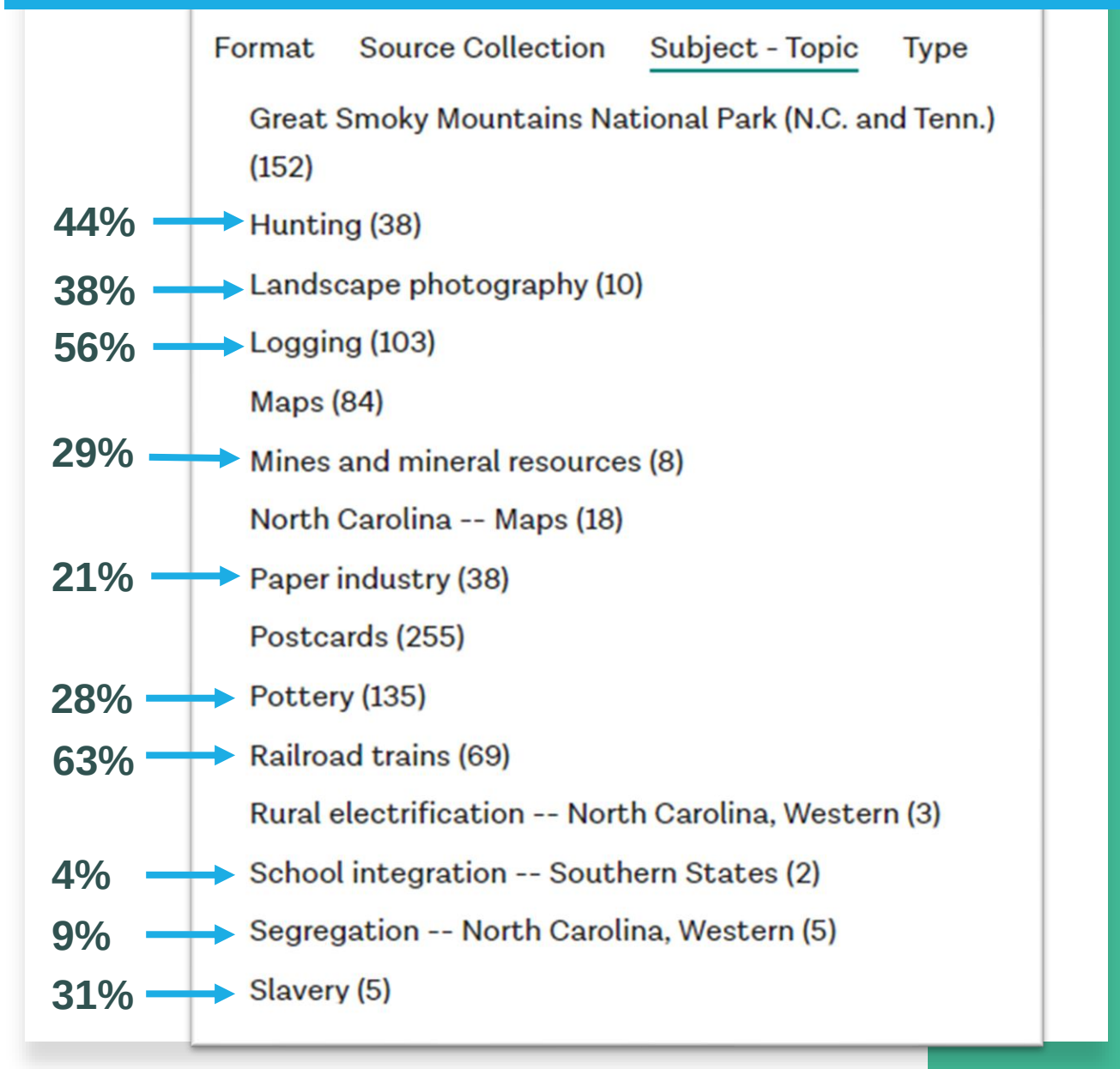
Having several terms represent the same topic obstructs collocation and inhibits discoverability through pre-selected subject filters and subject index searching.

