

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

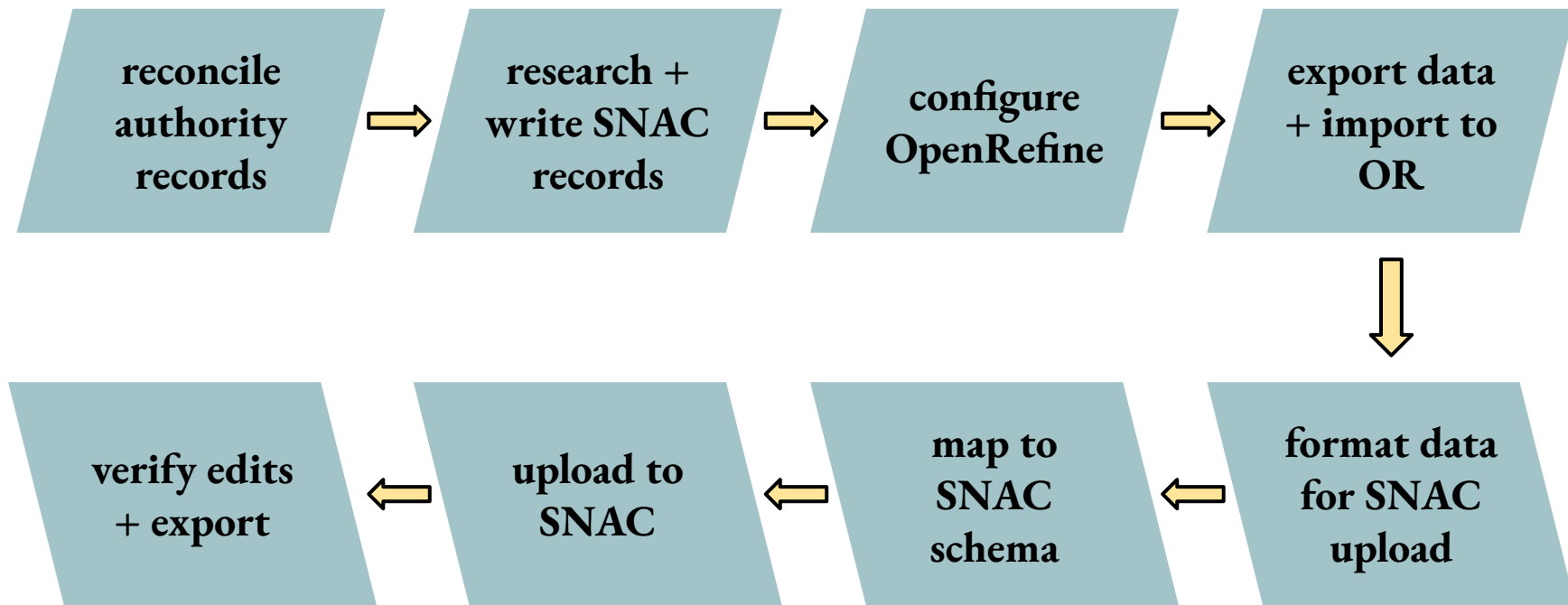
project background

- ❖ 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



batch upload workflow: overview

repeats 3x: CPFs, resources, relationships

reconcile
authority
records



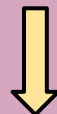
research +
write SNAC
records



configure
OpenRefine



export data
+ import to
OR



verify edits
+ export



upload to
SNAC

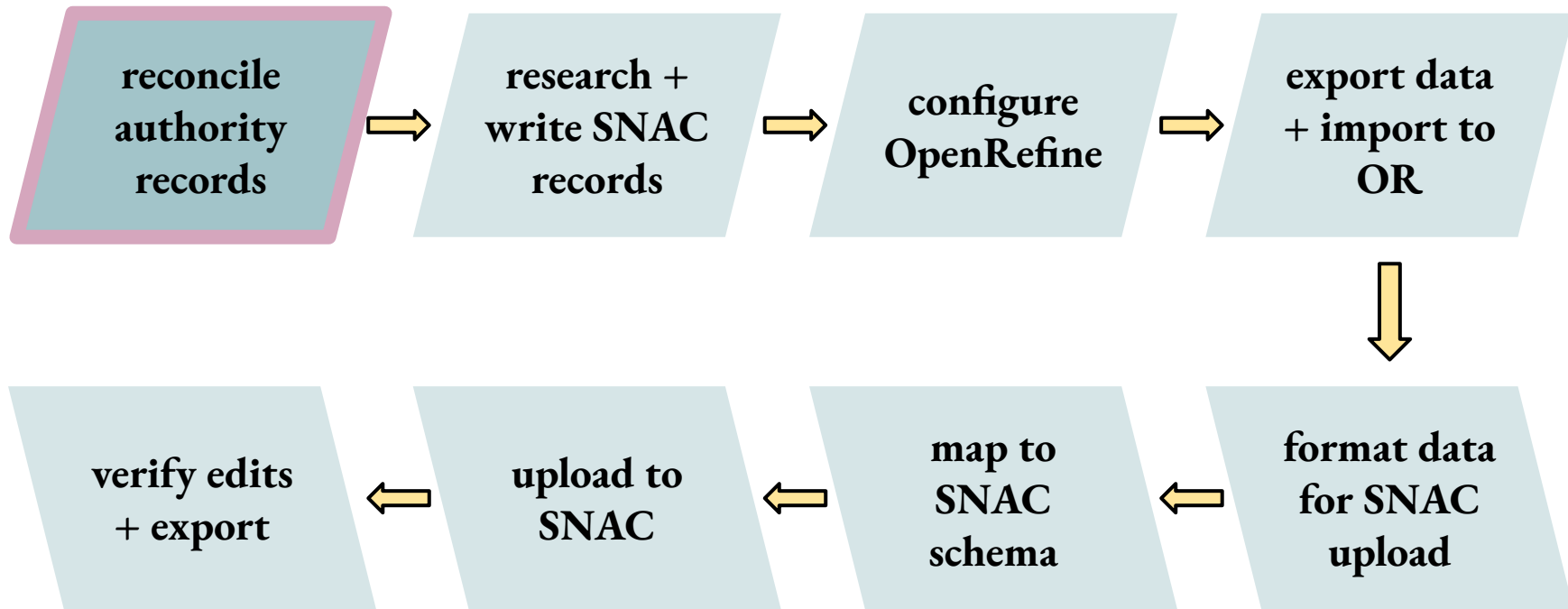


map to
SNAC
schema



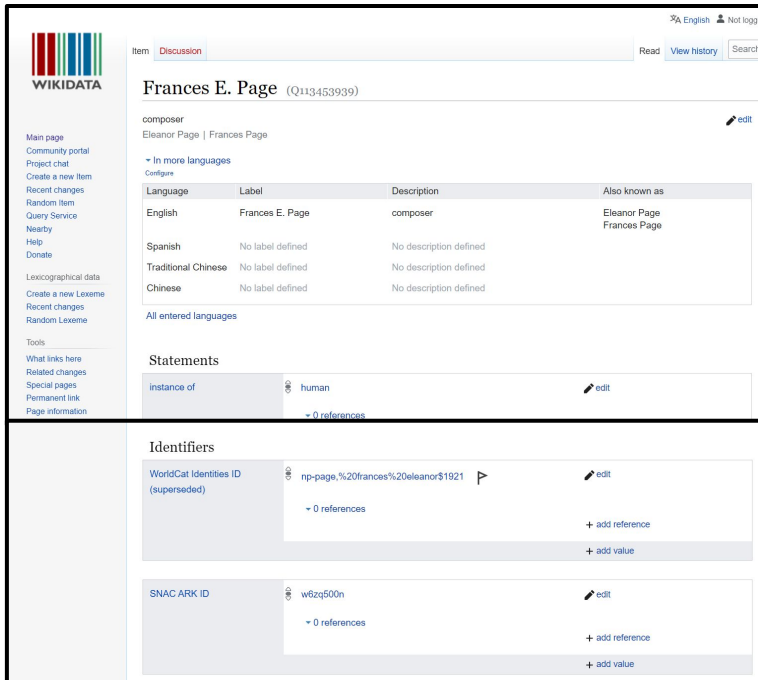
format data
for SNAC
upload

