Creating/Updating SNAC Authorities

Unlocking the Airwaves and *Broadcasting A/V Data* Workflow Documentation

Related documentation:
- SNAC schema formatting and alignment guide
- SNAC reconciliation
- Transforming SNAC records

A. Create SNAC Holding Repositories

Before new SNAC Resources can be created, SNAC needs to have an entry for all repositories which will be attached to those resources. *Only one repository can be attached to one resource*.

Collect data from the target resource (finding aid or other descriptive resource) to fill in the fields of the new resource. Copying and pasting from the resource's web page will save time.

SNAC also prompts users to designate the **Language** and **Script** of the resource, and to enter a value for the **Holding Repository**.

For example, Maryland Institute for Technology in the Humanities has a standard SNAC CPF record, but because it is designated as a holding repository, it also has a tab titled "Holdings" that lists the content attached to it in SNAC.

B. Create SNAC Resources

The next phase is to create, refine, and upload a list of SNAC resources. These will be at various levels of granularity:

- A Program **Series**
- A **Folder** of Documents
- An **Item** from a Folder of documents

Creating/Updating SNAC Authorities

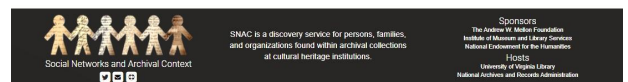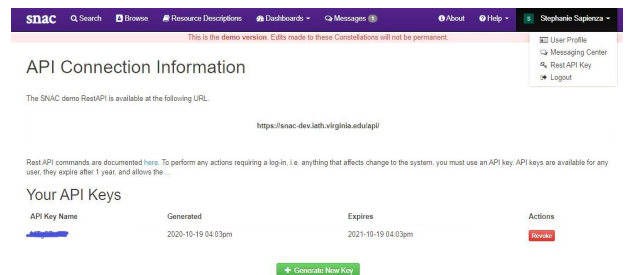*Unlocking the Airwaves* and *Broadcasting A/V Data* Workflow Documentation

---

The steps involved in this process include:

1. Export a list of resource data from Airtable as as CSV file:
   a. Program Series (from Series table): Series Title; Series Description; Slug; URL; # of programs
   b. Folder of Documents: Box #, Folder #, Title, URL, Description, # page scans
   c. Item from Folder: Box #, Folder #, Item #, Title, URL, Description, # page scans

2. Prep/refine data for OpenRefine in Excel to end up with the following six fields:

   a. Title
      i. Ex 1 (program series): National Association of Educational Broadcasters (NAEB) audio tapes | National Educational Radio Network Broadcasts, 1952-1973 | Listen America
      ii. Ex. 2 (folder): National Association of Educational Broadcasters Records, 1925-1977 (bulk 1950-1970) | Board of Directors, Correspondence, 1954 (Box 20, Folder 1)
      iii. Ex. 3 (item): National Association of Educational Broadcasters Records, 1925-1977 (bulk 1950-1970) | Publications, 1930-1939 | Proposed Aims and Purposes of the National Association of Educational Broadcasters (September 1939) (Box 101, Folder 1)
   b. Link
      i. Ex. 1 (program series): https://unlockingtheairwaves.org/programs/listen-america/
      ii. Ex. 2 (folder): https://unlockingtheairwaves.org/document/naeb-b020-f01/
      iii. Ex. 3 (item): https://unlockingtheairwaves.org/document/naeb-b101-f01-13/

   c. Abstract
      i. Ex. 1 (program series): A series of 13 programs featuring the works of selected contemporary American authors.
      ii. Ex. 2 (folder): 1954 correspondence involving the NAEB board of directors and staff, especially executive director Harry Skornia. Includes financial reports and surveys used for board votes on issuTracy Kessleres such as committee names and the NAEB's position on subscription television. Correspondence addresses topics including the hiring of Harold Hill as associate director and Morris Novik as a personnel consultant.
      iii. Ex. 3 (item): Short document detailing the aims and purposes of the NAEB.

---

    d. Extent

        i. Ex. 1 (program series): 37 digital audio files

        ii. Ex. 2 (folder): 173 digital page scans

        iii. Ex. 3 (item): 12 digital page scans

    e. Holding Repository (See Section A, ex. 85373668)

    f. Resource Type (DigitalArchivalResource)

3. Create project in OpenRefine

    a. Configure SNAC Extension

        i. general [OpenRefine extension instructions](#)

        ii. [download SNAC extension](#)

    b. If you have multiple downloaded CSVs of resource data, you can open all 3 into one OpenRefine project. Fields across different CSVs with the same name will be automerged into a single column, but any fields that will map to the same schema element with different names in different CSVs will have to be merged in OpenRefine.

    c. Configure SNAC API key

        i. Log into SNAC account, click on Profile, 'Rest API Key'

        ii. Generate New Key, copy the key

        iii. Navigate back to OpenRefine, click on the SNAC pulldown, Manage API Key, and paste the key

        iv. Save this key somewhere, such as in a password manager. It will eventually expire, the expiration date is recorded in your SNAC account.

