Broadcasting A/V Data to the SNAC universe (and beyond!)

September 19, 2023

Emily Frazier (she/her)
formerly: Maryland Institute for Technology in the Humanities
currently: University of Maryland Libraries
project background

❖ follow-up project to Unlocking the Airwaves (virtual reunification of NAEB collection)
❖ centered on archival authorities--aka CPF records, people and organizations, or entities
❖ goal was to enhance discoverability of authorities and the fully-digitized radio programs in which they appear or helped create
❖ 2 sides:
  ➢ project website, with links to other authority records and streaming audio as well as data visualization
  ➢ linked data on the web--SNAC, Wikidata, and Wikipedia
two SNAC processes

1. getting data OUT of SNAC for use on BAVD project website and to know what data to submit back to SNAC

2. uploading data TO SNAC using the OpenRefine + SNAC extension to surface authority records of people and organizations in early educational radio to the web
batch upload workflow: overview

1. reconcile authority records
2. research + write SNAC records
3. configure OpenRefine
4. export data + import to OR
5. verify edits + export
6. upload to SNAC
7. map to SNAC schema
8. format data for SNAC upload
batch upload workflow: overview

reconcile authority records
research + write SNAC records
configure OpenRefine
export data + import to OR

verify edits + export
upload to SNAC
map to SNAC schema
format data for SNAC upload

repeats 3x: CPFs, resources, relationships
batch upload workflow: reconciling

1. **reconcile authority records**
2. research + write SNAC records
3. configure OpenRefine
4. export data + import to OR
5. format data for SNAC upload
6. map to SNAC schema
7. upload to SNAC
8. verify edits + export
batch upload workflow: reconciling

Frances Eleanor Page
composer for radio station KUT

method 1: pull SNAC ID from Wikidata

method 2: SNAC reconciliation in OpenRefine

method 3: searching in SNAC
batch upload workflow: OR data formatting

1. reconcile authority records
2. research + write SNAC records
3. configure OpenRefine
4. export data + import to OR

5. verify edits + export
6. upload to SNAC
7. map to SNAC schema

format data for SNAC upload
batch upload workflow: OR data formatting

<table>
<thead>
<tr>
<th>OpenRefine</th>
<th>SNAC Records SNAC prod CPFs March2023 csv Permalink</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Image of OpenRefine interface" /></td>
<td></td>
</tr>
</tbody>
</table>

**30 rows**

<table>
<thead>
<tr>
<th>Show as: rows records Show: 5 10 25 50 100 500 1000 rows</th>
</tr>
</thead>
<tbody>
<tr>
<td>![OpenRefine interface showing data] (image)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNAC ID</th>
<th>Entity Type</th>
<th>Linked CPF Authority</th>
<th>Date of Birth</th>
<th>Date of Death</th>
<th>Active Date 1</th>
<th>Active Date 2</th>
<th>Subject(s)</th>
<th>Places</th>
<th>Occupation(s)</th>
<th>Activities</th>
<th>Biographical or Historical</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>person</td>
<td>&quot;Bidack, Cecil S.&quot;</td>
<td>July 2, 1903</td>
<td>October 16, 1995</td>
<td></td>
<td></td>
<td>Public broadcasting, Radio Educational broadcasting, Public radio/Radio broadcasting</td>
<td>Ohio, Columbus (Ohio)</td>
<td>Engineer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>person</td>
<td>&quot;Clayton, John Strother&quot;</td>
<td>April 17, 1925</td>
<td>April 25, 1929</td>
<td></td>
<td></td>
<td>Public broadcasting, Radio Educational broadcasting, Public radio/Radio broadcasting</td>
<td>Chapel Hill (N.C.), North Carolina</td>
<td>Radio Executive</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

an example starting SNAC CPF dataset in OpenRefine, before data formatting
### Batch Upload Workflow: OR Data Formatting

<table>
<thead>
<tr>
<th>SNAC OR Schema Field Name</th>
<th>Name Entry</th>
<th>D</th>
<th>Existing Date</th>
<th>Existing Date Type</th>
<th>Subject</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airwaves Field Name(s)</td>
<td>Linked CPF Authority</td>
<td>Date of Birth; Date of Death; Active Date 1; Active Date 2</td>
<td>Date of Birth; Date of Death; Active Date 1; Active Date 2</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
</tr>
<tr>
<td>Airwaves Table(s)</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
</tr>
<tr>
<td>BAVD Field Name(s)</td>
<td>SAMPLE</td>
<td>birthDate; deathDate</td>
<td>birthDate; deathDate</td>
<td>subjects</td>
<td>subjects</td>
<td>subjects</td>
<td>subjects</td>
</tr>
<tr>
<td>BAVD Table(s)</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
<td>SNAC Records</td>
</tr>
</tbody>
</table>

#### Data Formatting and Schema Alignment

- **OR Transformations Required**: To strip quotes, use this GREL expression: `value smartSplit("\\") join("\"\")`
- **Ordering of actions**: Transpose cells across columns into rows. 2 columns (key: Date Type, value Date). Then replace field names with "Birth", "Death", or "Active".
- **Manual Transformations Required**: Use the GREL expression: `value smartSplit("\\") join("\"\")`, then split multi-valued cells on separator #.

