SERVEI DE SISTEMES
D'INFORMACIÓ GEOGRÀFICA
I TELEDETECCIÓ
Universitat de Girona

Centro de Estudos de Geografia e Planeamento Regional
Faculdade de Ciências Socias e Humanas . Universidade Nova de Lisboa

Terra Communis (tComm): A free data provider for historical census micro-data

António M. Rodrigues, Bruno Neves, Carla Rebelo

Creative Commons 3.0

VII Jornadas De SIG Libre

What is tComm?

Spatio-temporal socio-economic data for all.

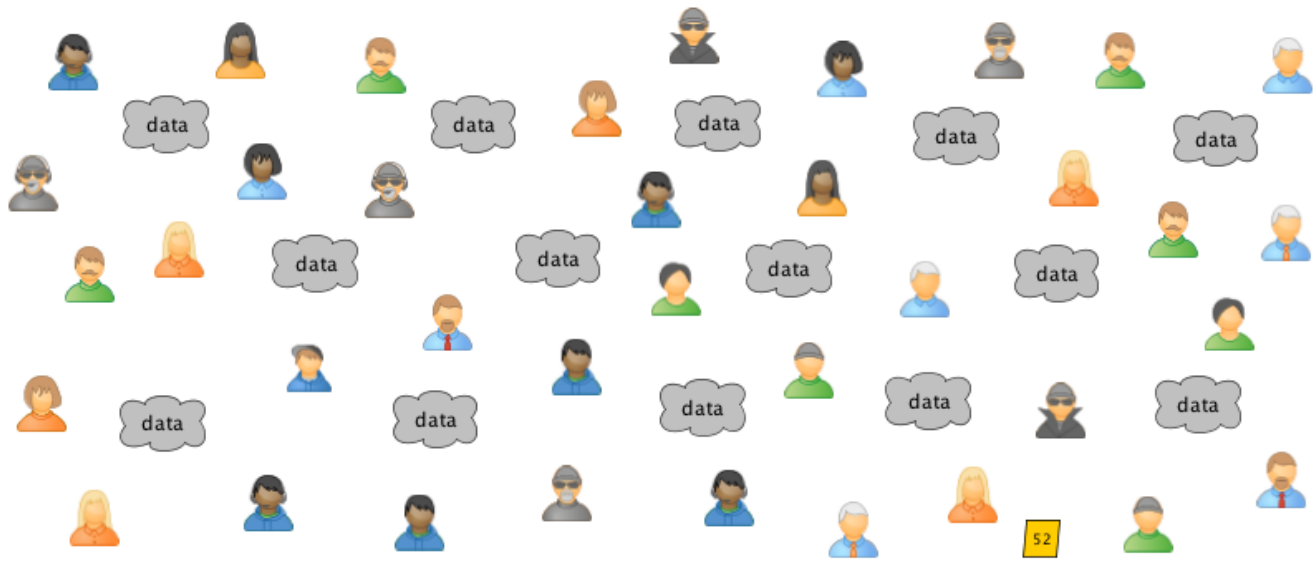
What's innovative in *tComm*?

1. A large scale comprehensive time-space database is built for Portuguese census data
2. the fact that data is made available online contributes to the goal of making knowledge symmetrically available
3. the fact the only Free and Open Source Software (FOSS) is used means that, other than man-hours, the project is costs' free.

Why? (1)

CONCEPTUAL FRAMEWORK: Abstract individual realities

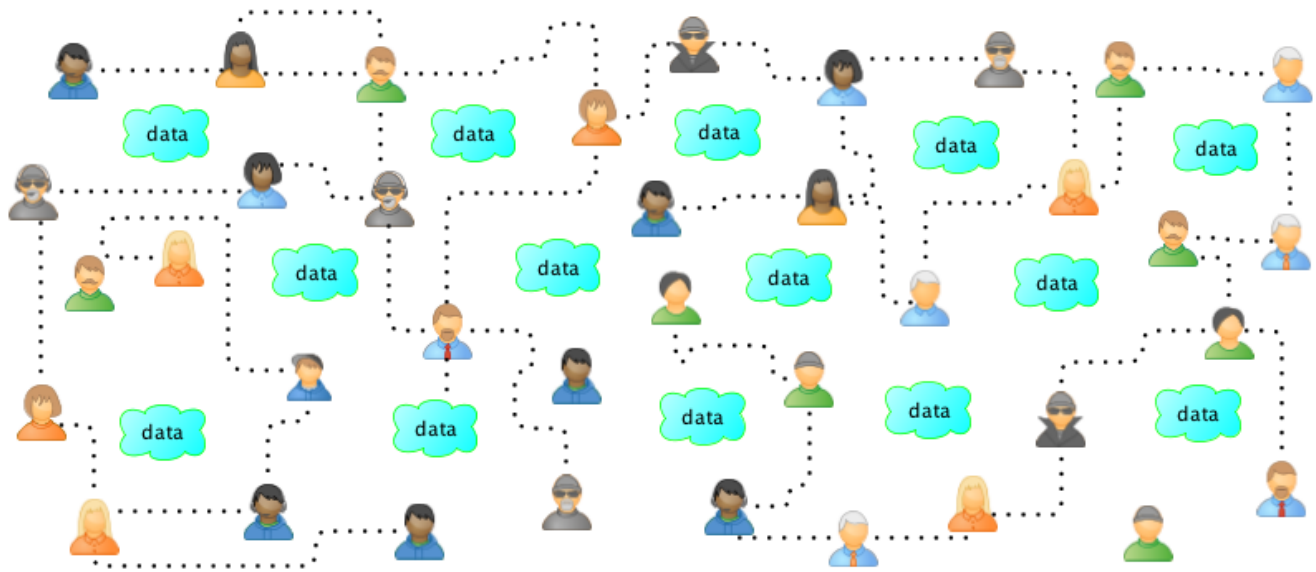
Lot's of (free) information; individual agents; no linkages.



