

Reusing Open Data from various sources to assess gender inequality: from street and school eponyms, to researchers and research

Miquel Duran and Sílvia Simon

University of Girona
Amical Wikimedia

@quelet @silviasimonr

Open Data are still scarce, but anyway opportunities arise. In Catalonia, we are able to obtain up-to-date data on:

- Street names for municipalities
- Schools (elementary and secondary)
- Researcher names, research groups, research projectes, research units (e.g., universities and research centers)

About streets: In general, only heritage-noteworthy streets are in Wikipedia and Wikidata.

There is no clear definition of street... even in OpenStreetMap a street uses to be a connected set of segments, not a single entity.

Errors arise now and then for names, articles, connectors, ...

What did we do from a list of streets for each municipality?

- single spreadsheet file, order by name, count: single list of streets in Catalonia with its count number, e.g.
- We focused in streets bearing an eponym, like "Martin Luther Jr. street" (indeed, this may containing slight variations and must be clean up)
- We entered eponym (if available) in Wikidata, for streets existing at least in two municipalities.

Most streets are not noteworthy to deserve a Wikipedia page. **Or are they?**

About **schools**:

Besides Open Data as a current list of schools having various information (location, phone, name, reference code, etc.), we have used other open data listings such as:

- Number of teachers, male and female
- Number of students
- Number of groups at each education level
- Number of students entering university soon
- For private schools: number of students financed by public money

Indeed, we did our best to enter eponyms.

Here a relevant question:

Does a person having an eponym in a street or a school deserve a Wikipedia page? Is he/her notable enough?

We think that the answer should be **YES**, especially for Wikipedias in minority languages like Catalan.

About **researcher names**:

From an Open Data listing, we are currently facing a strong difficulty: many researchers have already been entered in Wikidata, but there are redundancies (entered three times, e.g.) errors, discrepancies, ...

There is some important work to be done with researcher names in Wikidata.

Big question: **who is notable enough to be acceptable in Wikipedia?** (question asked by colleagues)

The authors are actually chemists (Institute of Computational Chemistry, University of Girona).

Relevant projects:

- **Wikidata:WikiProject_Chemistry**
- **Scholia**

Scholia needs clean, curated data for researchers... Wikiproject Chemistry needs clear, curated data for chemical compounds, of course.

See please "*Chemistry in Wikidata visualized with Scholia*", ACS Meeting San Francisco, 2023-08-13, Egon Willighagen

Wikidata + OpenStreetMap

- Wikidata is a Knowledge Graph that answers "Where is this?"
- OSM is a geographical database that answers "What is here?"
- They are interwinkled: osm.wikidata.link tool

WD and OSM

- WD: instance of (P31) street or educational center
 - Eponym: named after someone (P138)
- OSM tags: amenity=school / highway=*
 - Eponym: name:etymology:wikidata=Qxyz

Importing in WD, mapping in OSM

- Open Data: regional governments
- Availability in Catalonia:
 - complete list of streets
 - complete list of schools

But...

- WD has almost no streets
- WD may have schools (Catalonia, València: complete)
- OSM: all streets, but incomplete collection of schools

