Parsing JSON to CSV with OpenRefine

Example: University of Minnesota Libraries UMedia Repository records

Broadcasting A/V Data project

Open the JSON file with program records. In this case, the JSON file is showing the program records for the 20 objects on page 3 of results.

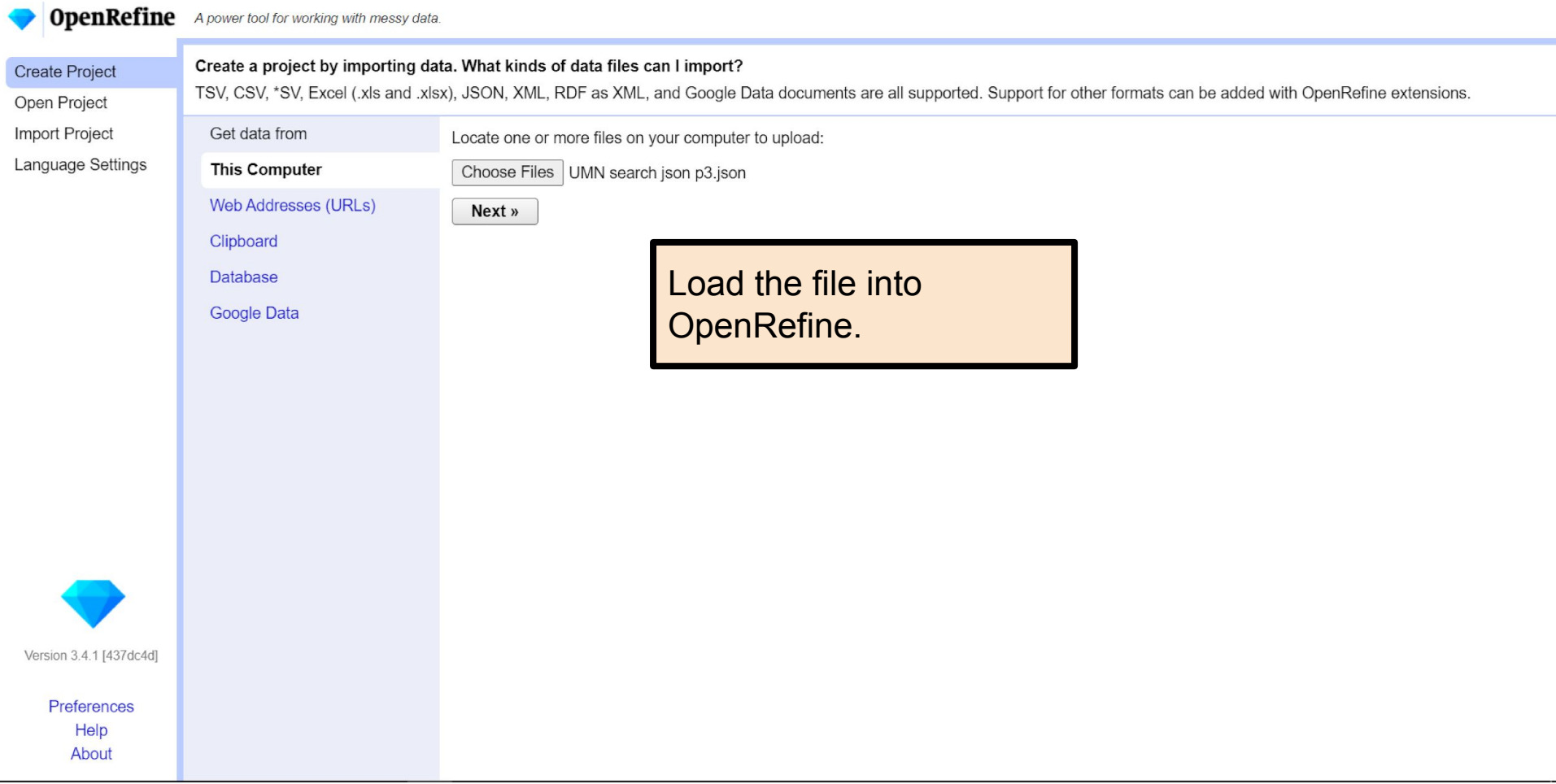
Maggie Kuhn and John Brantner, "The Sickness of Ageism"