batch upload workflow: reconciling



batch upload workflow: reconciling

Frances Eleanor Page composer for radio station KUT



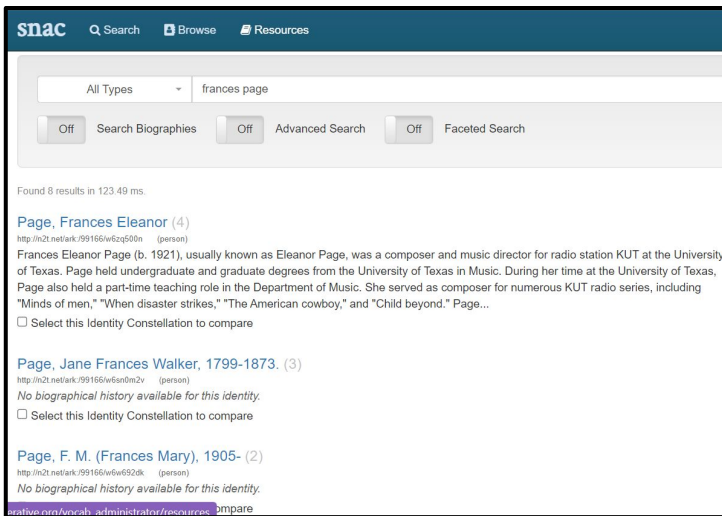
The screenshot shows the Wikidata page for Frances E. Page (Q113453939). The page is in English and shows the item's details, including its role as a composer. A table lists the item's labels in various languages: English (Frances E. Page), Spanish (No label defined), Traditional Chinese (No label defined), and Chinese (No label defined). The 'Statements' section shows the item is an instance of 'human'. The 'Identifiers' section lists two identifiers: 'WorldCat Identities ID (superseded)' with the value 'np-page.%20frances%20eleanor%201921' and 'SNAC ARK ID' with the value 'w6zq500n'. Both identifiers have '0 references' and options to 'add reference' or 'add value'.

method 1: pull SNAC ID from Wikidata

method 2: SNAC reconciliation in OpenRefine



The screenshot shows the 'Add standard service' dialog in OpenRefine. It prompts the user to 'Enter the service's URL' and shows the URL 'https://openrefine.snaccooperative.org/'. There are 'Add service' and 'Cancel' buttons at the bottom right.