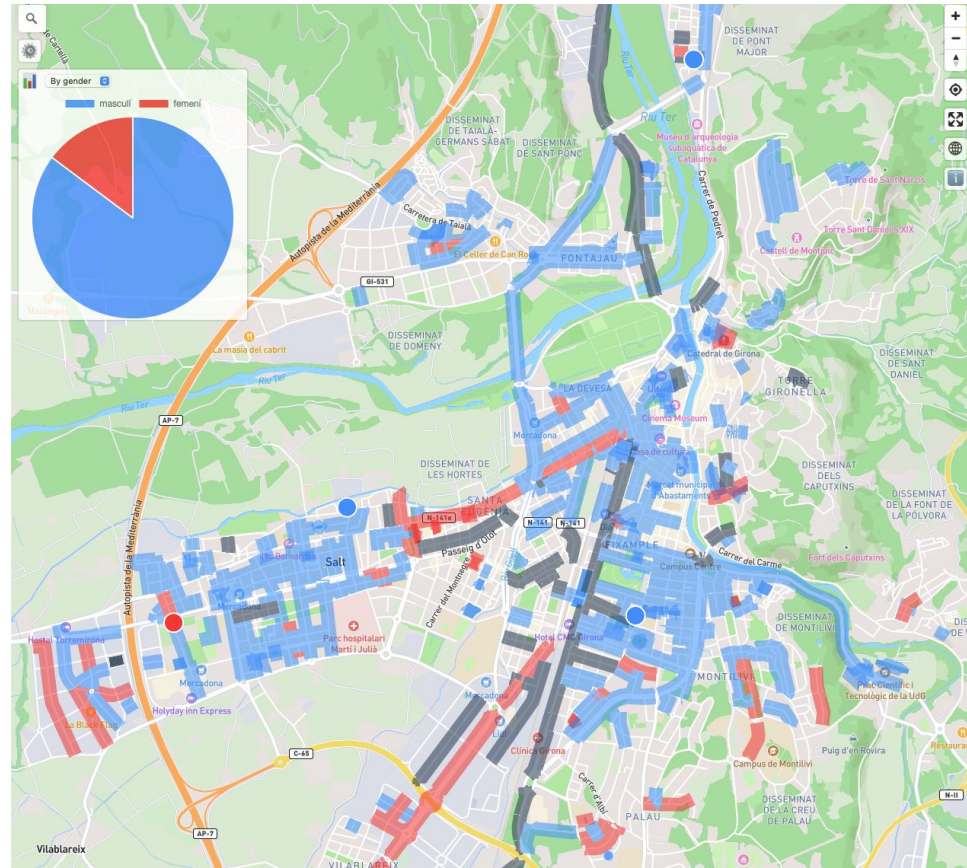
How are items and eponyms entered in WD and OSM?

- Streets:
 - WD: not applicable yet; OSM: already there
 - Eponyms: WD: n/a; OSM: via Python program and OSM APIs.
- Schools:
 - WD: Openrefine; OSM: manually (either nodes or ways)
 - Eponyms: WD: OpenRefine; OSM: via Python program and OSM APIs



A nice tool there: Open Etymology Map

- Displays streets having etymology, assigning color to type of eponym: name, place, other.
- <https://etymology.dsantini.it>



