



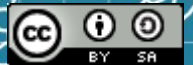
DameGender

Towards an international and free dataset
about name, gender and frequency

(David Arroyo Menéndez)
davidam@gmail.com

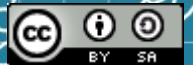


Equality of gender is the fifth
objective of sustainable
development for
United Nations





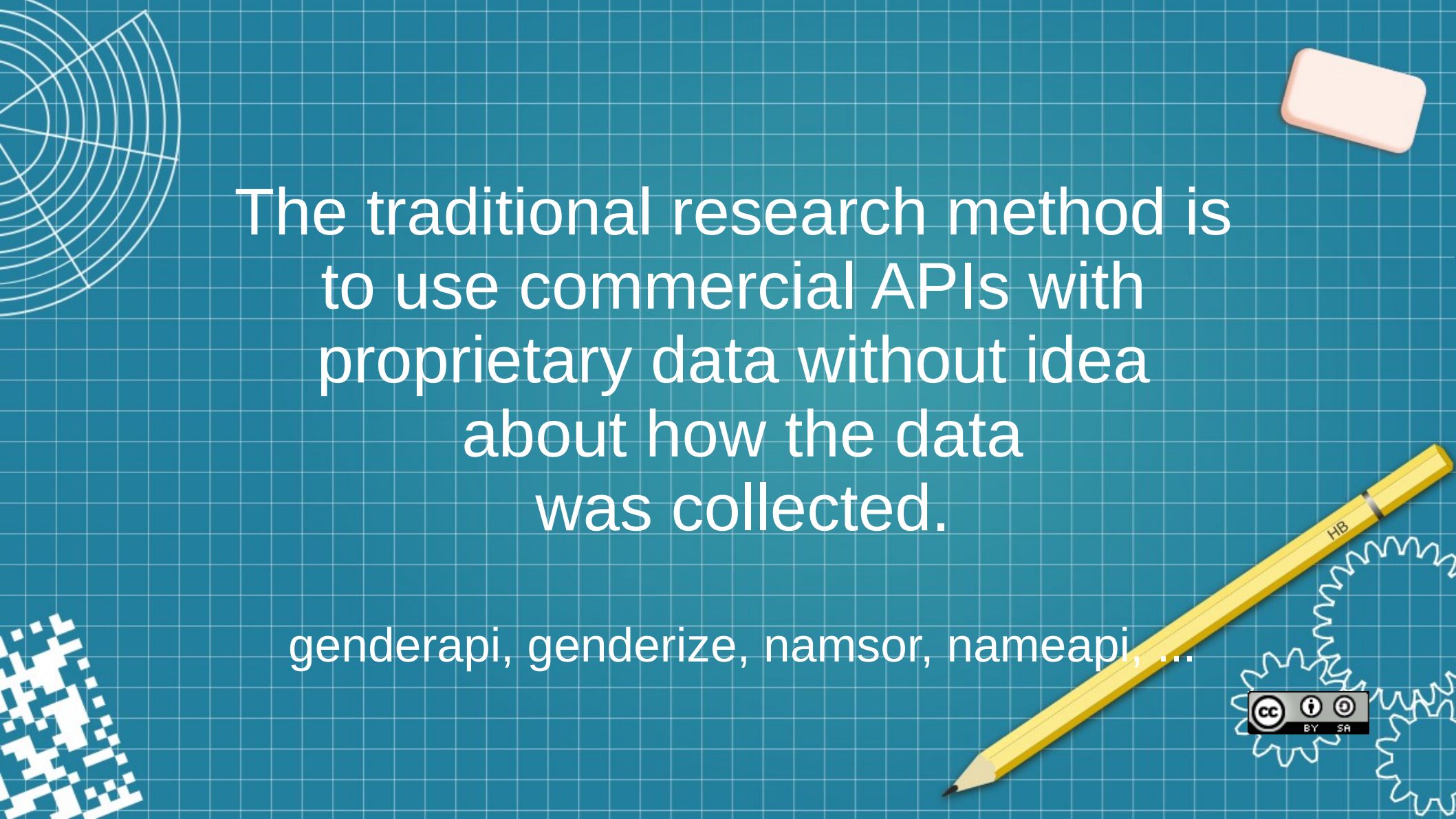
This equality can be reached by
measuring and analyzing data and
making good politics with the
results.