Why? (2)

CONCEPTUAL FRAMEWORK: Abstract individual realities

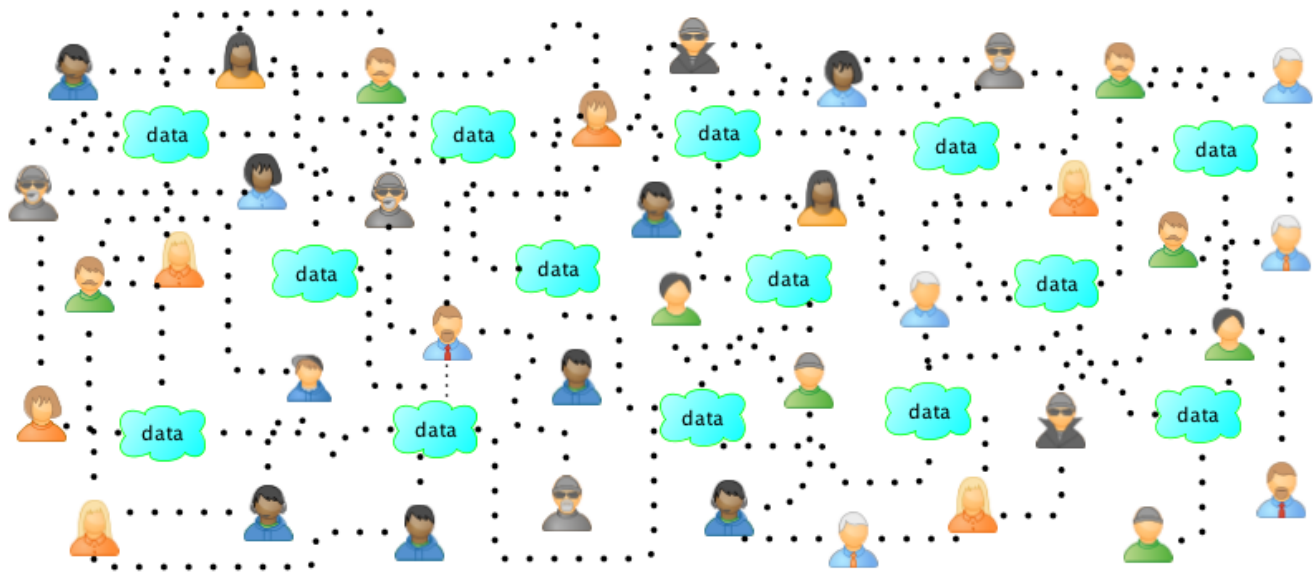
Free, yet unreachable data wharehouses; individual linkages.



Why? (3)

CONCEPTUAL FRAMEWORK: Abstract individual realities

Free Data wharehouses; Rich networks; Multi-thread linkages.



About the Portuguese National Census...