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

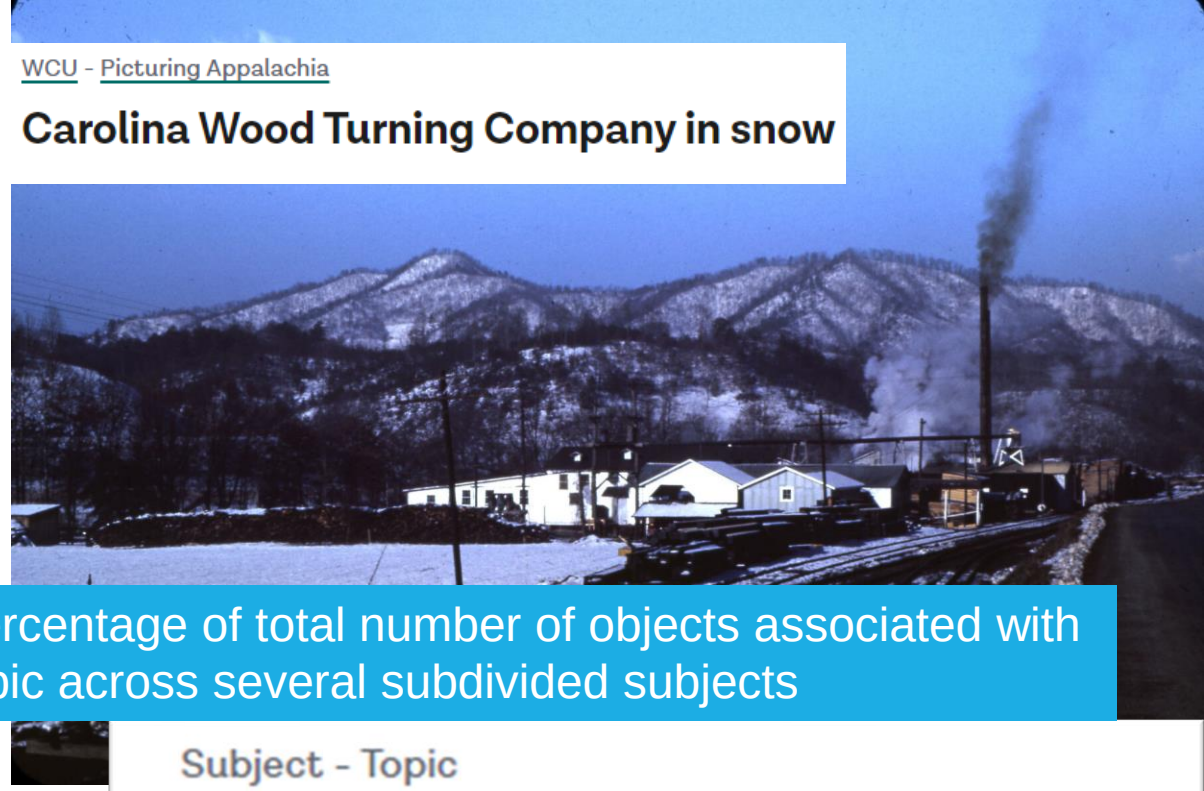
With a non-faceted vocabulary, subject filters link to a relatively low percentage of the total number of digital objects associated with a topical subject.

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



Subject index searching is limited

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



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

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

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.