Many gender studies count males and females based on their names, for instance, research papers, job positions, streets, etc.

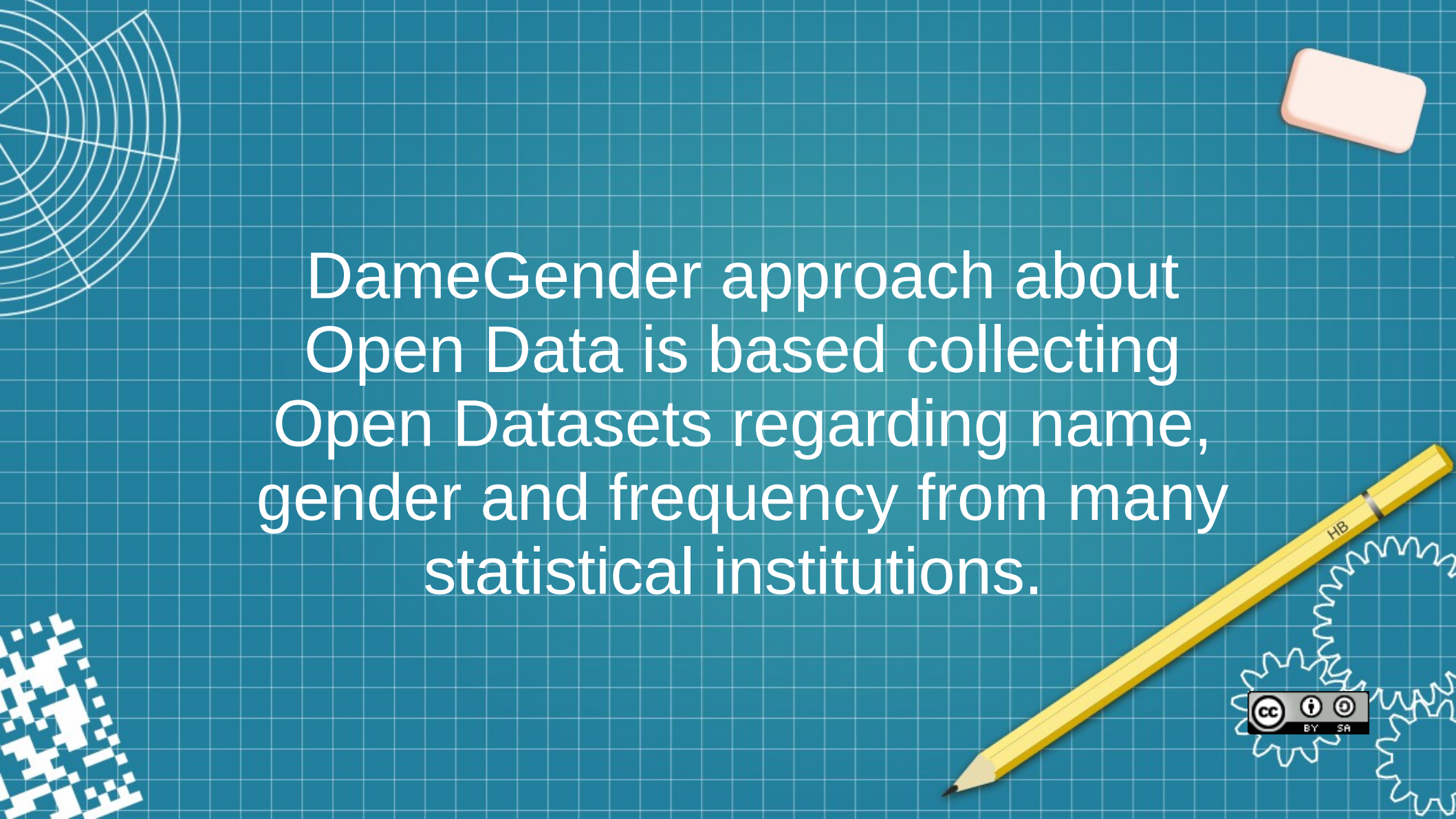




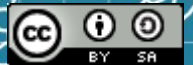
The traditional research method is
to use commercial APIs with
proprietary data without idea
about how the data
was collected.

genderapi, genderize, namsor, nameapi, ...