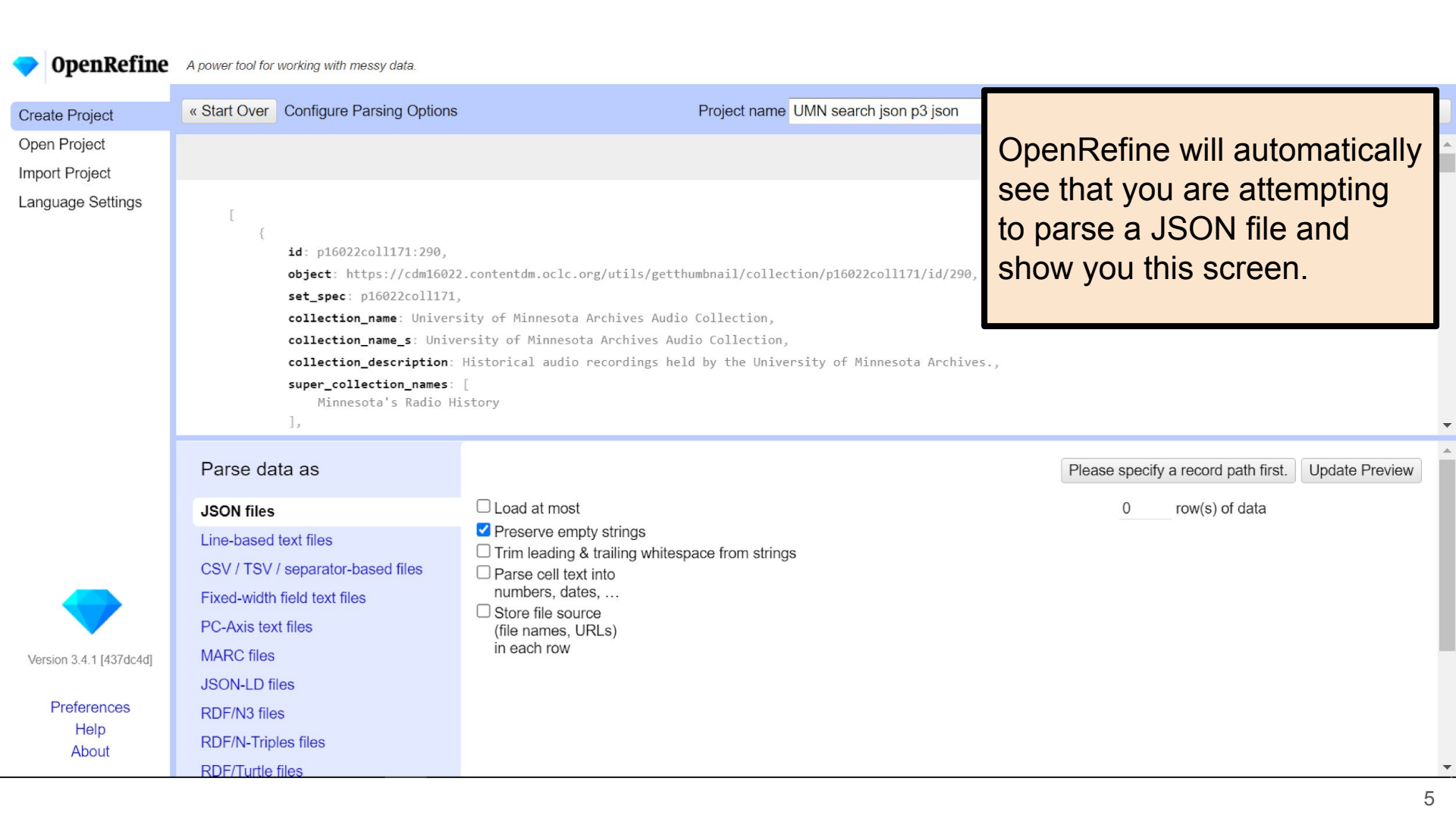
Description: That was Maggie Kuhn, National Conveener of the Gray Panthers. Earlier you heard University of MINnesota psychiatry professor John Brantner.
Creator: University of Minnesota. Radio Station KUOM
Contributor: Kuhn, Maggie; Brantner, John; Larson, Curtis (Engineer); Marlow, Andrew (Producer)
Created: 1981-07-21
Contributed By: University of Minnesota Libraries, University Archives.
Last Updated: 2017-11-28

https://umedia.lib.umn.edu/search.json?facets%5Bsuper_collection_name_ss%5D%5B%5D=Minnesota%27s+Radio+History&page=3&q=%2A