http://www.ine.pt

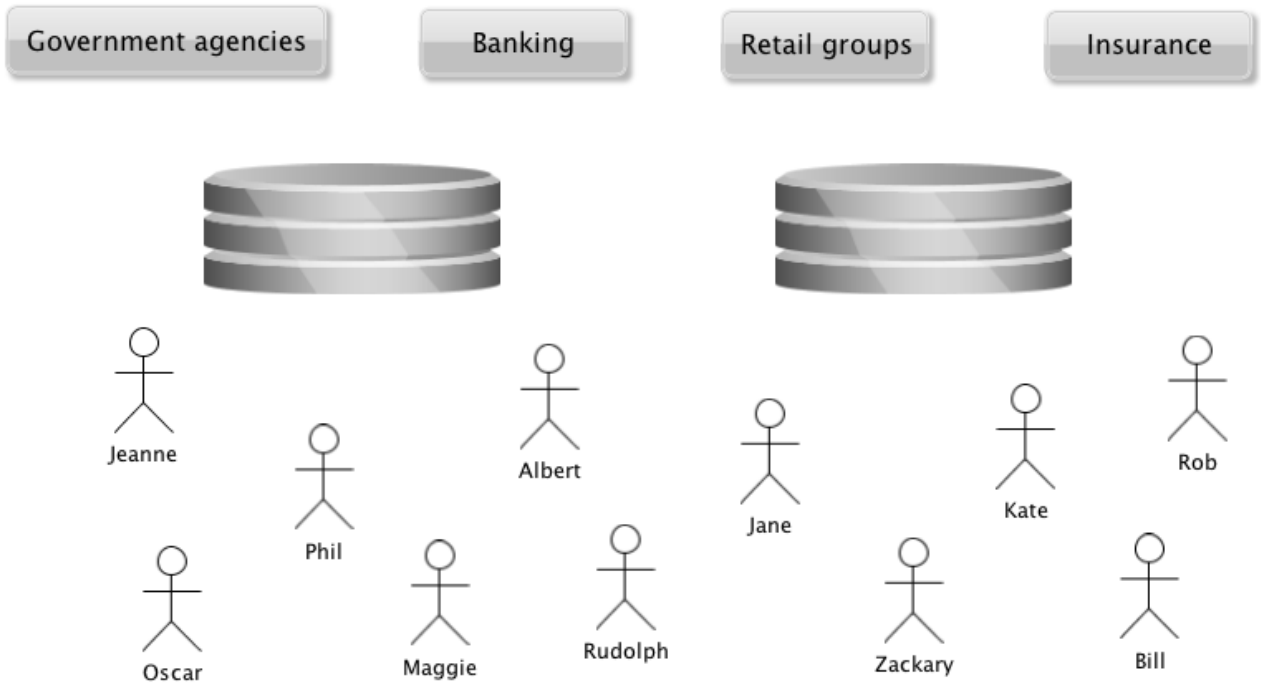
1. O que são os Censos ? Os Recenseamentos da População e da Habitação, (...) são uma contagem da população e do parque habitacional (...) permite saber quantos somos, como somos e como vivemos.
2. Porque é que é necessário que existam Censos ? (...) disponibilizam informação de grande utilidade para a definição de políticas e para a planificação de serviços em áreas como a educação, a saúde ou os transportes.

Portuguese Census data - brief notes

The evolution of the small statistical areas through the Census

Census	Number of Small areas Secções	Number of Small areas Subsecções	100% Portugal	Digital formal	Street centreline	Buildings
1981	Major cities	Major cities	No	No	No	No
1991	13 705	106 626	Yes	Yes	No	No
2001	16 094	177 893	Yes	Yes	Yes – With some street names	Partial – 400 000
2011	18 078	266 000	Yes	Yes	Yes, with street names, door numbers and postal code	Yes – 3 500 000

Free and acessible geo/demographic data



Ideally, it should be possible for a user to easily perform tasks such as:

- **Discover** the existence of data
- **Access** the data for research and analysis
- **Find** detailed information describing the data and its production processes
- **Access** the data sources and collection instruments from which and with which the data was collected, compiled, and aggregated
- **Effectively communicate**with the agencies involved in the production, storage, distribution of the data
- **Share** knowledge with other users

http://www.ama.pt

AGÊNCIA PARA
A MODERNIZAÇÃO
ADMINISTRATIVA
PRESIDÊNCIA DO CONSELHO DE MINISTROS

(...) constitui-se como **principal desafio da AMA** contribuir para a transformação profunda da relação entre a Administração Pública portuguesa e aqueles que a justificam, os Cidadãos e as Empresas, estruturando a sua actuação em torno de um princípio: **disponibilizar a informação correcta, no formato certo, à pessoa certa, no tempo exacto**.