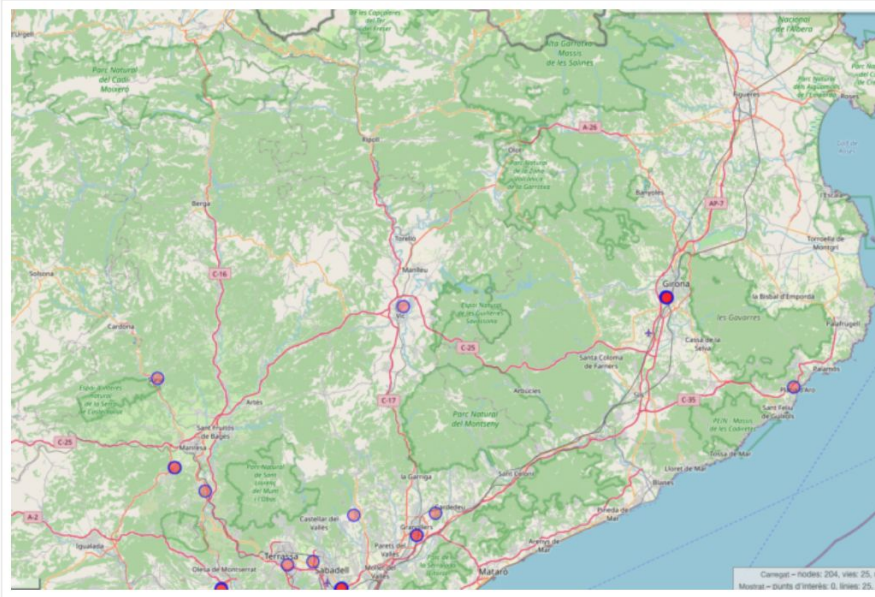


1600 Male names – only 70
STEM

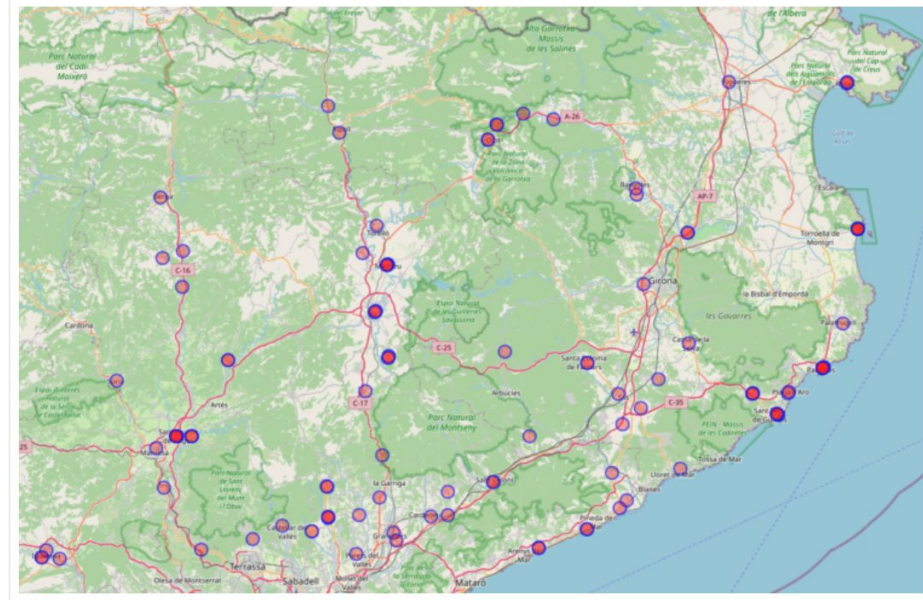
150 Female names – few STEM

- Marie Skłodowska Curie **28**
- Martina Castells **5**

Scientists and street names in Catalonia



Carrers de Marie Curie



Carrers d'Alexander Fleming



Schools

699 (cat) + 340 (PV) + 68 (IB)