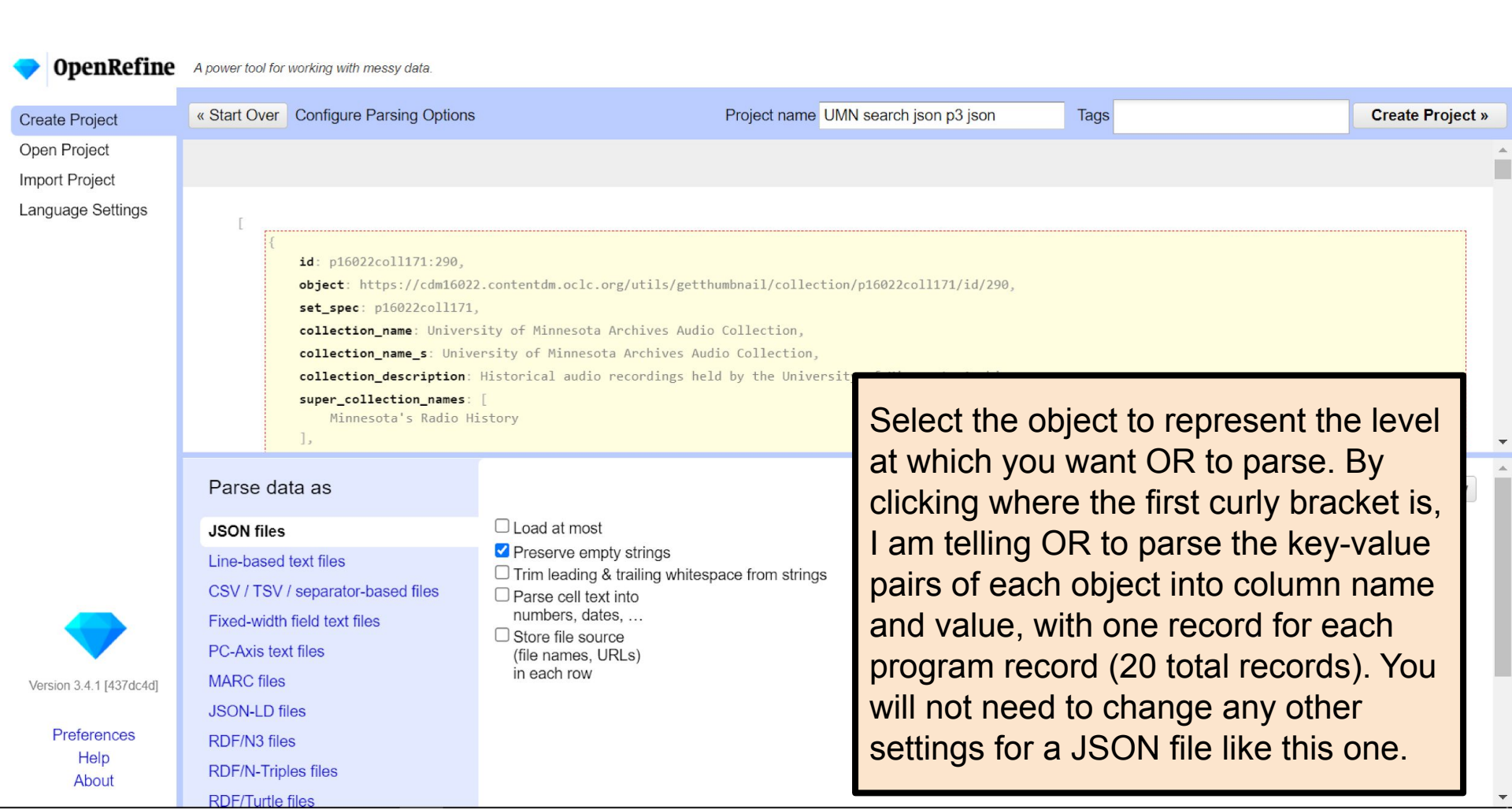
```
[
  {
    "id": "p16022coll171:290",
    "object": "https://cdm16022.contentdm.oclc.org/utils/getthumbnail/collection/p16022coll171/id/290",
    "set_spec": "p16022coll171",
    "collection_name": "University of Minnesota Archives Audio Collection",
    "collection_name_s": "University of Minnesota Archives Audio Collection",
    "collection_description": "Historical audio recordings held by the University of Minnesota Archives.",
    "super_collection_names": [
      "Minnesota's Radio History"
    ],
    "super_collection_name_ss": [
      "Minnesota's Radio History"
    ],
    "super_collection_set_specs": [
      "p16022coll434"
    ],
    "super_collection_descriptions": [
      "This project makes available audio recordings of radio programs produced by the oldest operating radio station in Minnesota, KUOM."
    ],
    "title": "Carter Woodson",
    "title_s": "Carter Woodson",
    "title_t": "Carter Woodson",
    "title_search": "Carter Woodson",
    "title_sort": "carterwoodson",
    "description": "10:50 - 11:00 a.m.",
    "creator": [
      "University of Minnesota. Radio Station KUOM"
    ],
    "creator_ss": [
      "University of Minnesota. Radio Station KUOM"
    ],
    "creator_sort": "universityofminnesotaradiostationkuom",
    "types": [
      "Sound"
    ],
    "format": [
      "Radio programs | http://vocab.getty.edu/aat/300387763"
    ],
    "format_name": [
      "Radio programs"
```