This Photo by Unknown Author is licensed under CC BY-SA

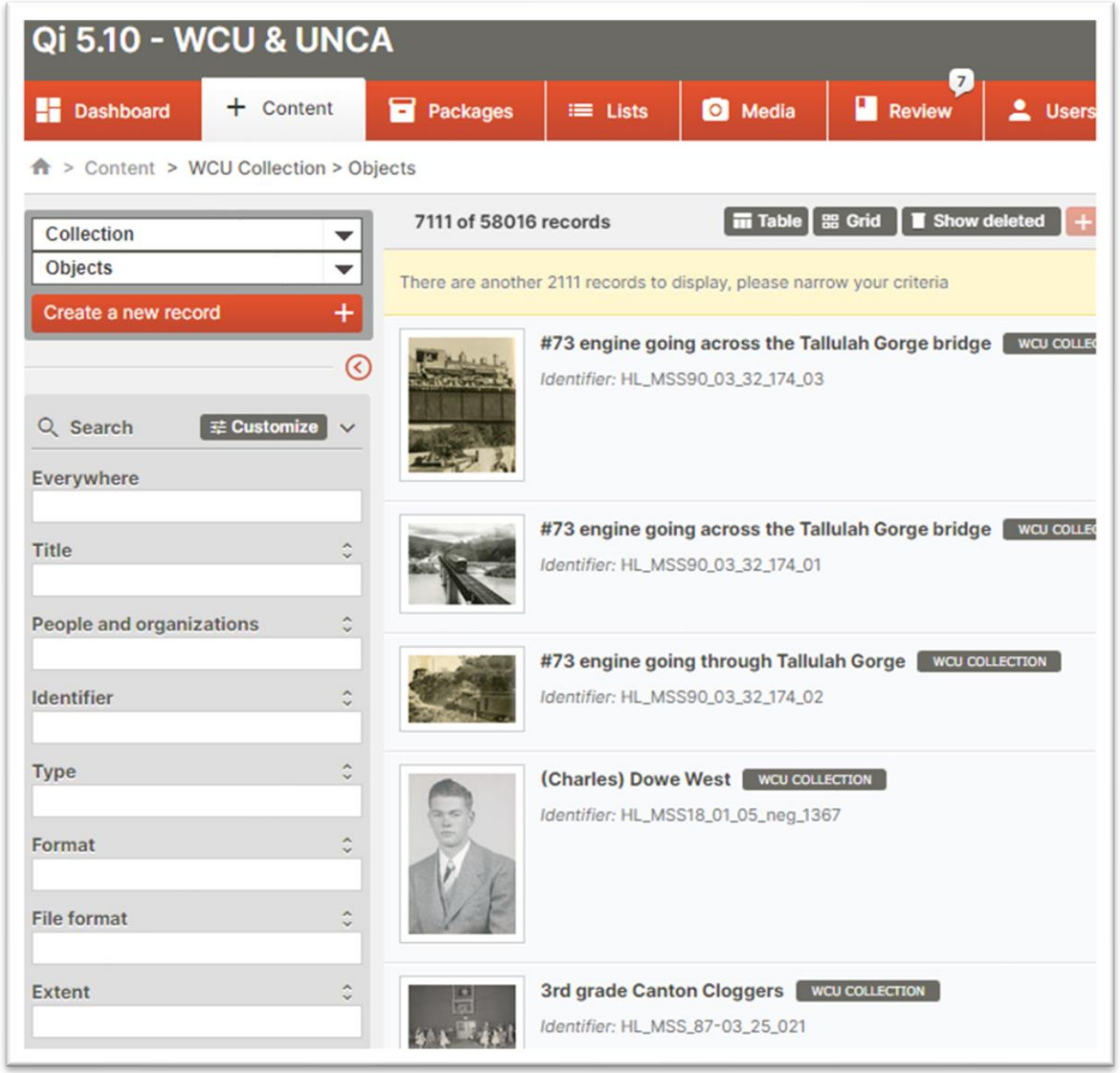
Phase 4: Conversion to faceted subjects

- 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