265 + 146 + 28 Person

name

61 Female name

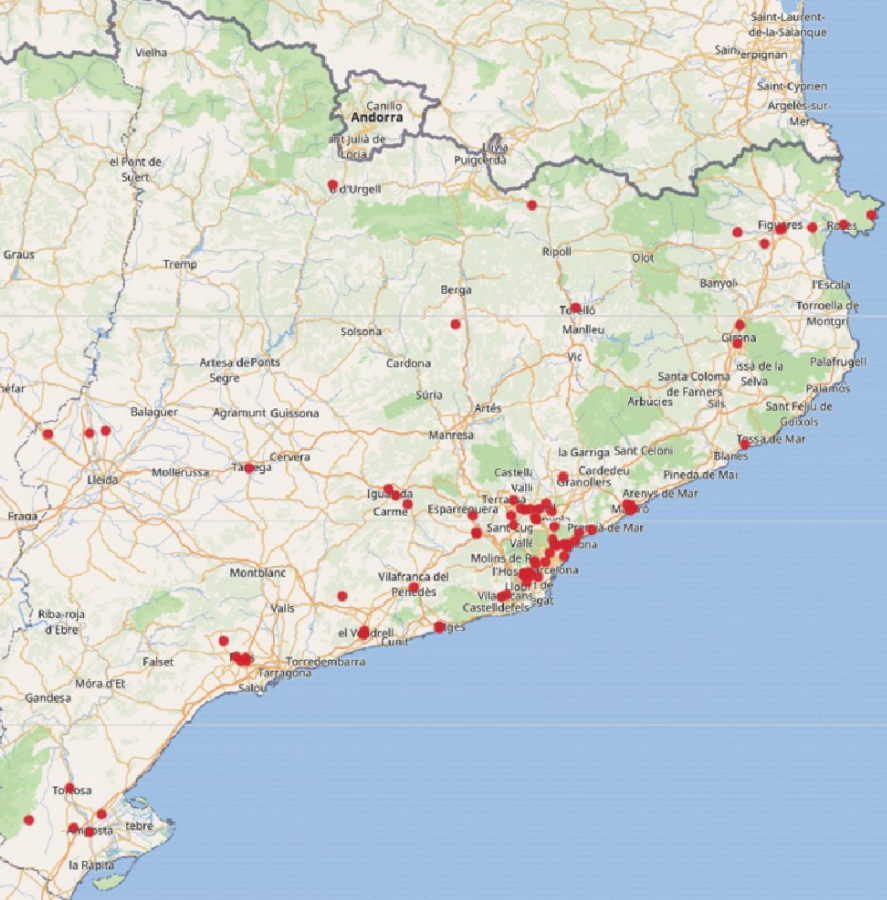
15 NO Wikipedia data

#WikiSciW



#WikiSciW@c4dUdG

**WIKIMANIA
SINGAPORE**



Schools

1634 Primary Schools

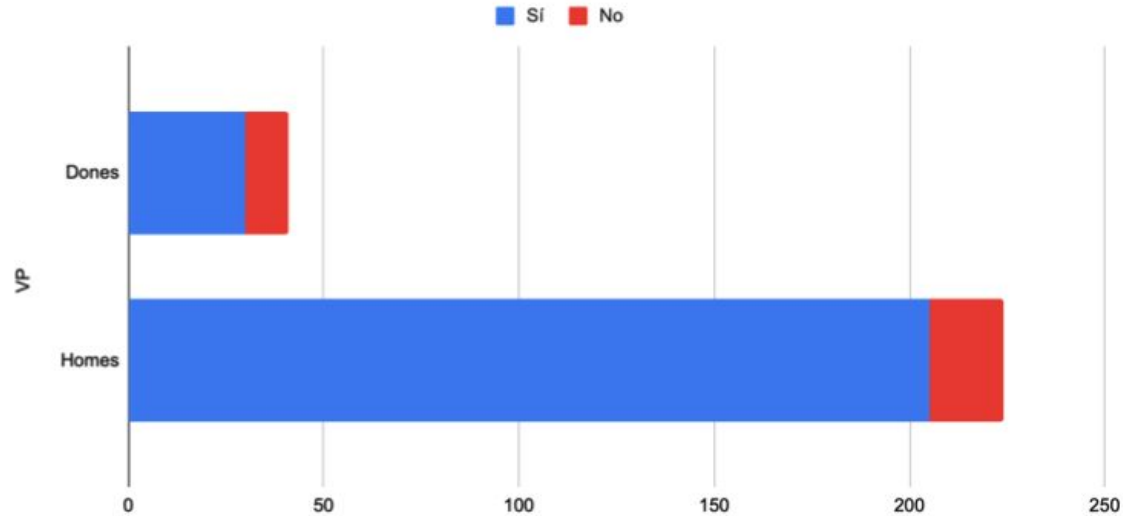
587 Person Name (36%)

80 Female name (14%)

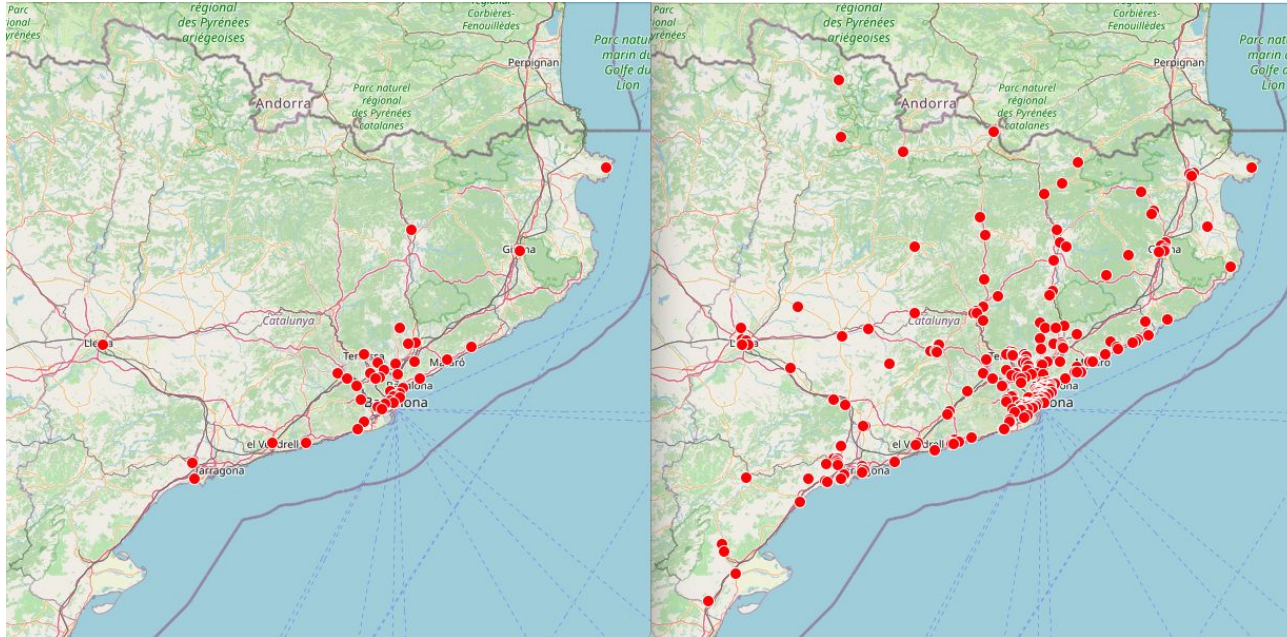
30 No Wikipedia data(38%)