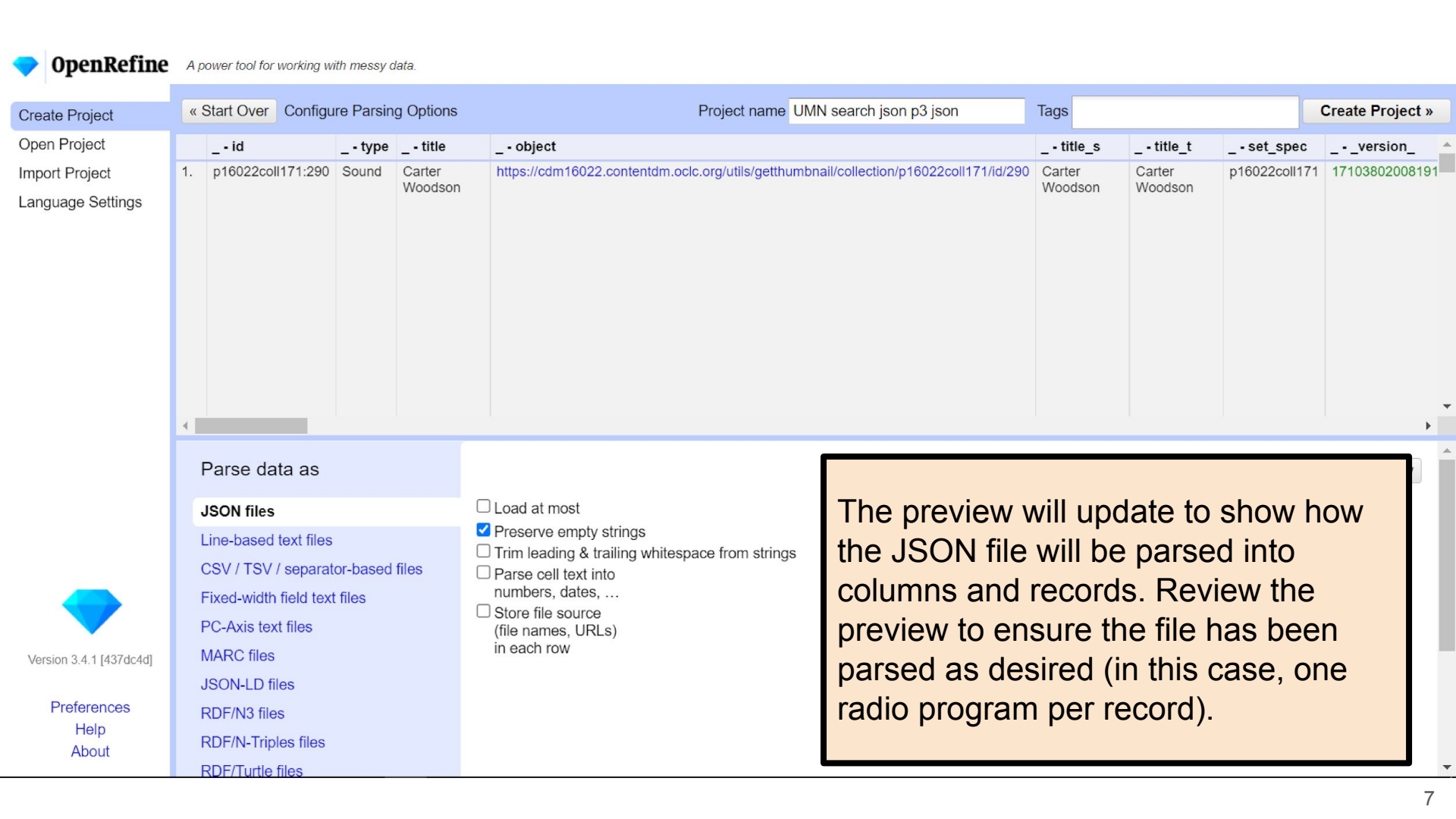

Download the JSON file with 20 individual radio program records.


OpenRefine
A power tool for working with messy data.
Create Project
Open Project
Import Project
Language Settings
Create a project by importing data. What kinds of data files can I import?
TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions.
Get data from
This Computer
Web Addresses (URLs)
Clipboard
Database
Google Data
Locate one or more files on your computer to upload:
Choose Files
UMN search json p3.json
Next »
Load the file into OpenRefine.
Version 3.4.1 [437dc4d]
Preferences
Help
About