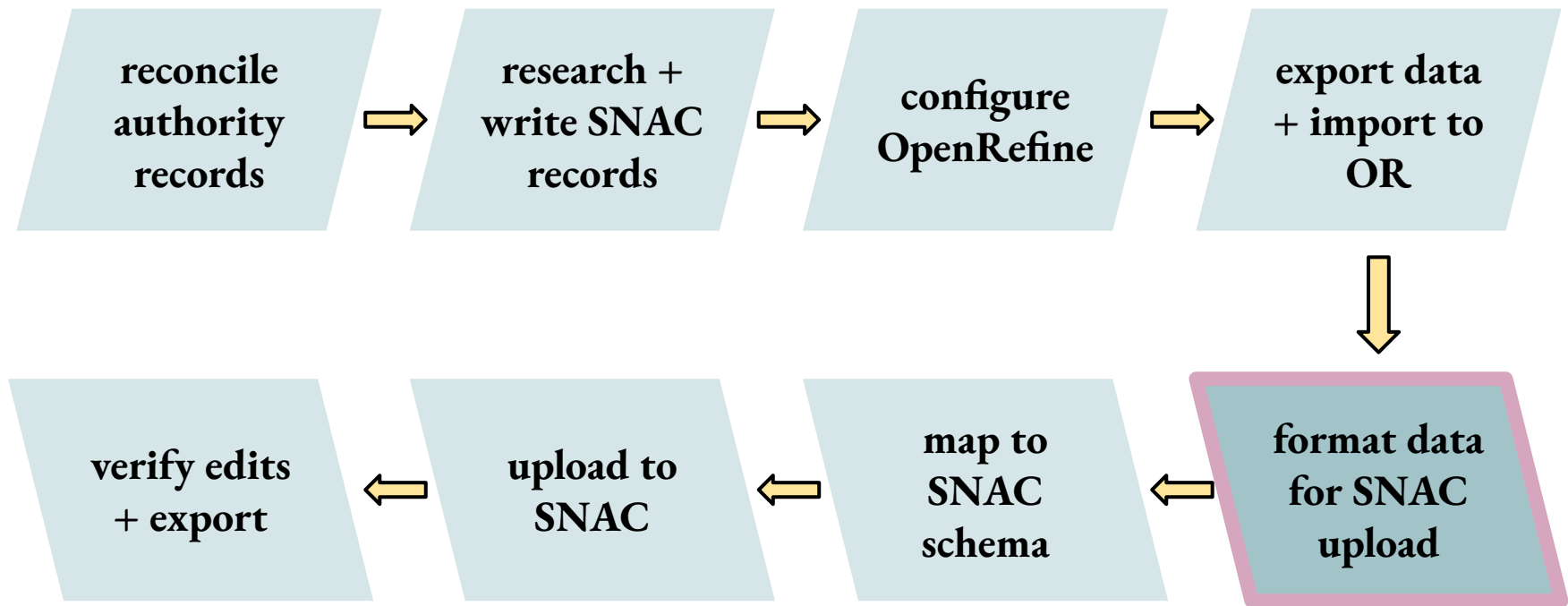


The screenshot shows the SNAC search interface. The search term 'frances page' has been entered, and 8 results are shown in 123.49 ms. The results list three entries:

- Page, Frances Eleanor (4)**
http://n2t.net/ark:/99166/w6zq500n (person)
Frances Eleanor Page (b. 1921), usually known as Eleanor Page, was a composer and music director for radio station KUT at the University of Texas. Page held undergraduate and graduate degrees from the University of Texas in Music. During her time at the University of Texas, Page also held a part-time teaching role in the Department of Music. She served as composer for numerous KUT radio series, including "Minds of men," "When disaster strikes," "The American cowboy," and "Child beyond." Page...
 Select this Identity Constellation to compare
- Page, Jane Frances Walker, 1799-1873. (3)**
http://n2t.net/ark:/99166/w6sn0m2v (person)
No biographical history available for this identity.
 Select this Identity Constellation to compare
- Page, F. M. (Frances Mary), 1905- (2)**
http://n2t.net/ark:/99166/w6w52z0k (person)
No biographical history available for this identity.
 Select this Identity Constellation to compare

method 3:
searching in
SNAC

batch upload workflow: OR data formatting



batch upload workflow: OR data formatting

The screenshot shows the OpenRefine interface with a dataset titled "SNAC Records SNAC prod CPFs March2023 csv". The interface includes a top navigation bar with "Open...", "Export", and "Help" buttons. Below the title, it indicates "30 rows" and provides options to "Show as: rows records" and "Show: 5 10 25 50 100 500 1000 rows". A table with 11 columns is displayed, showing two rows of data. The columns are: All, SNAC ID, Entity Type, Linked CPF Authority, Date of Birth, Date of Death, Active Date 1, Active Date 2, Subject(s), Places, Occupation(s), Activities, and Biographical or Histor. The first row (ID 3) is for Cecil S. Bidlack, born July 2, 1903, died October 16, 1995, an Engineer in Ohio, Columbus (Ohio). The second row (ID 4) is for John Strother Clayton, born April 17, 1925, died April 25, 2019, a Radio Executive in Chapel Hill (N.C.), North Carolina. The biographical text in the last column is partially visible and contains HTML tags like <p> and </p>.