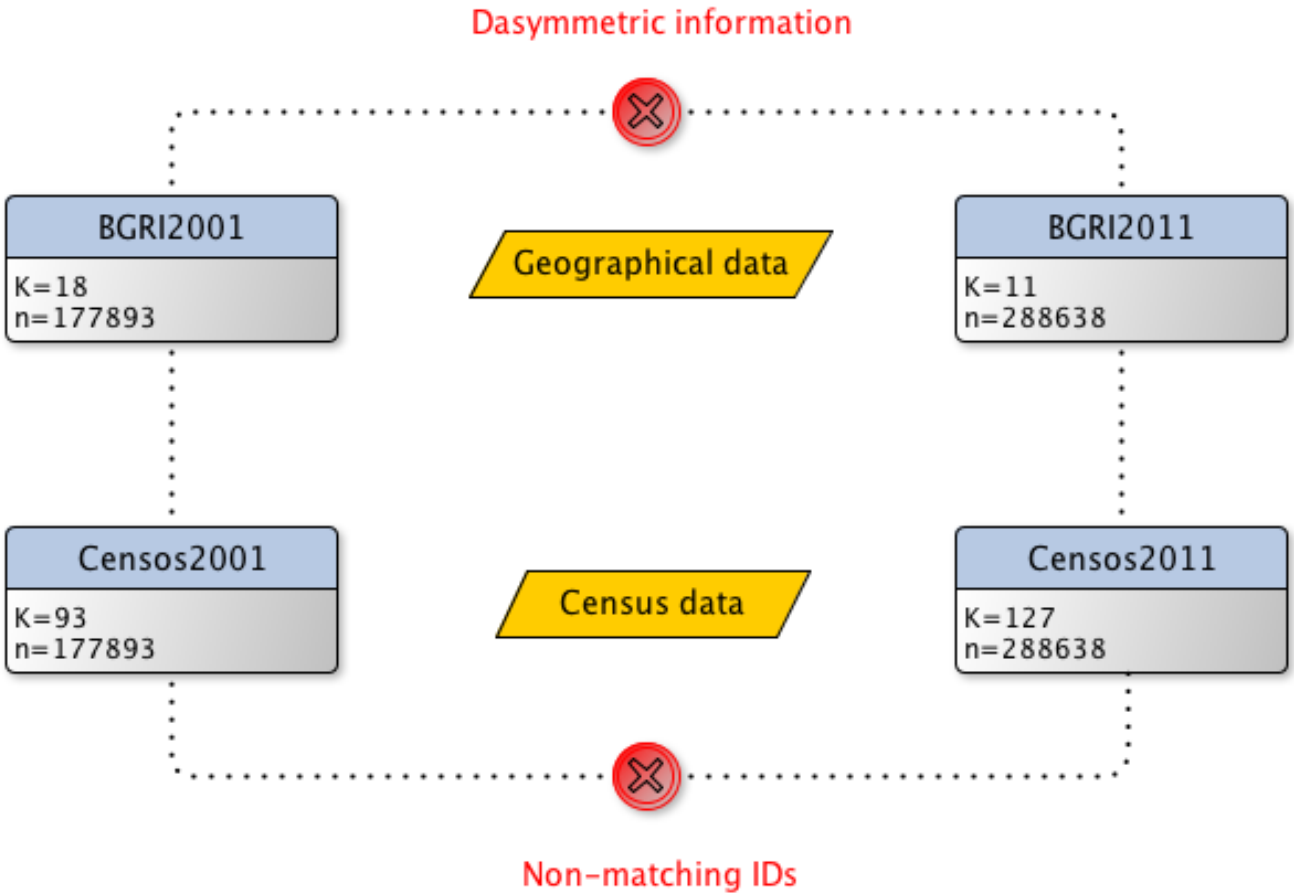
Let's imagine some questions...

Who are the potential customers of my grocery shop?

How many buildings existed in my neighbourhood prior to 1945?

What was the population change in my village between 2001 and 2011?

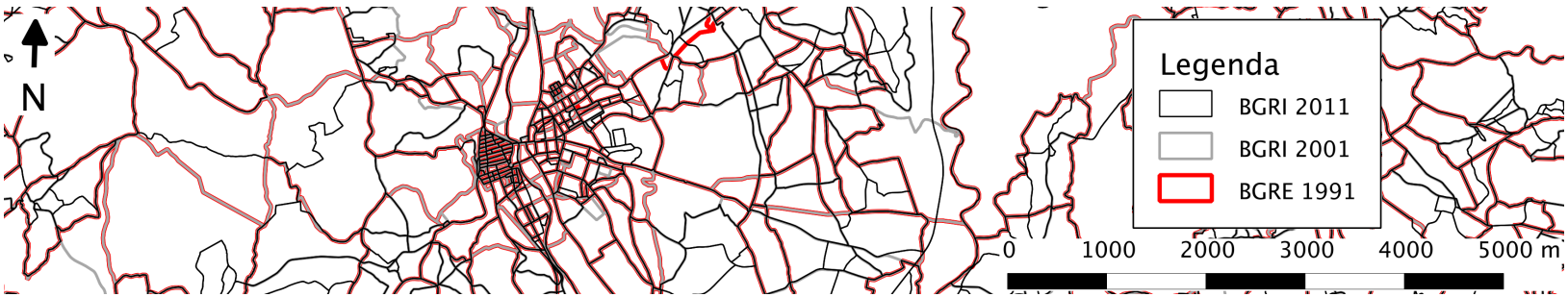
Limitations of census datasets (1)