High school eponyms and Wikipedia

Dones i Homes als noms dels Instituts de Catalunya: Presència a la Viquipèdia



High schools with female eponym in Catalonia



Female eponym fact

- Only 14% of high school eponyms in Catalan-speaking areas are female

	Total	Total amb epònim	% epònims	Epònim masc	Epònim fem	% ep femen	Manquen fem a vp	%manquem fem vp
Instituts Cat	699	265	38%	224	41	15%	11	27%
IES PV	340	146	43%	128	18	12%	4	22%
IES IB	68	28	41%	26	2	7%	0	0%
Esc Cat								
Prov Girona	232	81	35%	70	11	14%	3	27%
Prov Lleida	219	62	28%	57	5	8%	4	80%
Prov TGN	244	84	34%	61	13	15%	6	46%
Prov BCN	939	360	38%	309	51	14%	17	33%
Total Cat	1634	587	36%	507	80	14%	30	38%

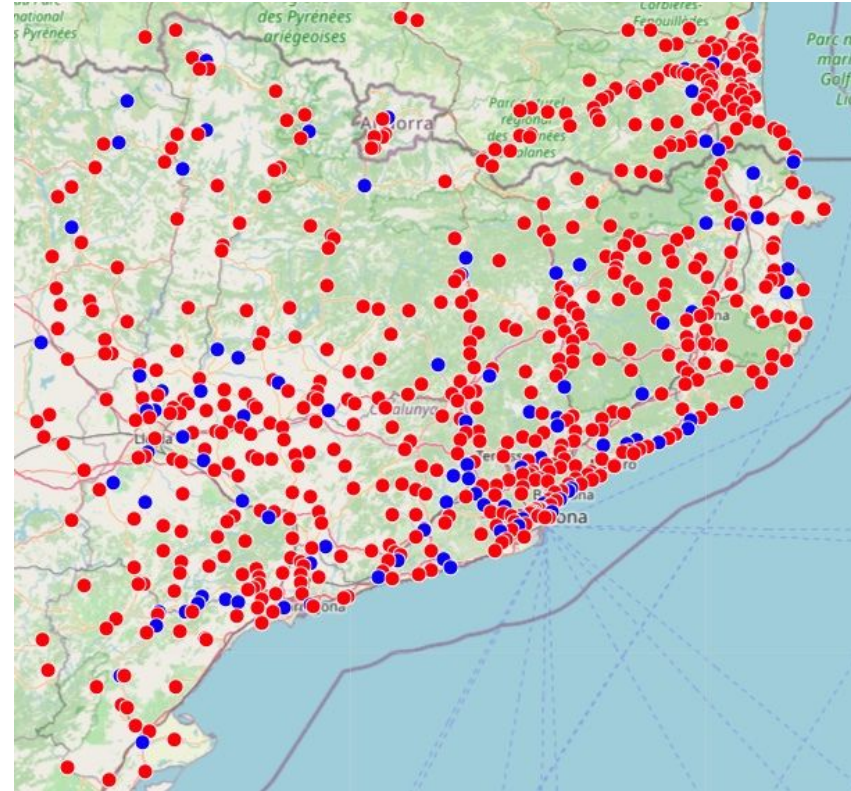
Academic study in Nature Group journal: Notability

Laouenan et al, A cross-verified database of notable people, 3500BC-2018AD, Scientific Data 9, 290 (2022)

Viewer:

<https://tjukanovt.github.io/notable-people>

Figure: Places where man is the most notable person (red) vs woman (blue)



More on notability

- Only 197 places in Catalonia has notable scientists (177 men, 45 women)
- Scientists have a hard time to be notable
- Women have a much harder time to be notable
- 71 men and 12 women are the MOST notable in Catalan places

Contributions

- <https://c4dudg.wordpress.com/category/wikisciw/>
- Social networks: #WikiSciW
- Communications to meetings
- Round tables, Wikimedia editathons, etc.

What's next?

- Higher education (this is really tough)
- Archives, Libraries, Museums, civic centers, etc.
- Sports, Health facilities
- Gender studies of researcher presence in WD
- Areas
 - In Catalan-speaking area
 - Outside Catalan-speaking area
- See https://ca.wikipedia.org/wiki/Viquiprojecte:Més_dones
- See #WikiSciW at c4d.udg.edu

#WikiSciW

Thank you!



Miquel Duran, Sílvia Simon

WIKIMANIA SINGAPORE