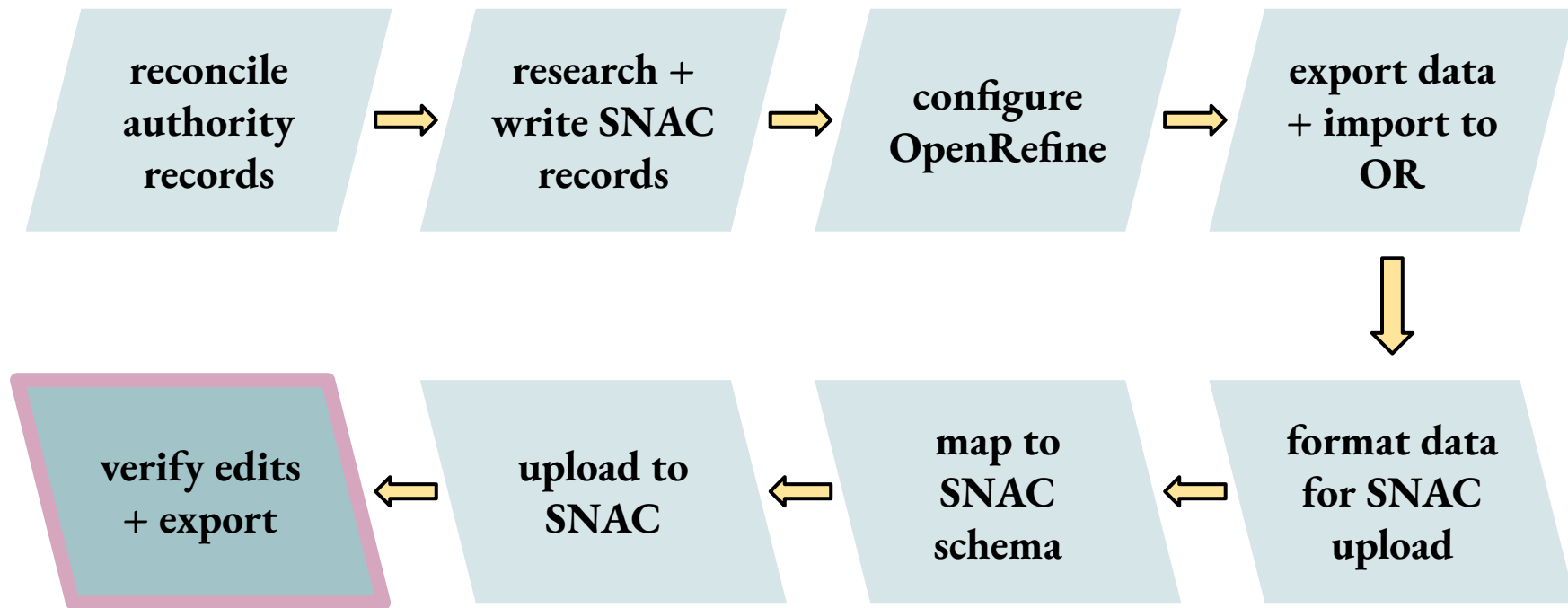
All	SNAC ID	Entity Type	Linked CPF Authority	Date of Birth	Date of Death	Active Date 1	Active Date 2	Subject(s)	Places	Occupation(s)	Activities	Biographical or Histor
☆	3. [Get ID 3]	person	"Bidlack, Cecil S."	July 2, 1903	October 16, 1995			Public broadcasting,Radio,Educational broadcasting,Public radio,Radio broadcasting	Ohio,Columbus (Ohio)	Engineer		</citation> <p>Cecil S. Bidlack was bo 1903. He attended Ohio Ste University, which he gradua in 1925. In 1928, he werkec Assistant Announcer and O for WEAO radio station at C University. Bidlack worked z Television Engineer for the Association of Educational Broadcasters from at least 1958. He authored the TV T Tips column in the NAEB's i newsletter, and in 1958 alsca Editor for the NAEB Eng Newsletter. Bidlack died on 16, 1995.</p></citation>Fro combined records of the Na Association of Educational Broadcasters (NAEB), held University of Maryland and Wisconsin Historical Societ, compiled as part of the Unk Airwaves project (unlockingtheairwaves.org).</citation>
☆	4. [Get ID 4]	person	"Clayton, John Strother"	April 17, 1925	April 25, 2019			Public broadcasting,Radio,Educational broadcasting,Public radio,Radio broadcasting	Chapel Hill (N.C.),North Carolina	Radio Executive		<p>Dr. John Strother Claytc born April 17, 1925 in Arkan began his career in radio wi Forces Radio in France afte War II. Clayton was writa director for many successfu programs produced by the l of North Carolina around th

