

Linked Data at UMD Workshop Formulas and Notes

DATA:

https://drive.google.com/drive/folders/13y2PwUNSR1PWFa_PJkyLLE1HTI0q2dPy?usp=sharing

For reference:

[CPF Authorities data in Airtable](#)

[Geographic Authorities data in Airtable](#)

[Subjects data in Airtable](#)

To clean up PEOPLE names:

Copy and paste the formula below into the GREL box in OpenRefine

```
value.replace(/\s\(.*\)/,"").replace(/, [0-9]{4}-([0-9]{4})*/,"").match(/(.*) (.*)/).reverse().join(" ").trim()
```

What this does:

- removes anything in parentheses in the name
- removes years that follow names
- reverses names to "natural" order (Firstname Lastname)
- trims off any leading and trailing spaces

To clean up ORGANIZATION/CORPORATE BODY names:

Copy and paste the formula below into the GREL box in OpenRefine

```
value.replace(/\s\(.*\)/,"")
```

What this does:

- removes anything in parentheses in the name

Reconciliation formulas and property IDs:

cell.recon.match.id = GREL formula to get Wikidata ID

case-sensitive!

General

Len = Wikidata label

Den = Wikidata short description

Aen = Wikidata alternate names

Location

P625 = Coordinate location

P2046 = Area occupied (size)

P281 = Postal code

P30 = Continent

P17 = Country

P131 = Located in the administrative territorial entity (location)

P19 = Place of birth

P20 = Place of death

Biographical

P106 = Occupation

P101 = Field of work

P108 = Employer

Dates

P569 = Date of birth

P570 = Date of death

P571 = Inception date (organizations)

Authority IDs

P244 = Library of Congress authority ID (geographic, people/orgs, subjects)

P214 = VIAF ID

P3430 = SNAC ARK ID

P1225 = NARA ID