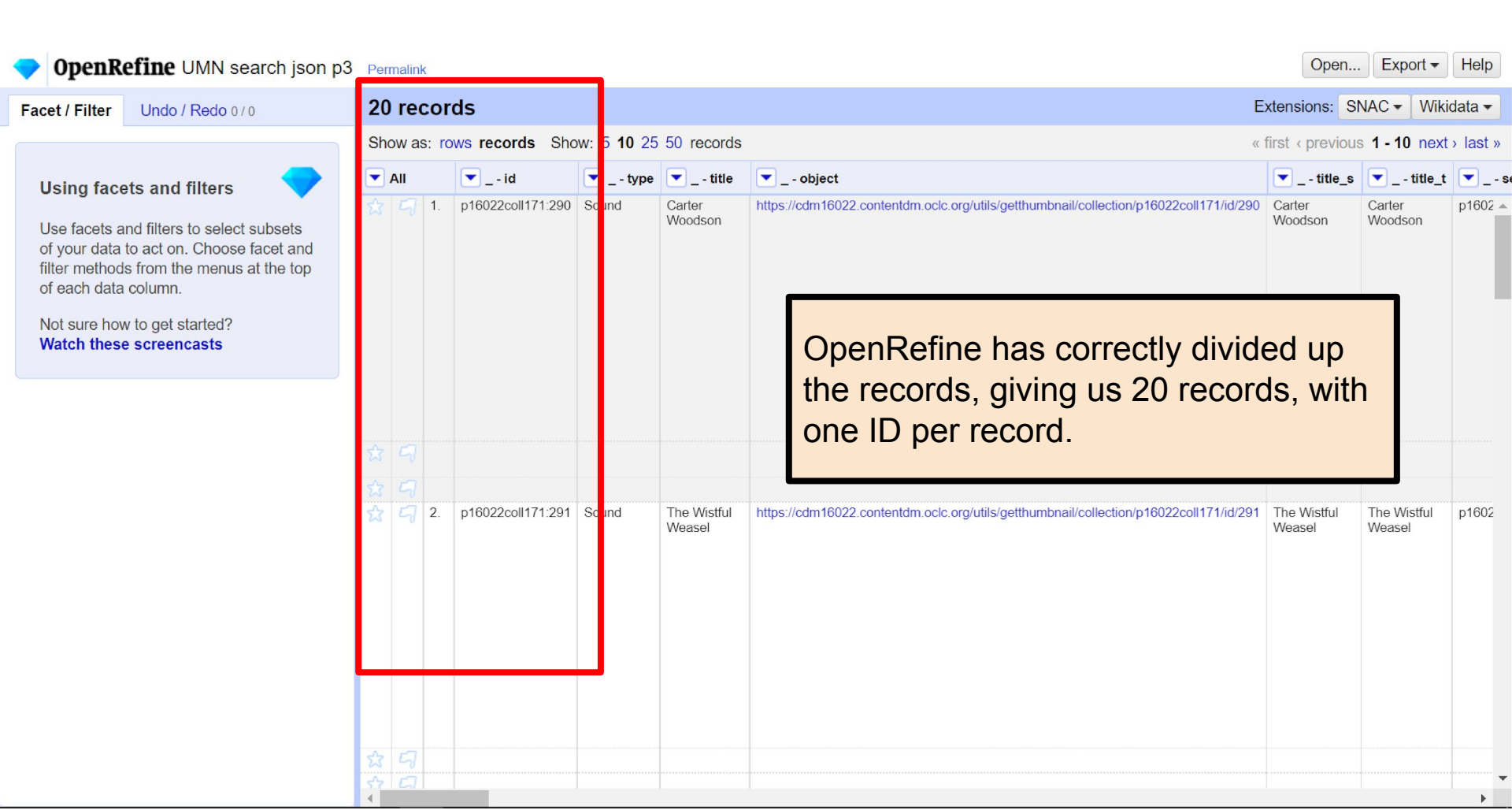


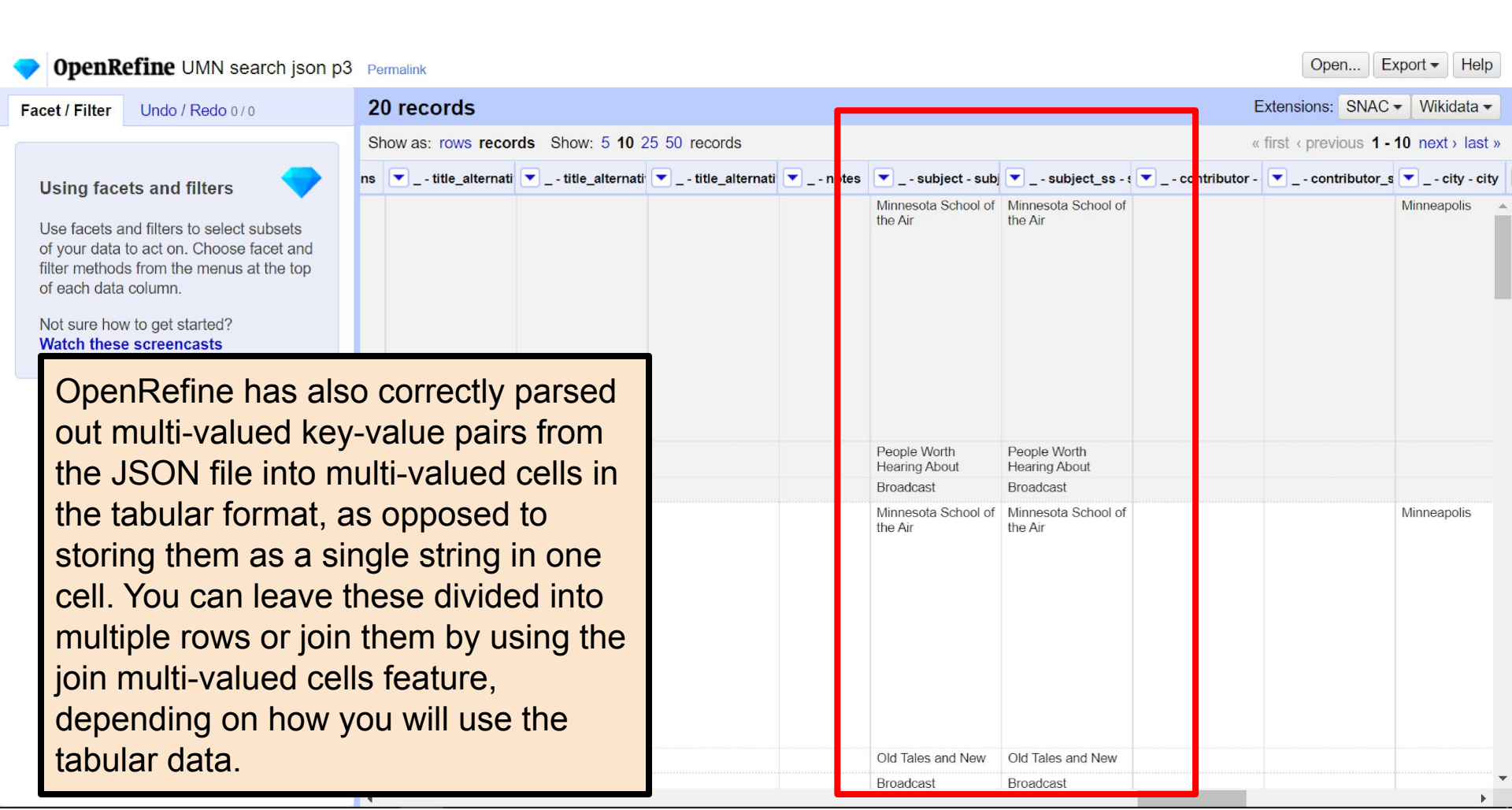
OpenRefine will automatically see that you are attempting to parse a JSON file and show you this screen.