an example starting SNAC CPF dataset in OpenRefine, before data formatting

batch upload workflow: OR data formatting

	A	D	E	F	G	
1	SNAC OR Schema Field Name	Name Entry	Exist Date	Exist Date Type	Subject	Pla
2	Airwaves Field Name(s)	Linked CPF Authority	Date of Birth; Date of Death; Active Date 1; Active Date 2	Date of Birth; Date of Death; Active Date 1; Active Date 2	Subject(s)	Pla
3	Airwaves Table(s)	SNAC Records	SNAC Records	SNAC Records	SNAC Records	SN
4	BAVD Field Name(s) SAMPLE DATA	bavdCPF	birthDate; deathDate	birthDate; deathDate	subjects	plac
5	BAVD Table(s)	SNAC Records	SNAC Records	SNAC Records	SNAC Records	SN
6	Related Fields		Date Type	Date		Pla
7	For Use With	All	All	All	All	All
8	OR Transformations Required	To strip quotes, use this GREL expression: value.smartSplit(",").join("#")		Transpose cells across columns into rows, 2 columns (key: Date Type; value: Date); Then replace field names with "Birth", "Death", or "Active"	Use this GREL expression: value.smartSplit(",").join("#"), then split multi-valued cells on separator #	Use valu mul
9	Ordering of actions					
10	Manual Transformations Required					
11	Example 1	WCAU-TV (Television Station : Philadelphia, Penn.)	1932-03-15	Active	Radio stations	Ala
12	Example 2	Press, O. Leonard	1956	Birth		
13	Example 3	Duquesne University	1968-09	Death		Wa
14	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		SNAC now uses FAST-aligned headings, so there should be no subordination/pre-coordination of subject terms. E.g., no "--" between terms	Do in A form
15	<i>Release v1.0.0-beta-18</i>					

batch upload workflow: exporting IDs



batch upload workflow: exporting IDs

OpenRefine SNAC Records BAVD SNAC prod Join March2023 csv [Permalink](#) Open... Export Help

236 rows Extensions Wikibase

Show as: rows records Show: 5 10 25 50 100 500 1000 rows « first « previous 1 next » last »