4. Map fields to SNAC Schema
    a. Click on the SNAC pulldown, select 'Edit SNAC Schema'
    b. Map fields to SNAC Schema
    c. Click 'Save schema'

Creating/Updating SNAC Authorities

Unlocking the Airwaves and *Broadcasting A/V Data* Workflow Documentation

5. Submit Resources to SNAC using OpenRefine Extension
   a. Click 'Preview' tab to preview data prior to submission
   b. Click on the SNAC pulldown, select Upload Edits to SNAC
   c. In the 'Upload your edits' modal window, select 'Upload Using Key'
   d. Refresh the OpenRefine project in your browser. The submission process should result in the data set updating with new SNAC IDs for new resources

6. Download new SNAC IDs; ingest into Airtable using CSV Import
   a. Click on Export button in OpenRefine, select CSV download
   b. You will need to export 3 separate CSVs for each of the 3 tables in Airtable where you need to import the resulting IDs. Use facets in OpenRefine to do this.
   c. Use the CSV Import app in Airtable to import the new SNAC Resource IDs to the three tables: Document Metadata-Folders; Document Metadata-Items; Programs Metadata.

C. Create or Update SNAC CPF Entities

The next phase is to create, refine, and upload a set of CPF Entities to SNAC. The CPF entities tracked by Airtable are located in the CPF Entities table. Entities have been given the following designations: **Crucial**, **Somewhat**, or **Mild**.

Crucial statuses mean that this particular entity has been deemed highly important to the NAEB collections. This may be because the entity is associated with a high amount of content (documents and program series), or it may be because the person or organization is particularly important in the NAEB's history. Entities with the Crucial status will receive a **Rich Profile** in SNAC. Rich Profiles contain all of following fields:

Creating/Updating SNAC Authorities

Unlocking the Airwaves and *Broadcasting A/V Data* Workflow Documentation

Tag/Field	Notes
tag: <entityType> Airtable field: Entity Type	Person or Corporate Body
tag: <nameEntry> <preferredForm> Airtable field: Linked CPF Authority	Name(s)
tag: <existDates> <fromDate> Airtable field: Date of Birth/Active Date 1	Date of Birth (or Active Start for Corporate Bodies)
tag: <existDates> <toDate> Airtable field: Date of Death/Active Date 2	Date of Death (or Active End for Corporate Bodies)
tag: <place> <placeEntry> <placeRole> = Associated With Airtable field: Places	Places Lived, Located, or Worked
tag: <biogHist> Airtable field: Biographical or Historical Note	2-3 paragraph biography or history
tag: <occupations> Airtable field: Occupation(s)	Occupations (People only)
tag: <function> <term> Airtable field: Activities	Activities (Corporate bodies only)
tag: <localDescription> <term> @localType = Associated Subject Airtable field: Subject(s)	Subject Terms
tag: <relations> <cpfRelation> <relationEntry> @xlink:arcrole = Associated With	Related CPFs

Creating/Updating SNAC Authorities

Unlocking the Airwaves and *Broadcasting A/V Data* Workflow Documentation

Airtable field: Relations (Associated With) SNAC IDs	
tag: \<relations> 　　　\<cpfRelation> 　　　\<relationEntry> 　　　@xlink:arcrole = Same As Airtable field: VIAF URL; LOC ID URL; WorldCat URL; Wikidata URL; NARA URL	Related external CPF records (other authority control sources)

Somewhat statuses mean that this entity is still important to the NAEB collections, but is either less heavily represented in the collections or isn't as much of a key figure in the organization's history. Entities with the **Somewhat** status will receive a **Minimal Profile** in SNAC. Minimal Profiles contain the following fields:

Tag/Field	Notes
tag: \<entityType> Airtable field: Entity Type	Person or Corporate Body
tag: \<nameEntry> 　　　\<preferredForm> Airtable field: Linked CPF Authority	Name(s)
tag: \<place> 　　　\<placeEntry> 　　　\<placeRole> = Associated With Airtable field: Places	Places Lived, Located, or Worked
tag: \<localDescription> 　　　\<term> 　　　@localType = Associated Subject Airtable field: Subject(s)	Subject Terms
tag: \<relations> 　　　\<cpfRelation> 　　　\<relationEntry> 　　　@xlink:arcrole = Associated With Airtable field: Relations (Associated With) SNAC IDs	Related CPFs

Creating/Updating SNAC Authorities

Unlocking the Airwaves and *Broadcasting A/V Data* Workflow Documentation

tag: <relations> <cpfRelation> <relationEntry> @xlink:arcrole = Same As Airtable field: VIAF URL; LOC ID URL; WorldCat URL; Wikidata URL; NARA URL	Related external CPF records (other authority control sources)