OpenRefine
A power tool for working with messy data.
Create Project
Open Project
Import Project
Language Settings
« Start Over
Configure Parsing Options
Project name
UMN search json p3 json
Tags
Create Project »
[
{
id: p16022coll171:290,
object: https://cdm16022.contentdm.oclc.org/utils/getthumbnail/collection/p16022coll171/id/290,
set_spec: p16022coll171,
collection_name: University of Minnesota Archives Audio Collection,
collection_name_s: University of Minnesota Archives Audio Collection,
collection_description: Historical audio recordings held by the Universit
super_collection_names: [
Minnesota's Radio History
],
Parse data as
JSON files
Line-based text files
CSV / TSV / separator-based files
Fixed-width field text files
PC-Axis text files
MARC files
JSON-LD files
RDF/N3 files
RDF/N-Triples files
RDF/Turtle files
Load at most
Preserve empty strings
Trim leading & trailing whitespace from strings
Parse cell text into numbers, dates, …
Store file source (file names, URLs) in each row
Version 3.4.1 [437dc4d]
Preferences
Help
About
Select the object to represent the level at which you want OR to parse. By clicking where the first curly bracket is, I am telling OR to parse the key-value pairs of each object into column name and value, with one record for each program record (20 total records). You will not need to change any other settings for a JSON file like this one.



The preview will update to show how the JSON file will be parsed into columns and records. Review the preview to ensure the file has been parsed as desired (in this case, one radio program per record).


OpenRefine UMN search json p3
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo 0 / 0
20 records
Extensions: SNAC Wikidata
Show as: rows records Show: 5 10 25 50 records
« first ‹ previous 1 - 10 next › last »
All
_ - id
_ - type
_ - title
_ - object
_ - title_s
_ - title_t
Using facets and filters
Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.
Not sure how to get started?
Watch these screencasts
1. p16022coll171:290 Sound Carter Woodson https://cdm16022.contentdm.oclc.org/utils/getthumbnail/collection/p16022coll171/id/290 Carter Woodson Carter Woodson p1602
2. p16022coll171:291 Sound The Wistful Weasel https://cdm16022.contentdm.oclc.org/utils/getthumbnail/collection/p16022coll171/id/291 The Wistful Weasel The Wistful Weasel p1602
OpenRefine has correctly divided up the records, giving us 20 records, with one ID per record.


OpenRefine UMN search json p3
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo 0 / 0
20 records
Extensions:
SNAC
Wikidata
Show as: rows records
Show: 5 10 25 50 records
« first ‹ previous 1 - 10 next › last »
Using facets and filters
Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.
Not sure how to get started?
Watch these screencasts
_ - subject - subj
_ - subject_ss -
_ - contributor_s
_ - city - city
Minnesota School of the Air
Minneapolis
People Worth Hearing About
Broadcast
Old Tales and New
OpenRefine has also correctly parsed out multi-valued key-value pairs from the JSON file into multi-valued cells in the tabular format, as opposed to storing them as a single string in one cell. You can leave these divided into multiple rows or join them by using the join multi-valued cells feature, depending on how you will use the tabular data.