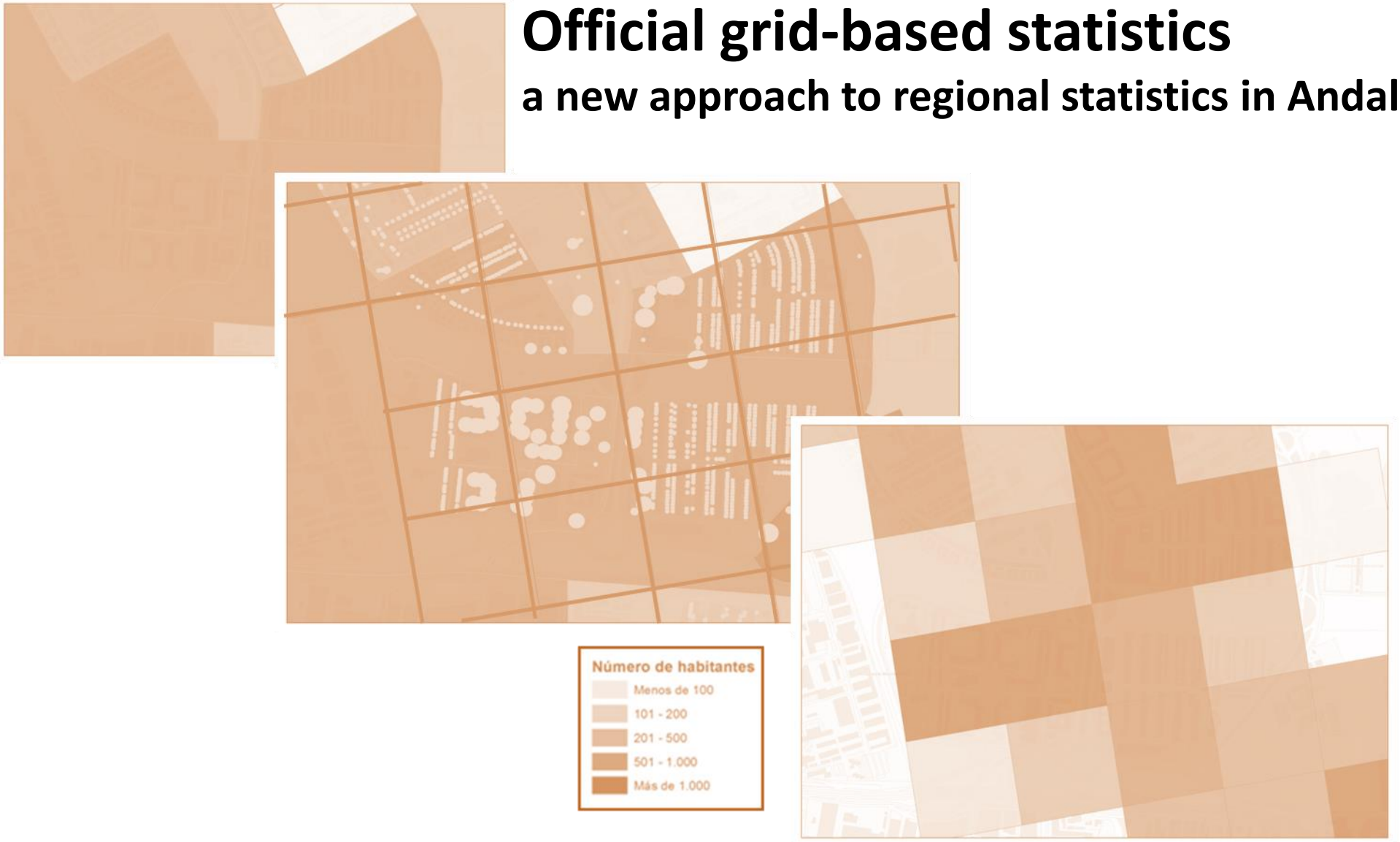
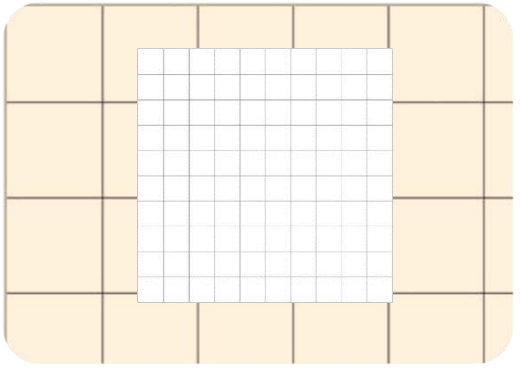
Instituto de Estadística y Cartografía de Andalucía
CONSEJERÍA DE ECONOMÍA Y CONOCIMIENTO