DameGender approach about
Open Data is based collecting
Open Datasets regarding name,
gender and frequency from many
statistical institutions.



Take a look about our mapamundi!



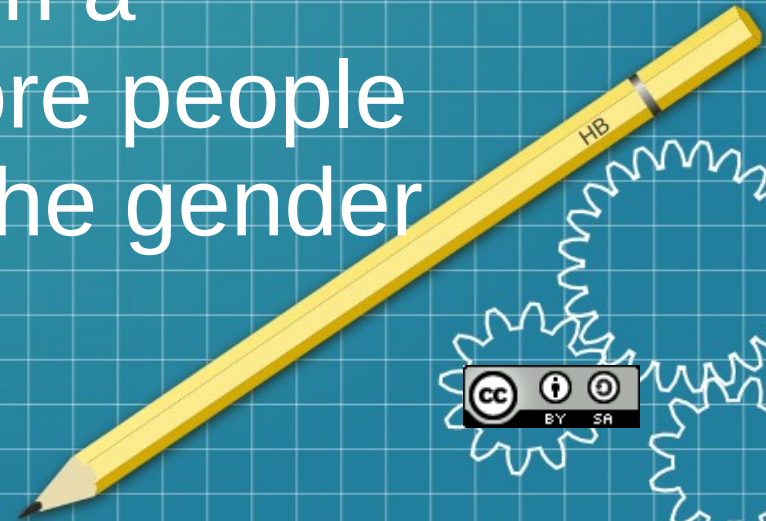
Take a look about our accuracies!

Dataset Test	Damegender Accuracy
Scientists Wikipedia	0.93
FIFA soccer	0.93
WTA tennis	0.91
National Leagues	0.91
Baby Names New York	0.98
Conseil Garonne	0.97
Paul Sebo	0.88
Lucia Santamaria	0.88

Comparing Damegender Free Dataset in number of names

Free Dataset	Number of males	Number of females
USA census (SSA)	91320	91320
Namdict (genderguesser)	48821	48821
NLTK	2943	5001
DameGender	257925	304553

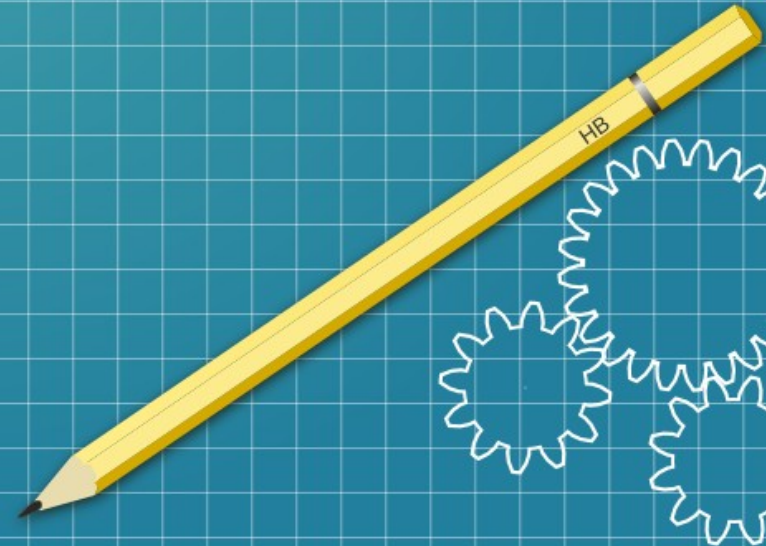
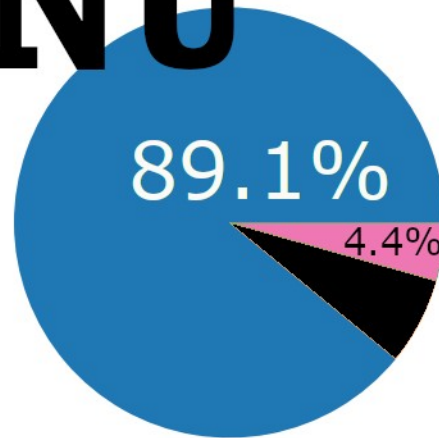
DameGender will allow to
measure gender gap to students
and academics
interested on the phenomenon
without costs and on a
reproducible way and more people
will be contributing to fix the gender
gap.