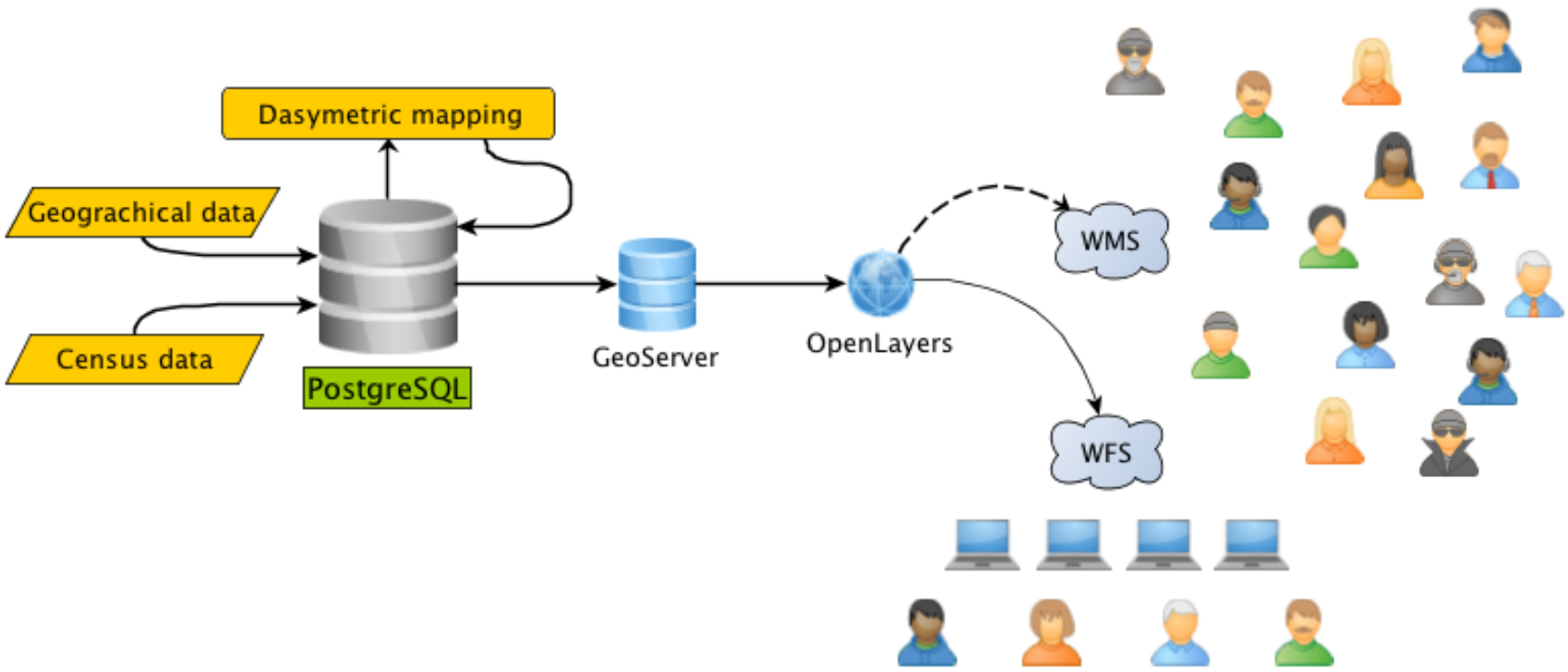
Limitations of census datasets (2)

Aggregation issues

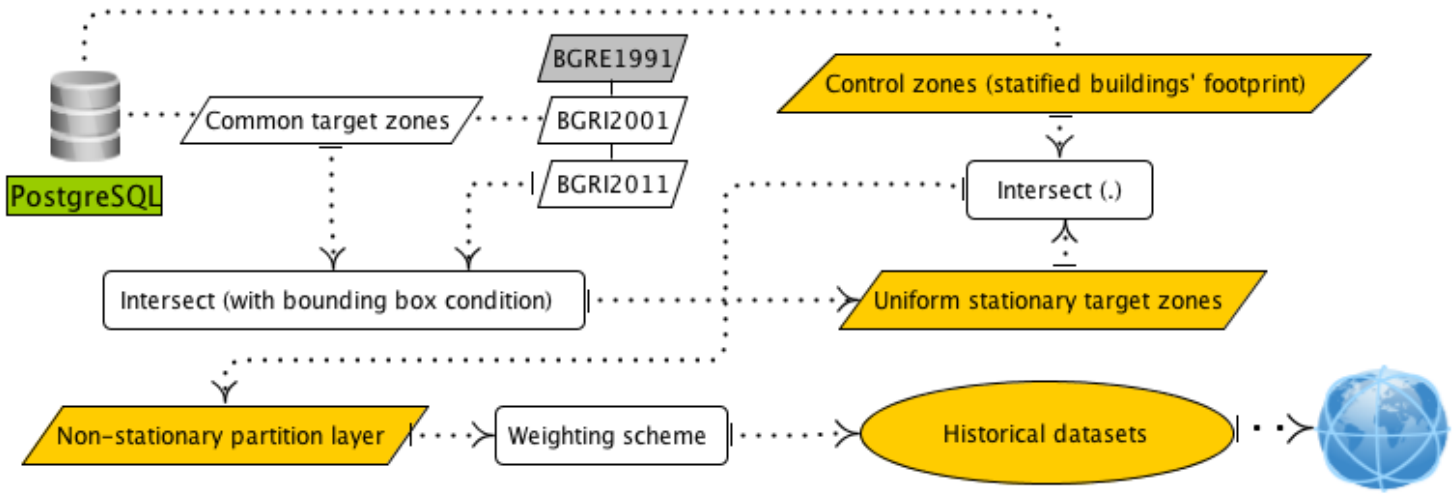
(d)asymetric information



workflow


Dasymetric mapping
Geograchical data
Census data
PostgreSQL
GeoServer
OpenLayers
WMS
WFS