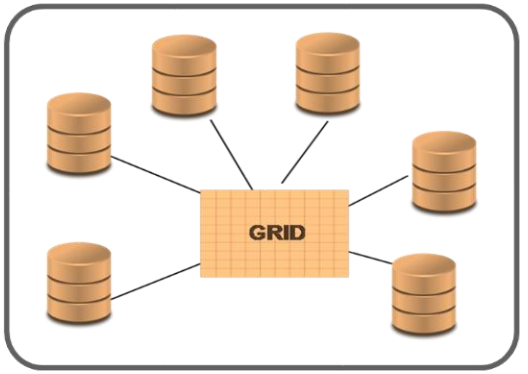
Official grid-based statistics

a new approach to regional statistics in Andalusia



European Forum for Geography and Statistics, 15-17 November, Paris 2016
Instituto de Estadística y Cartografía de Andalucía
Iria Enrique Regueira





Member of the Regional Government of Andalusia

Merge in 2011 of the Statistical Office of Andalusia and the Institute of Cartography of Andalusia

IECA works on integration and synergies Statistics & Geography

2013 Grid pilot project

2013 Population grid final results & further administrative registries linkage

Lessons learnt

Usability and data demand

Grid standard as reference, enlarge grid release domain

Need for infrastructure to integrate grid on structural statistics production

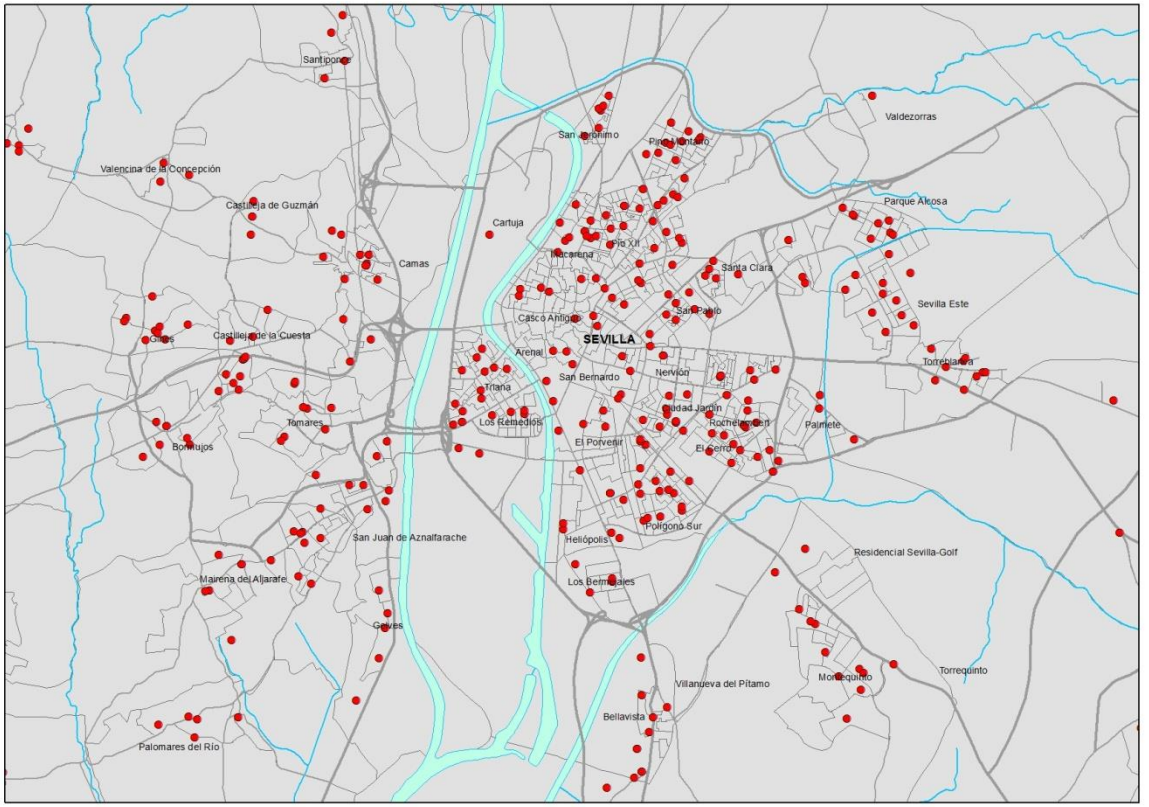
OVERVIEW

Regional official statistics are required to provide:

Relevant data for regional and local government actions: infrastructure planning, education services, health centres, social assitance units…

Useful information for events and phenomena analyses across the region

Public schools in Seville



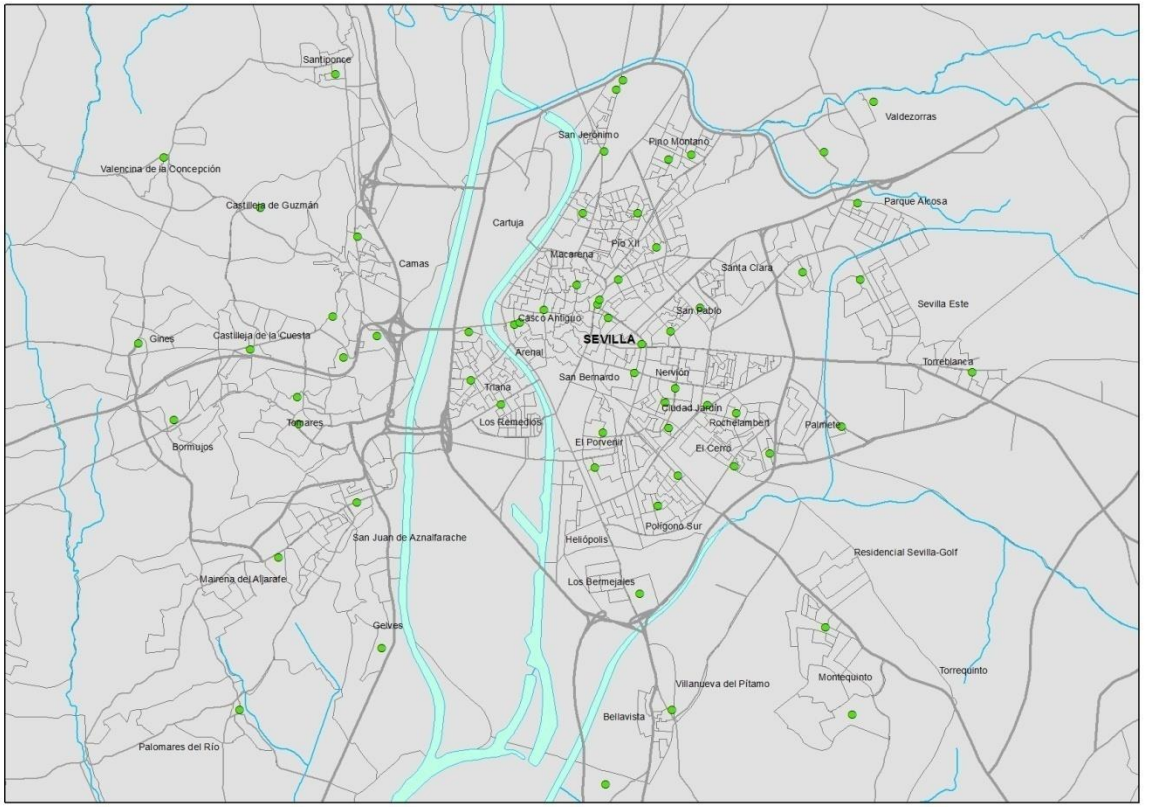
OVERVIEW

Regional official statistics are required to provide:

Relevant data for regional and local government actions: infrastructure planning, education services, health centres, social assitance units…