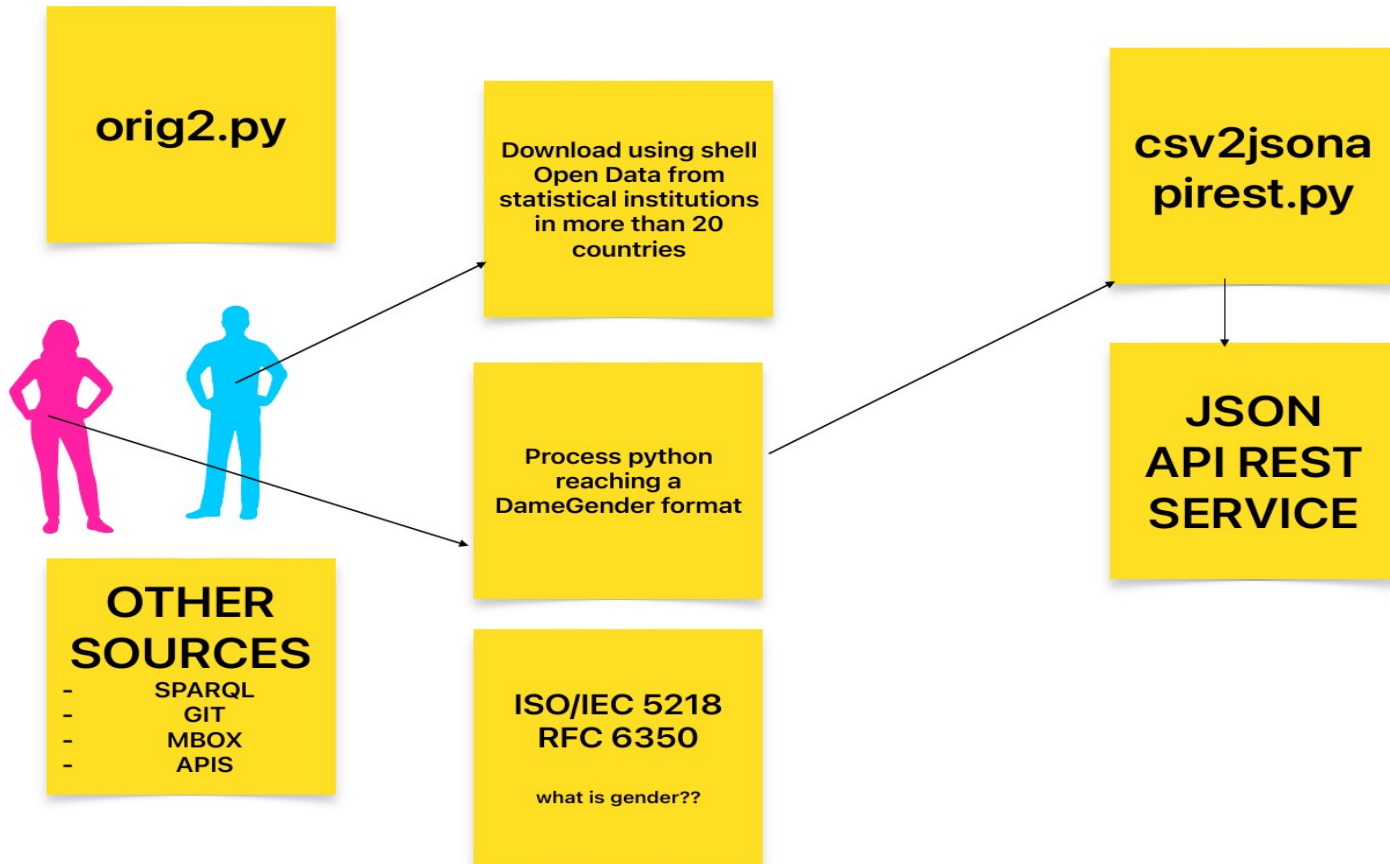
csv2gender: infer names to gender from a csv file

```
$ python3 csv2gender.py files/gnu-maintainers.csv  
  --first_name_position=0 --title="GNU"  
  --dataset="inter" --outcsv="files/gnu.gender.csv"  
  --outimg="files/debian.gender.png"
```

GNU



orig2.py and csv2jsonapi.py



Thanks!!