Dasymmetric mapping - workflow



Data processing - big data

$$177893 \times 288638 \tag{1}$$

Efficiency of the bounding box overlap operator (&&) when performing intersections in PostGIS.

Inefficiency of GIST indexes in big tables?

Some code...

```
01 CREATE TABLE tmp.teste AS
02 SELECT
03     tmp.tomar2001.bgri2001,
04     tmp.tomar2011.bgri2011,
05     tmp.tomar2001.p2001,
06     tmp.tomar2011.p2011,
07     tmp.tomar2001.area2001,
08     tmp.tomar2011.area2011,
09     ST_Intersection(tmp.tomar2001.the_geom, tmp.tomar2011.the_geom) as the_geom,
10     st_area(ST_Intersection(tmp.tomar2001.the_geom, tmp.tomar2011.the_geom)) as
   areaClip,
11     st_area(ST_Intersection(tmp.tomar2001.the_geom, tmp.tomar2011.the_geom)) /
12         tmp.tomar2001.area2001 AS weight2001,
13     st_area(ST_Intersection(tmp.tomar2001.the_geom, tmp.tomar2011.the_geom)) /
14         tmp.tomar2011.area2011 AS weight20011,
15     tmp.tomar2001.p2001 * st_area(ST_Intersection(tmp.tomar2001.the_geom,
   tmp.tomar2011.the_geom)) /
16         tmp.tomar2001.area2001 AS p2001prop,
17     tmp.tomar2011.p2011 * st_area(ST_Intersection(tmp.tomar2001.the_geom,
   tmp.tomar2011.the_geom)) /
18         tmp.tomar2011.area2011 AS p2011prop,
19     st_isvalid(ST_Intersection(tmp.tomar2001.the_geom, tmp.tomar2011.the_geom))
20 FROM
21     tmp.tomar2001,
22     tmp.tomar2011
23 WHERE
24     tmp.tomar2001.the_geom && tmp.tomar2011.the_geom;
```

Intersection

Weights

proportions

BB restriction

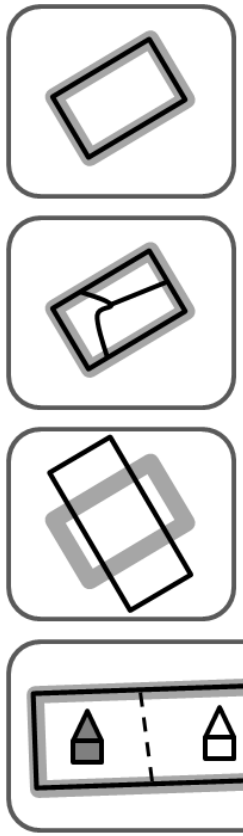
Dasymmetric mapping techniques (1)