Useful information for events and phenomena analyses across the region

Public health services centers in Seville



High demand for more detailed and geographically disaggregated information

OPPORTUNITIES

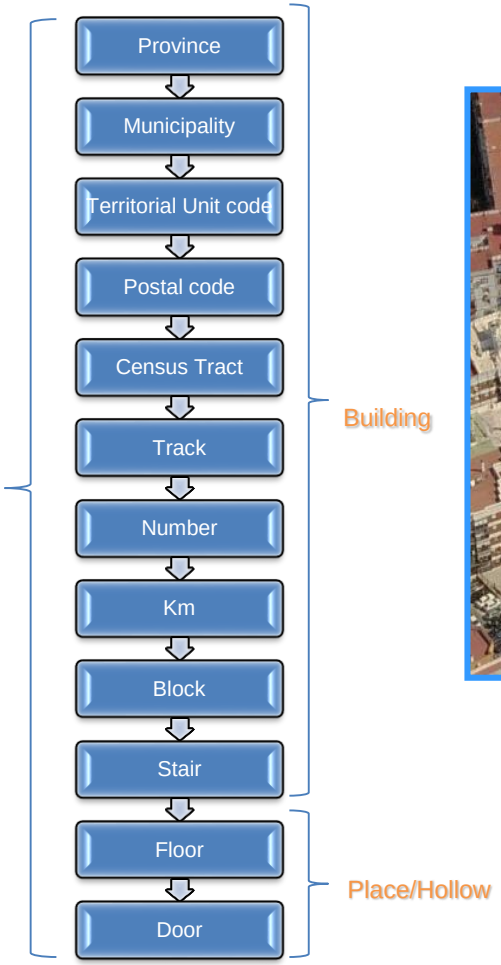
Data availability from multiple sources:

Surveys, census, registers…administrative files

Thematic & geographic disaggregation varies among sources

Sources such as census or administrative files usually contain postal addresses

Opportunity to **geocode** and **georeferenced**



Postal Address

OPPORTUNITIES

Geographically detailed data available

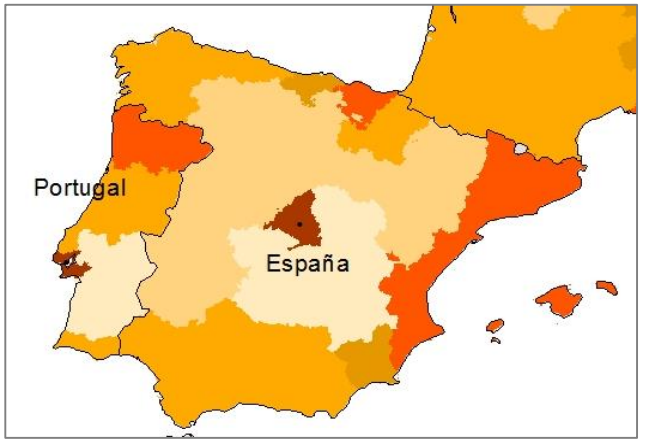
Official statistics usually reported according to a hierarchical system of administrative units