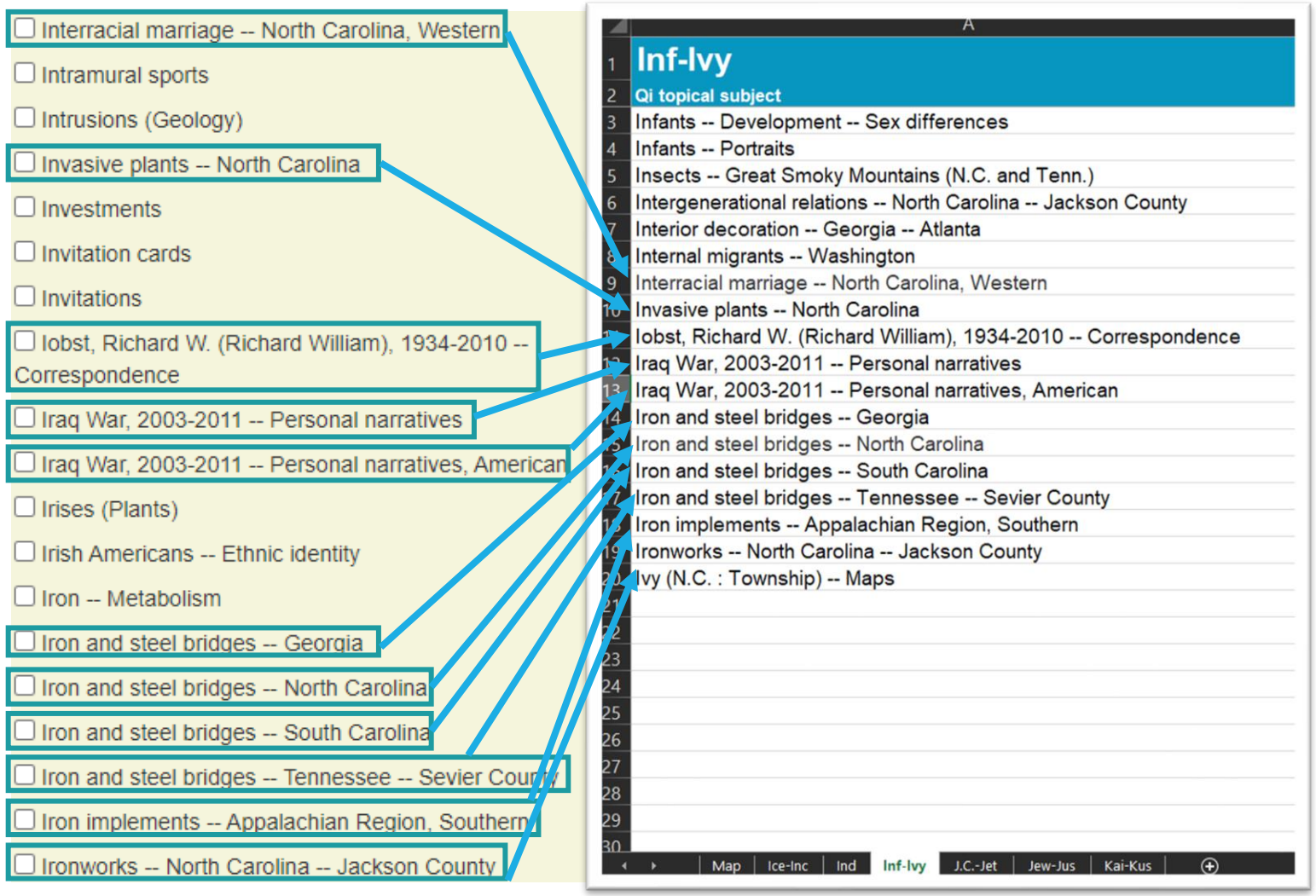


Phase 4 Conversion Procedure in Qi

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