With the following fields populated only when easily/immediately available from another reconciled source like Wikidata/VIAF, etc:

Tag/Field	Notes
tag: <existDates> <fromDate> Airtable field: Date of Birth/Active Date 1	Date of Birth (or Active Start for Corporate Bodies)
tag: <existDates> <toDate> Airtable field: Date of Death/Active Date 2	Date of Death (or Active End for Corporate Bodies)
tag: <biogHist> Airtable field: Biographical or Historical Note	2-3 paragraph biography or history
tag: <occupations> Airtable field: Occupation(s)	Occupations (People only)
tag: <function> <term> Airtable field: Activities	Activities (Corporate bodies only)

All of the above work will be completed within the Airwaves Application Database, in the SNAC Records table. After the entities have been marked as complete (either with the Rich or Minimal profile), they can be downloaded and refined/prepped for upload to OpenRefine and SNAC. The steps involved in this process include:

1. Export a list of CPF data from Airtable as as CSV file:
   a. Linked CPF Authority
   b. Entity Type
   c. Biographical or Historical Note
   d. Date of Birth

---

        e. Date of Death

        f. Active Date 1

        g. Active Date 2

        h. Subject(s)

        i. Places

        j. Occupations

        k. Activities

        l. Wikidata URL

        m. VIAF URL

        n. WorldCat URL

        o. LOC URL

        p. NARA URL

2. Create project in OpenRefine
   a. Follow instructions above in Section B(3) to configure SNAC in OpenRefine

3. Prep/refine data in OpenRefine:
   a. Strip quotes from **Name Entry/Linked CPF Authority** by using the GREL expression: `value.smartSplit(",").join("#")`
   b. 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.
   c. 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.'
   d. Strip quotes from **Subjects** field by using the GREL expression: `value.smartSplit(",").join("#")`
   e. Split multi-valued cells in the **Subjects** field on the separator '#'
   f. Strip quotes from **Places** field by using the GREL expression: `value.smartSplit(",").join("#")`
   g. Split multi-valued cells in the **Places** field on the separator '#'
   h. 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
   i. Strip quotes from **Occupations** field by using the GREL expression: `value.smartSplit(",").join("#")`
   j. Split multi-valued cells in the **Occupations** field on the separator '#'
   k. Strip quotes from **Activities** field by using the GREL expression: `value.smartSplit(",").join("#")`
   l. Split multi-valued cells in the **Activities** field on the separator '#'

---

     m. Click on the first of the External Related CPF URL fields **(VIAF URL; LOC ID URL; WorldCat URL; Wikidata URL; NARA URL)**, and select Transpose. Transpose cells across columns into rows, into 1 column. Select all the External Related CPF URL fields as the fields to transpose in the option window.

4. Map fields to SNAC Schema
     a. Click on the SNAC pulldown, select 'Edit SNAC Schema'
     b. Map fields to SNAC Schema
     c. Click 'Save schema'

5. Submit Resources to SNAC using OpenRefine Extension
     a. Click 'Preview' tab to preview data prior to submission
     b. Click on the SNAC pulldown, select Upload Edits to SNAC
     c. In the 'Upload your edits' modal window, select 'Upload Using Key'
     d. Refresh the OpenRefine project in your browser. The submission process should result in the data set updating with new SNAC IDs for new resources

6. Download new SNAC IDs; ingest into Airtable using CSV Import
     a. Click on Export button in OpenRefine, select CSV download
     b. Use the CSV Import app in Airtable to import the new SNAC Resource IDs to the SNAC records table

D. Create Relationships (CPF to Resource; CPF to CPF)

The final phase of creating SNAC authority constellations allows you to connect the CPFs and Resources that you created in the previous 2 steps, as well as connect CPFs to one another. These actions can be done at the same time, as they use the same infrastructure within the SNAC OpenRefine extension.

In order to cleanly export the relationships between these components, there is some **initial work that must be done in Airtable** specific to the Unlocking the Airwaves and Broadcasting A/V Data projects. Skip these instructions if you are not using Airtable and/or already have matched up your CPF to Resource relations and CPF to CPF relations using their IDs:

1. The CPF IDs and Resource IDs that you have just created and/or will be using in your relationships must be listed in Airtable with the records for the CPFs and Resource records.