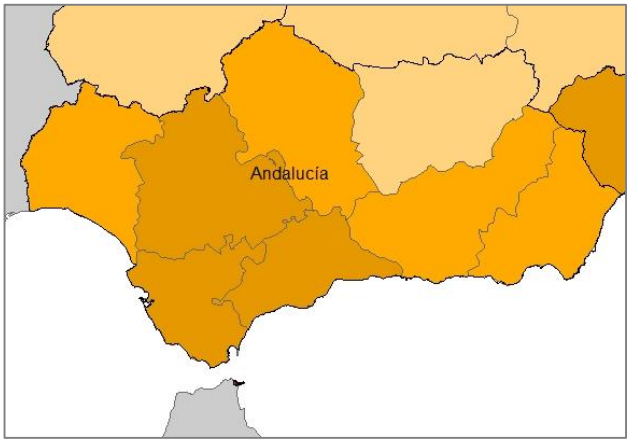
Modifiable areal unit

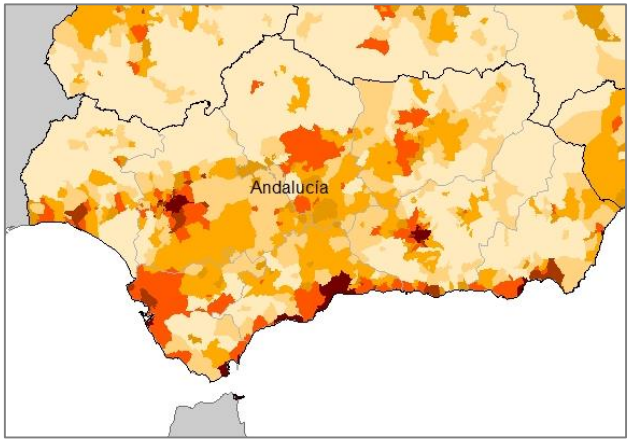
Nowadays regional planning and managing relies on municipal information or census tract data at best

The largest census tract in Andalusia, Andújar, covers 88.148 ha.

The shortest census tract in Andalusia, Aznalfarache, covers 0,37 ha.







High geographically detailed info, independent from administrative boundaries crucial for regional and local administration

OPPORTUNITIES

Geographically detailed data available, grid data dissemination

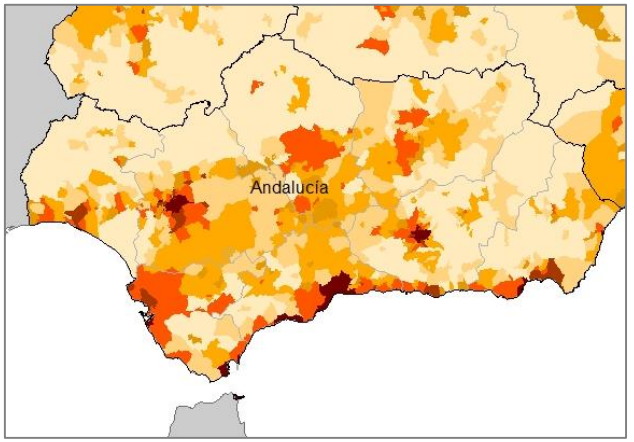
Equal size allowing for easy comparison

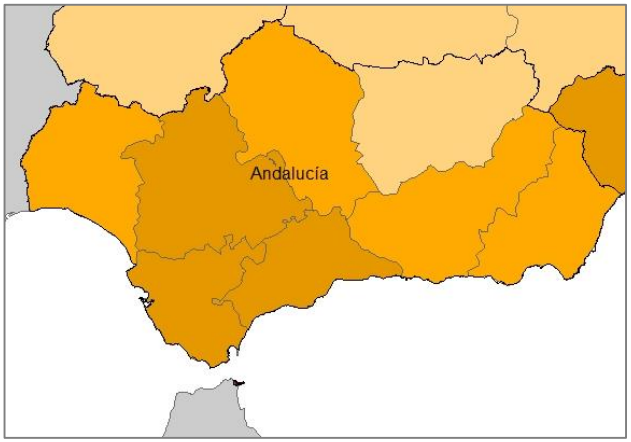
Stability over time

Easy integration with other scientific data

Grids can be built hierarchically in terms of cell size, thus matching the study area