All	snacID	entityType	BAVD SNAC Resource ID (Contributors) (from bavidCPF)	BAVD SNAC Resource ID (ReferencedIn) (from bavidCPF)	BAVD SNAC Resource ID (Creators) (from bavidCPF)	snacIDassociatedWith
5.	87813837	person	11670994, 11671352, 11672402, 11672402		11671797, 11671717, 11671940, 11671950, 11671955	
6.	87813853	person	11670680, 11670681, 11670682, 11670683, 11670684, 11670685, 11670686, 11670687, 11670688, 11670689, 11670690, 11670691, 11670692, 11670693, 11670694			52986448 87813909
7.	87813867	person	11669845, 11670250		11672334, 11672334, 11672334, 11672334, 11672334, 11672334, 11672334, 11672334, 11672334, 11672334, 11672334	59459688, 13198872
8.	87813880	person	11671351, 11670357			52986448
9.	87813894	person	11672249, 11669505, 11672336, 11672336, 11672334, 11672334, 11672334, 11672334, 11672334, 11672334, 11669547, 11669471, 11669467, 11672252, 11672251, 11669491, 11669565, 11669692, 11672283, 11669813, 11669828, 11669840, 11669848, 11669850, 11669861, 11669872, 11669883, 11669887, 11669888, 11669902, 11669912, 11669915, 11669917, 11669920, 11669925, 11669927, 11669938, 11669943, 11669945, 11669948, 11669964, 11669981, 11669986, 11669989, 11669998, 11670001, 11670008, 11670010, 11670016, 11670021, 11670023, 11670027, 11670032, 11670041, 11670057, 11670063, 11670086, 11670099, 11670103, 11670105, 11670115, 11670121, 11670136, 11670137, 11670163, 11670172, 11670176, 11670177, 11670185, 11670186, 11670189, 11670196, 11670212, 11670244, 11670253, 11670261, 11670263, 11670276, 11670277, 11670278, 11670318, 11670323, 11670329, 11670356, 11670378, 11670379, 11670382, 11670392, 11670399, 11670401, 11670402, 11670407, 11670411, 11670414, 11670421, 11670446, 11670449, 11670451, 11670459, 11670473, 11670477, 11670479, 11670496, 11670499, 11670503, 11670510, 11670517, 11670537, 11670544, 11670565, 11670566, 11670572, 11670574, 11670579, 11670581, 11670601, 11670615, 11670617, 11670621, 11670628, 11670630, 11670631, 11670632, 11670638, 11670647, 11670659, 11669857, 11670492, 11669951, 11670210, 11670422, 11670481		13198872, 29231884, 12725674	
10.	87813909	corporateBody	11670665, 11670666, 11670667, 11670668, 11670669, 11670670, 11670671, 11670672, 11670673, 11670674, 11670675, 11670676, 11670677, 11670678, 11670679, 11670680, 11670681, 11670682, 11670683, 11670684, 11670685, 11670686, 11670687, 11670688, 11670689, 11670690, 11670691, 11670692, 11670693, 11670694			

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

```

1  const fetch = require("node-fetch")
2  const fs = require("fs")
3  const path = require("path")
4  const { parse } = require("json2csv")
5
6  const fields = ["SNAC ID", "SNAC AKA", "Related IDs", "CPF Type", "Name Entry", "Subject", "Occupation", "Activities", "BlogList", "Date (from)",
7
8  const getSnacs = async () => {
9
10  const allData = []
11  const allDataJSON = []
12

```

SNAC data harvesting script, from [workflow](#)