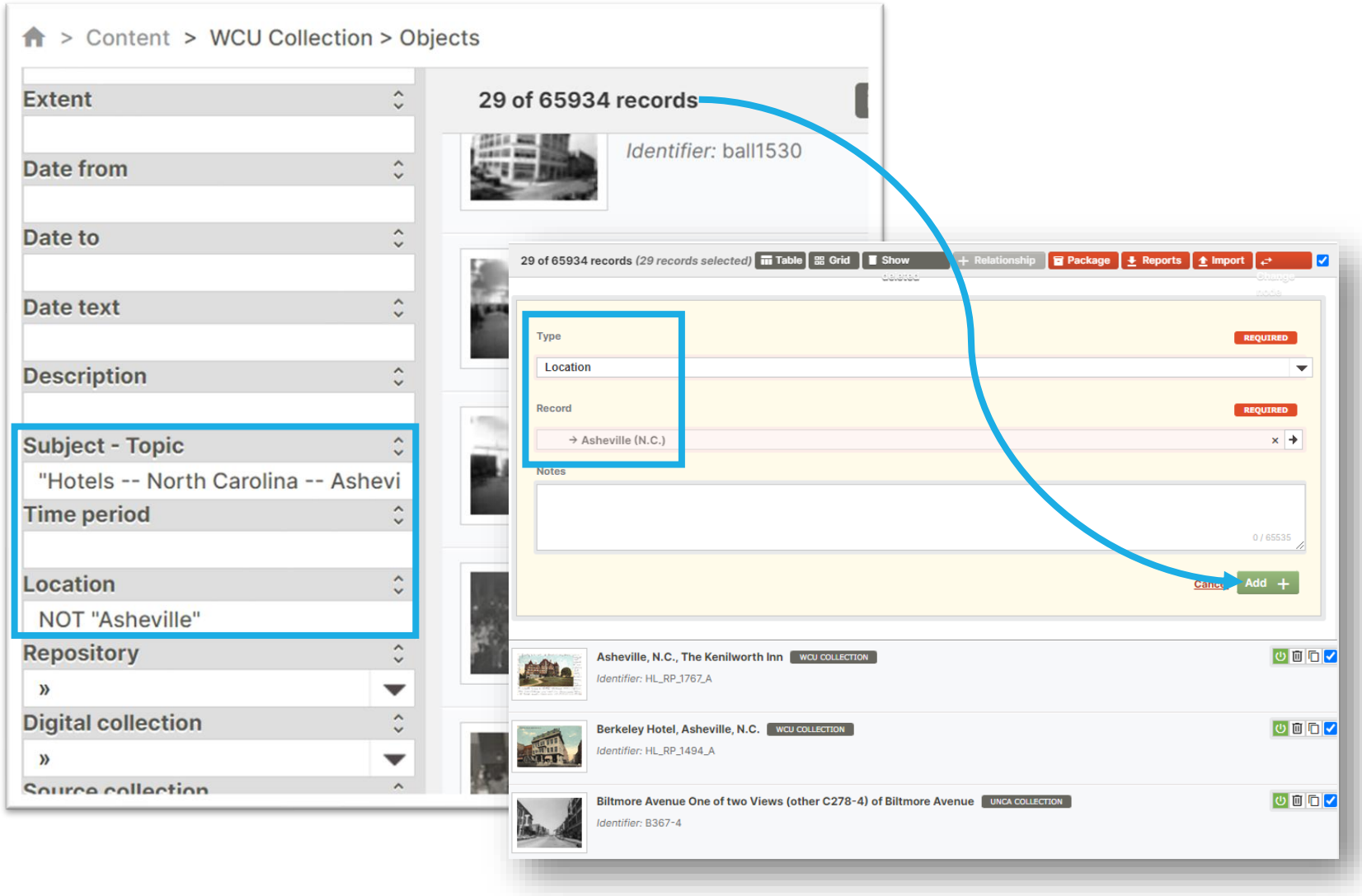
1. Identify and record subdivided subjects for conversion to faceted vocabulary

- 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.

