Transforming SNAC Records

This workflow was used for the Broadcasting A/V Data project, a related project to Unlocking the Airwaves. The Social Networks and Archival Context data needs and usage were nearly identical for the two projects. For BA/VD, the team elected to collect and transform SNAC data in-house, using the SNAC Application Programming Interface (API).

Code by Raff Viglianti, documentation by Emily Frazier

For the *Broadcasting A/V Data* project, data from Social Networks and Archival Context (SNAC) was harvested for display on the BAVD project website using the SNAC application programming interface (API). API documentation is available on the SNAC website, as is a test page for trying out commands. Before writing the code that interfaces with the API to gather the SNAC data, a list of requested fields was created, along with any specifications regarding formatting or separators between multiple items in a field. This code was used to connect to the SNAC API and fetch the records.

For this project, the following fields were collected from SNAC constellations:

- SNAC ID
- SNAC ARK URL
- Same as relations (external authority control URLs)
- CPF type (person/corporate body)
- Name entries (name/alternate names)
- Subjects
- Occupations
- Activities
- Biographical/Historical note
- Associated with relations (associated SNAC CPFs)
- Dates (start date and end date)
- Date types (birth, death, active)
- Places
- Place types (associated place)

The linked code above, in addition to fetching this data using a list of previously reconciled SNAC IDs, also does several custom transformations to "flatten" the constellations into a tabular format. It also cleans XML formatting from the biographical/historical notes. The following fields are altered as part of the data fetching and cleaning process:

- Biographical/Historical note: formatting tags (<p>) and namespace declarations (xmlns=) are stripped out of text
- Associated with relations: names of associated CPFs are joined with the IDs of the associated CPFs using " && "
- Places: plain text place names and "geoplace" names are both added to this field

- In any field where multiple values existed, values were wrapped in quotes and separated by commas inside straight brackets for the whole string of values

The result of data fetching and cleaning is written to a CSV file, which for the BA/VD project was then further transformed in OpenRefine before being ingested into Airtable.

The code works by connecting to the SNAC API at https://api.snaccooperative.org/. You can see details about different SNAC API commands from the API documentation linked above. Note that an API key is not required to fetch data from SNAC, though it is required for other actions not covered here. The data is returned by the API as JSON. You must have a list of SNAC IDs for your CPF records in order to fetch data.