2. Next, you will need a "chain" of lookup fields to connect the SNAC Records table to the table(s) where your resources live. In this case, you will need to go to the table "CPF Authorities" and look up the value of the field with the SNAC resource ID for all of your

tables of resources and all of the entity roles within them. For example, 4 lookup fields that do the following:

    a. "SNAC_PROD_LinkedSeriesCreator-ResourceID" - looks up the resource ID for each radio SERIES where the CPF entity is a CREATOR

    b. "SNAC_PROD_LinkedSeriesContrib-Resource ID" - looks up the resource ID for each radio SERIES where the CPF entity is a CONTRIBUTOR

    c. "SNAC_PROD_LinkedDocsFolderContribs-ResourceIDs" - looks up the resource ID for each FOLDER where the CPF entity is a CONTRIBUTOR

    d. "SNAC_PROD_LinkedDocsItemsContrib-ResourceIDs" - looks up the resource ID for each PAPER ITEM where the CPF entity is a CONTRIBUTOR

3. You will also need a formula field that combines all of the fields with the same ROLE together, because you want them all in one field later on. In this case, you'd use a formula to combine the values for fields (b), (c), and (d) above.

4. The last step the chain is to look up your resource ID values in the SNAC Records table. To do this, you will need to create lookup fields in the SNAC Records table that look at:

    a. "SNAC_PROD_LinkedSeriesCreator-ResourceID" from step (2)(a) - this is your CREATOR role resource ID field

    b. The combined formula field that rolls up (2)(b)-(2)(d) above - this is your CONTRIBUTOR role resource ID field

5. The other data you will need to prepare in Airtable are the CPF to CPF relations. When you researched your new CPF records, you probably identified the related CPFs. If those CPFs already existed in SNAC, then you would have a field with their SNAC IDs in your SNAC Records table. Now that new CPF records have been created in SNAC in section C above, you may have CPF to CPF relations that need to be added to the related CPF IDs field. (e.g, perhaps you just created CPF records for two related entities and want to relate them in this step)

When the relations have all been pulled into/updated in the SNAC Records table, you can then **export and move into OpenRefine**. The steps involved in this process include:

1. Export a list of relationship data from Airtable as a CSV with the following fields:

    a. LinkedMaterialsContrib-ResourceIDs (from Linked CPF Authority) - CONTRIBUTOR role resource IDs

    b. LinkedSeriesCreator-ResourceID (from Linked CPF Authority) - CREATOR role resource IDs

    c. Relations (Associated With) SNAC IDs - CPF to CPF relations

    d. Entity Type

---

    e. SNAC ID

2. Open the file in a new OpenRefine project and do the following transformations on it to prepare it for upload:
    a. On fields (1)(a) and (1)(b) above (the resource ID creator and contributor fields), split multi-valued cells on the separator ", " (that's a comma followed by a space).
    b. On the leftmost field out of (1)(a) and (1)(b) above, select the option in the dropdown menu to transpose cells across columns into rows. Select only the columns (1)(a) and (1)(b) as your columns to transpose. Use the two-column result option, and set the key to "CPF to Resource Relation Type" and the value to "SNAC Resource ID". After transposing, replace the values in the "CPF to Resource Relation Type" column with "creatorOf" (for creator role) or "contributorOf" (for contributor role). This has given you an OpenRefine-readable list of what resources you will be connecting your CPF record to.
    c. On the field (1)(c) above (the CPF to CPF relations field), split multi-valued cells on the separator "," (that's just a comma, no space).
    d. Create a facet to only show rows that have a value in the field (1)(c), then create a column using "add column based on this column" that fills in all matching rows with the value "associatedWith". Title this column "CPF to CPF Relation Type".
    e. On the field (1)(e) above (SNAC ID), use the fill down tool in the dropdown menu to fill in the empty cells in that column. Do the same on field (1)(d) above (entity type).
    f. Move field (1)(e) (SNAC ID) to the front of your OpenRefine table using the reorder option. This is necessary because OpenRefine will not read it as the key column that defines whether your project is in rows or records mode unless it is first.

3. Next, map your fields to the SNAC schema, first changing the type to "SNAC join model". If you have used the field titles suggested above, all your fields should match those in the join model schema, except for:
    a. (your title) SNAC ID : (schema) SNAC CPF ID
    b. (your title) Entity Type : (schema) CPF type
    c. (your title) Relations (Associated With) SNAC IDs : (schema) Related SNAC CPF ID

4. Submit Resources to SNAC using OpenRefine Extension
    a. Click 'Preview' tab to preview data prior to submission
    b. Click on the SNAC pulldown, select Upload Edits to SNAC

---

      c. In the 'Upload your edits' modal window, select 'Upload Using Key' (refer to earlier instructions if you have not yet configured your SNAC API key)

5. The upload will take a long time, scaled according to how large your batch was. Review results on SNAC and in the SNAC results column that will be added to your OpenRefine project upon conclusion of the upload.
         a. If you have "failure" and/or "error" responses, you may view your editor dashboard on the SNAC web interface and find you have locked edits. You should be able to unlock them, though the problem will recur with a retry unless you track down the source of the issue.

6. Nothing to export from this step after upload concludes, because it was only connecting previously existing CPF and resource IDs.