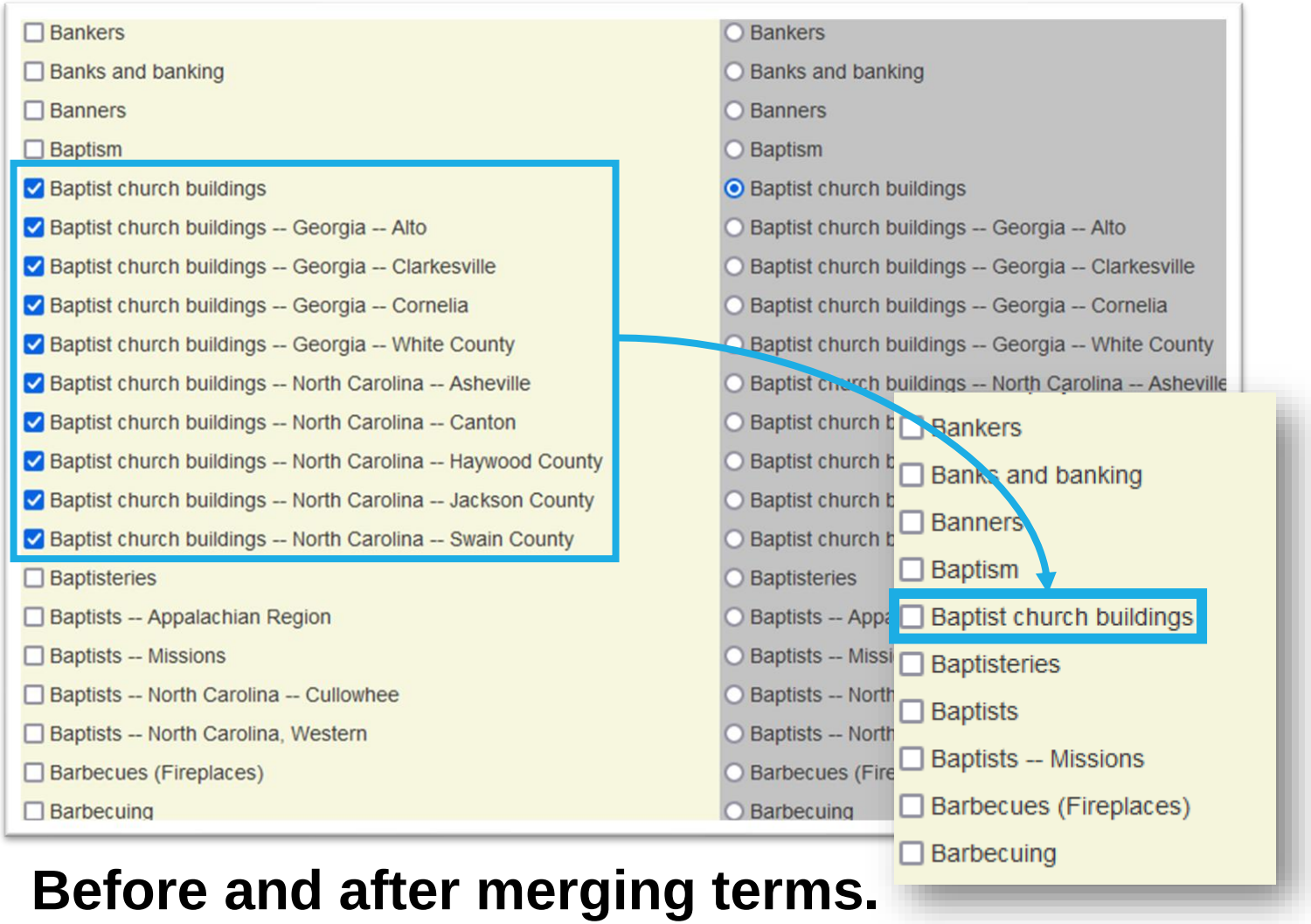


2. Search subject in Qi to add subdivision information to appropriate field

- 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.

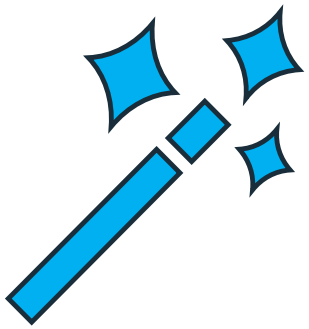


3. Rename or merge subject as appropriate



Before and after merging terms.

- 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.