	A	D	E	F	G	H	I
1	SNAC OR Schema Field Name	Name Entry	Exist Date	Exist Date Type	Subject	Place	Place Role
2	Airwaves Field Name(s)	Linked CPF Authority	Date of Birth, Date of Death, Active Date 1, Active Date 2	Date of Birth, Date of Death, Active Date 1, Active Date 2	Subjects	Places	n/a
3	BAVO Field Name(s) (EXAMPLE)	SNAC Records	SNAC Records	SNAC Records	SNAC Records	SNAC Records	n/a
4	BAVO Field Name(s) (EXAMPLE)	Linked CPF	birthDate, deathDate	birthDate, deathDate	subjects	places	n/a
5	BAVO Field Name(s)	SNAC Records	SNAC Records	SNAC Records	SNAC Records	SNAC Records	n/a
6	Related Fields	Date Type	Date	Date	Place Role	Place	Place
7	For Use With	All	All	All	All	All	All
8	OR Transformations Required	To strip quotes, use this GREL expression: <code>value smartSplit(" ", join " ")</code>		Transpose cells across columns into rows. 2 columns (any Date Type: value Date). Then replace field names with 'Birth', 'Death', or 'Active'	Use this GREL expression: <code>value smartSplit(" ", join " ")</code> , then split multi-valued cells on separator #	Use this GREL expression: <code>value smartSplit(" ", join " ")</code> , then split multi-valued cells on separator #	Facet to rows containing values in field; then add column based on column with "AssociatedPlace" as value.
9	Ordering of actions						After transforming Place field
10	Manual Transformations Required						
11	Example 1	WCALU-TV (Television Station - Philadelphia, Penn.)	1952-03-15	Active	Radio stations	Alabama	
12	Example 2	Press. O. Leonard	1956	Birth			
13	Example 3	Duquesne University	1968-09	Death		Washington (D.C.)	
14	Notes	Formatting in Airwaves is aligned with SNAC preferred formatting - if entry has LCNAP 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		SNAC now uses FAST-aligned headings, so there should be no subordination/pre-coordination of subject terms. E.g., no "-" between terms	Do not use commas when adding places in Airwaves; should be RDA-compliant format	
15	Release v1.0.0-beta-18						

[data formatting and schema alignment guide](#)

Creating/Updating SNAC Authorities

Unlocking the Airwaves and Broadcasting A/V Data Workflow Documentation

- Date of Death
 - Active Date 1
 - Active Date 2
 - Subject(s)
 - Places
 - Occupations
 - Activities
 - Wikidata URL
 - VIAF URL
 - WorldCat URL
 - LOC URL
 - NARA URL
- Create project in OpenRefine
 - Follow instructions above in Section B(3) to configure SNAC in OpenRefine
- Prep/refine data in OpenRefine:
 - Strip quotes from **Name Entry/Linked CPF Authority** by using the GREL expression: `value smartSplit(" ", join " ")`
 - Click on column **Date of Birth**, select Transpose. Select four columns (**Date of Birth**, **Date of Death**, **Active Date 1**, **Active Date 2**). Transpose to two columns. Name key column **Date Type**, name Value column **Date**.
 - Use Edit cells > Replace to change values in **Date Type**: 'Date of Birth' to 'Birth'; 'Date of Death' to 'Death'; 'Active Date 1' to 'Active'; 'Active Date 2' to 'Active.'
 - Strip quotes from **Subjects** field by using the GREL expression: `value smartSplit(" ", join " ")`
 - Split multi-valued cells in the **Subjects** field on the separator '#'
 - Strip quotes from **Places** field by using the GREL expression: `value smartSplit(" ", join " ")`
 - Split multi-valued cells in the **Places** field on the separator '#'
 - Facet to rows containing values in **Place** field, then add column based on **Place** column with "AssociatedPlace" as the value, and title it "Place Role", then clear facet

[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=s_haring
- ❖ SNAC technical resources hub: <https://portal.snaccooperative.org/node/489>
- ❖ Library Carpentry OpenRefine lesson: <https://librarycarpentry.org/lc-open-refine/>
- ❖ Programming Historian OpenRefine lesson:
<https://programminghistorian.org/en/lessons/cleaning-data-with-openrefine>
- ❖ SNAC-dev: <https://snac-dev.iath.virginia.edu/>
- ❖ Sample data:
 - CPFs: <https://airtable.com/appMq6PvC5azYIsdt/shr1QKR1k0eGgPI2b/tblQI1mC0vnrEwv2T>
 - resources: <https://airtable.com/appMq6PvC5azYIsdt/shrhRgNNWR1UewTEJ/tblpppSIjZmbCDWZh>
 - join table: <https://airtable.com/appMq6PvC5azYIsdt/shrI8TproJdnjHRlh/tblQI1mC0vnrEwv2T>

thank you! + contact info

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`