Same geometry 2001-2011

Aggregated geometry 2001-2011

Different geometry 2001-2011

Footprint not represented by control zones



2011

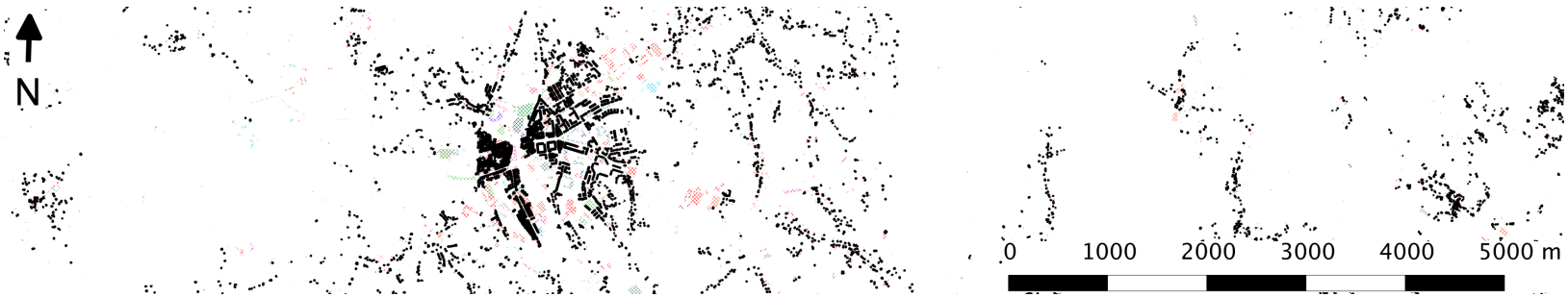
2001

Non-identified building

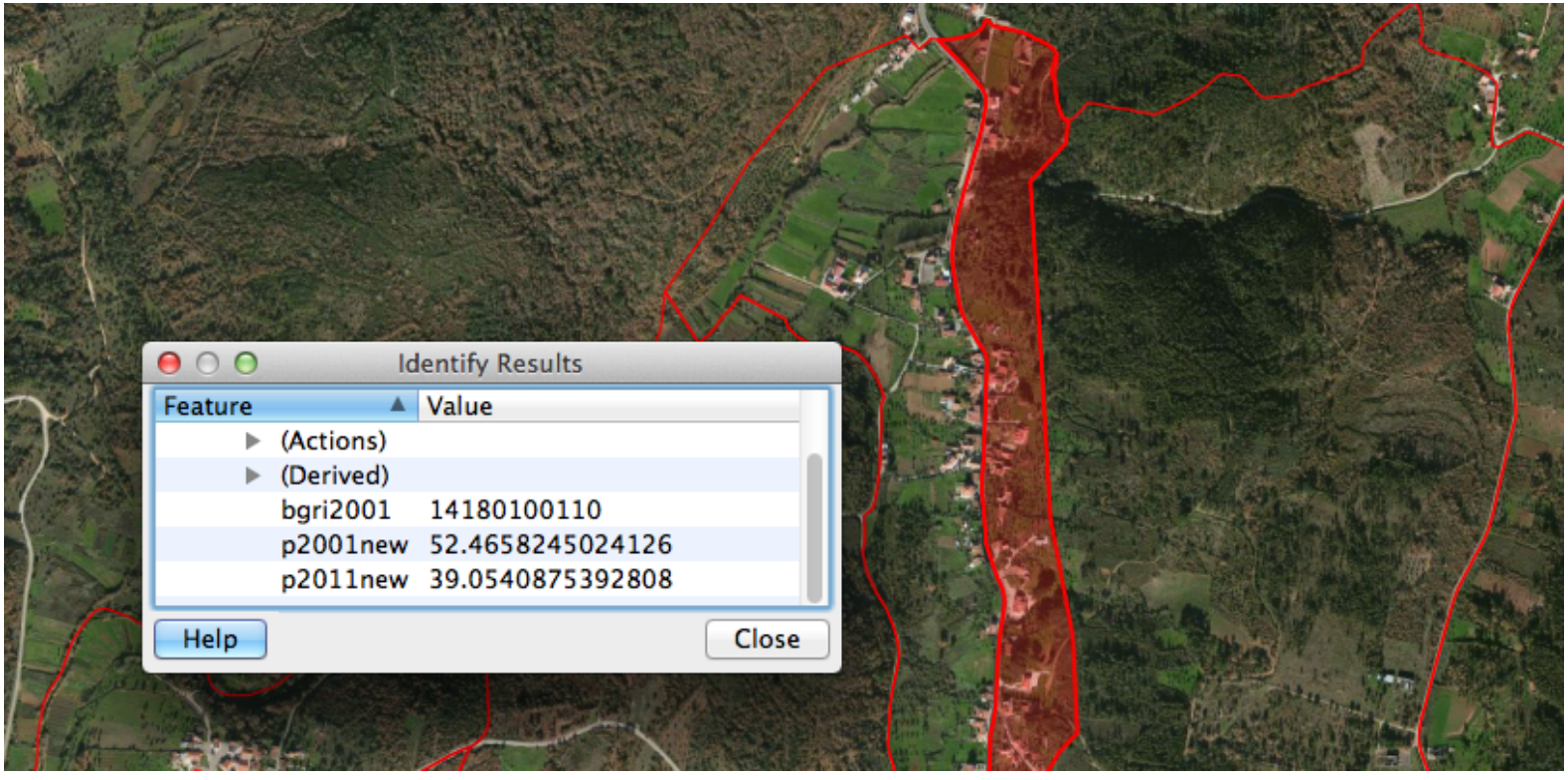
Identified building

Control zone

Dasymmetric mapping techniques (2)