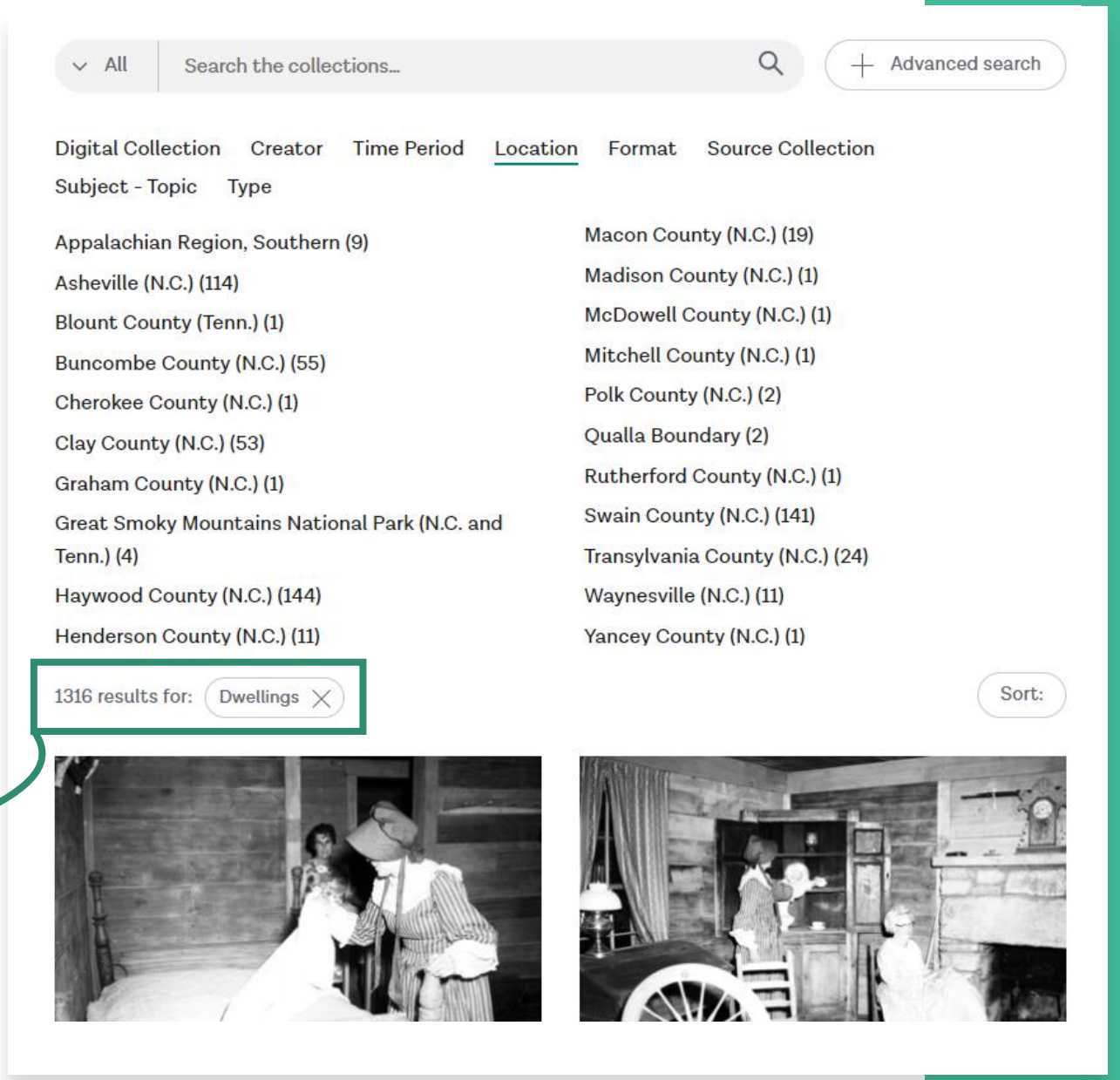


Conversion to faceted subjects: General principles

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



“Dwellings” formerly spanned 22 subdivided subjects.

Thank you!

Contact:

Rebecca Saunders (she/her)

rlsaunders@wcu.edu