#### Example 1
- **Example 1**: WCAU-TV (Television Station: Philadelphia, Penn.)
  - **Date**: 1932-03-15
  - **Status**: Active
  - **Type**: Radio stations

#### Example 2
- **Example 2**: Press, O. Leonard
  - **Date**: 1966
  - **Status**: Birth

#### Example 3
- **Example 3**: Duquesne University
  - **Date**: 1966-09
  - **Status**: Death

#### Notes
- Formatting in Airwaves is aligned with SNAC preferred formatting—if entity has LCNAF entry, that formatting can be used and seems to be SNAC aligned as well.
- All date formats will upload exactly as submitted, no need for reformatting.

---

**Quick Reference Guide**
batch upload workflow: exporting IDs

1. reconcile authority records
2. research + write SNAC records
3. configure OpenRefine
4. export data + import to OR
5. verify edits + export
6. upload to SNAC
7. map to SNAC schema
8. format data for SNAC upload
### Batch Upload Workflow: Exporting IDs

**OpenRefine**

SNAC Records BAVD SNAC prod Join March 2023 csv

| Rows | 236 | Show as: | rows | records | Show: | 5 | 10 | 25 | 50 | 100 | 500 | 1000 | rows |
|------|-----|----------|------|---------|------|---|----|----|----|-----|------|-------|------|------|
| **All** | | | **snaclID** | **entityType** | **BAVO SNAC Resource ID (Contributors) (from bavdCPF)** | **BAVO SNAC Resource ID (ReferencedIn) (from bavdCPF)** | **BAVO SNAC Resource ID (Creators) (from bavdCPF)** | **snaclAssociatedWith** |
| 6. | 87813837 | person | 11670964, 11671352, 11672402, 11672402 | | | | | |
| 6. | 87813853 | person | 11670680, 11670678, 11670984, 11670988, 11670991, 11671352, 11672402, 11672402 | | | | | |
| 7. | 87813985 | person | 11699845, 11670250 | | | | | |
| 8. | 87813880 | person | 11671351, 11670357 | | | | | |
| 9. | 87813894 | person | 11672249, 11665609, 11670236, 11672334, 1167234, 11672334, 11672334 | | | | | |
| 10. | 87813909 | corporateBody | 11670965, 11670699, 11670991, 11670682, 11670679, 11670680, 11670681, 11670682 | | | | | |

*An example starting SNAC relationships dataset in OpenRefine, before data formatting*
batch upload workflow: sticking points

❖ accidentally duplicating records and their contents
❖ figuring out unexpected behavior in OpenRefine
❖ development vs. production SNAC errors - FIXED!
❖ OpenRefine terminology - records mode vs. rows mode
batch upload workflow: reading the docs

SNAC data harvesting script, from workflow

data formatting and schema alignment guide

creating SNAC constellations workflow
resources

❖ Broadcasting A/V Data website: https://bavd.unlockingtheairwaves.org/
❖ SNAC OpenRefine extension: https://github.com/snac-cooperative/snac-openrefine-extension
❖ OpenRefine website: https://openrefine.org/
❖ BAVD documentation hub - see docs 9, 12, and 13: https://github.com/umd-mith/broadcasting/tree/main/workflows-docs
❖ SNAC OpenRefine extension documentation (by SNAC team): https://docs.google.com/document/d/1HjhKHAWXe6Xkh8F-qBnsl-psyApd2UPBLX2OgaVbA2Q/edit?usp=sharing
❖ SNAC technical resources hub: https://portal.snaccooperative.org/node/489
❖ Library Carpentry OpenRefine lesson: https://librarycarpentry.org/lc-open-refine/
❖ SNAC-dev: https://snac-dev.iath.virginia.edu/
❖ Sample data:
  ➢ CPFs: https://airtable.com/appMq6PvC5azYIsdt/shr1QKR1k0eGgPI2b/tblQI1mC0vnrEwv2T
  ➢ resources: https://airtable.com/appMq6PvC5azYIsdt/shrhRgNNWR1UewTEJ/tblpppSJiZmbCDWZh
  ➢ join table: https://airtable.com/appMq6PvC5azYIsdt/shrI8TproJdnjHRJh/tblQI1mC0vnrEwv2T
Emily Frazier
Collections Data Assessment Coordinator, University of Maryland Libraries
emilyf1(at)umd.edu

Stephanie Sapienza
Project Director, Broadcasting A/V Data Digital Humanities Archivist, Maryland Institute for Technology in the Humanities
sapienza(at)umd.edu