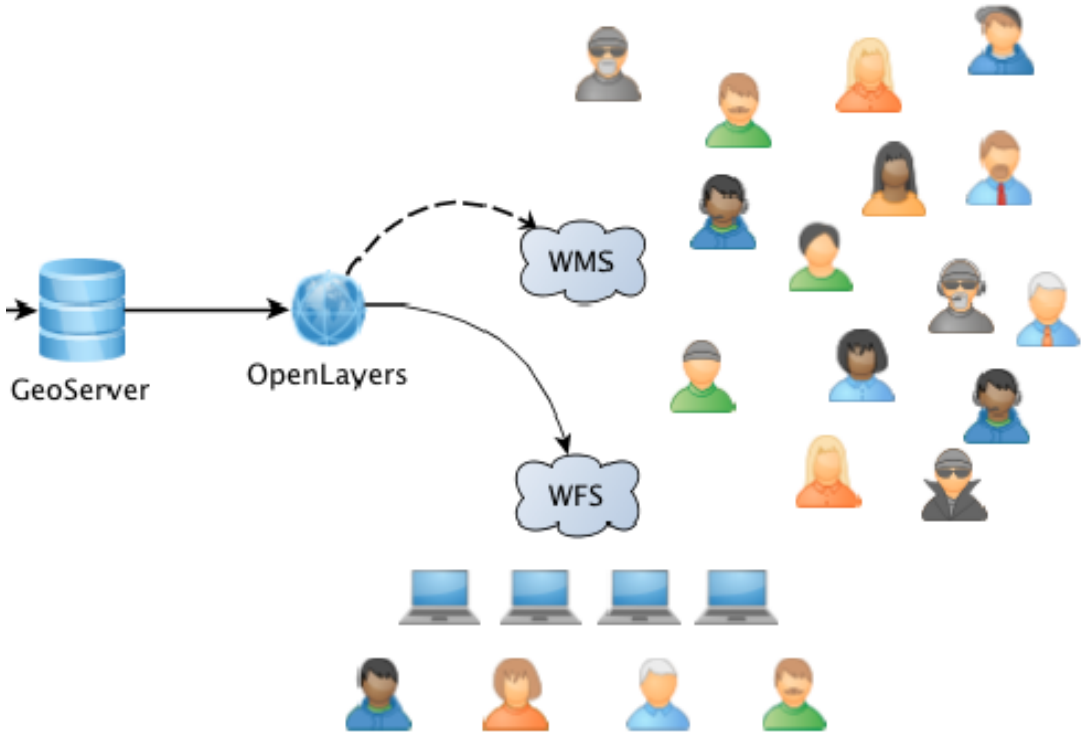


What we can do now... (Ceras)



What we can do now...

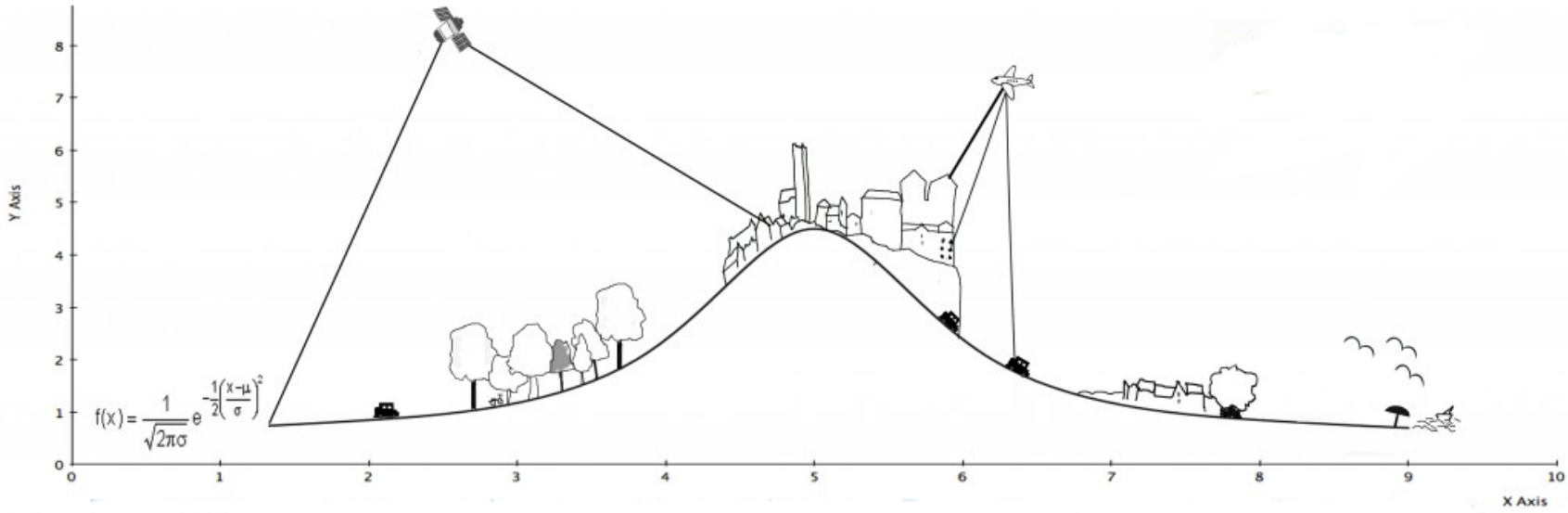
And in a few weeks...



In a couple of months...

PL/R functionality

Follow the next episodes in...



http://mgcotunl.wordpress.com

Gràcies!