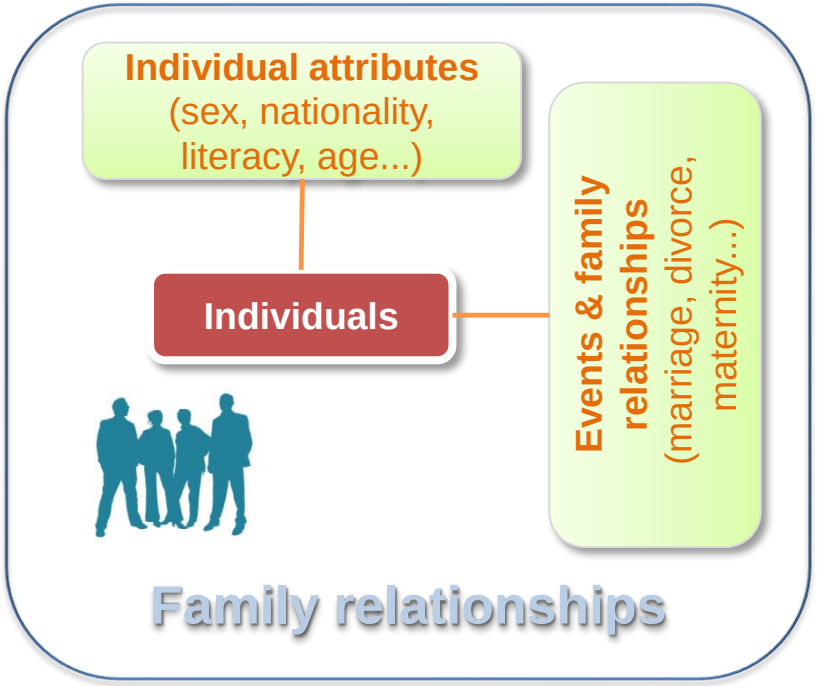
Allow for spatial modelling techniques



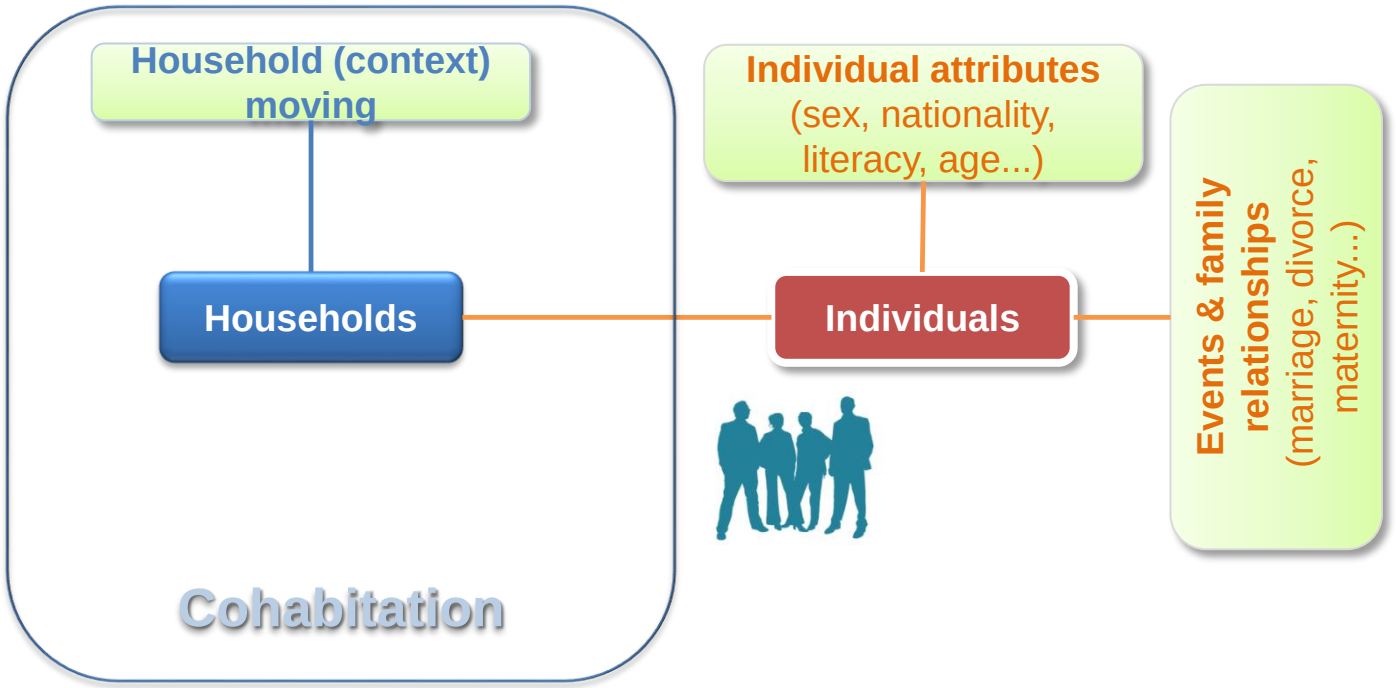


High geographically detailed info, independent from administrative boundaries crucial for regional and local administration

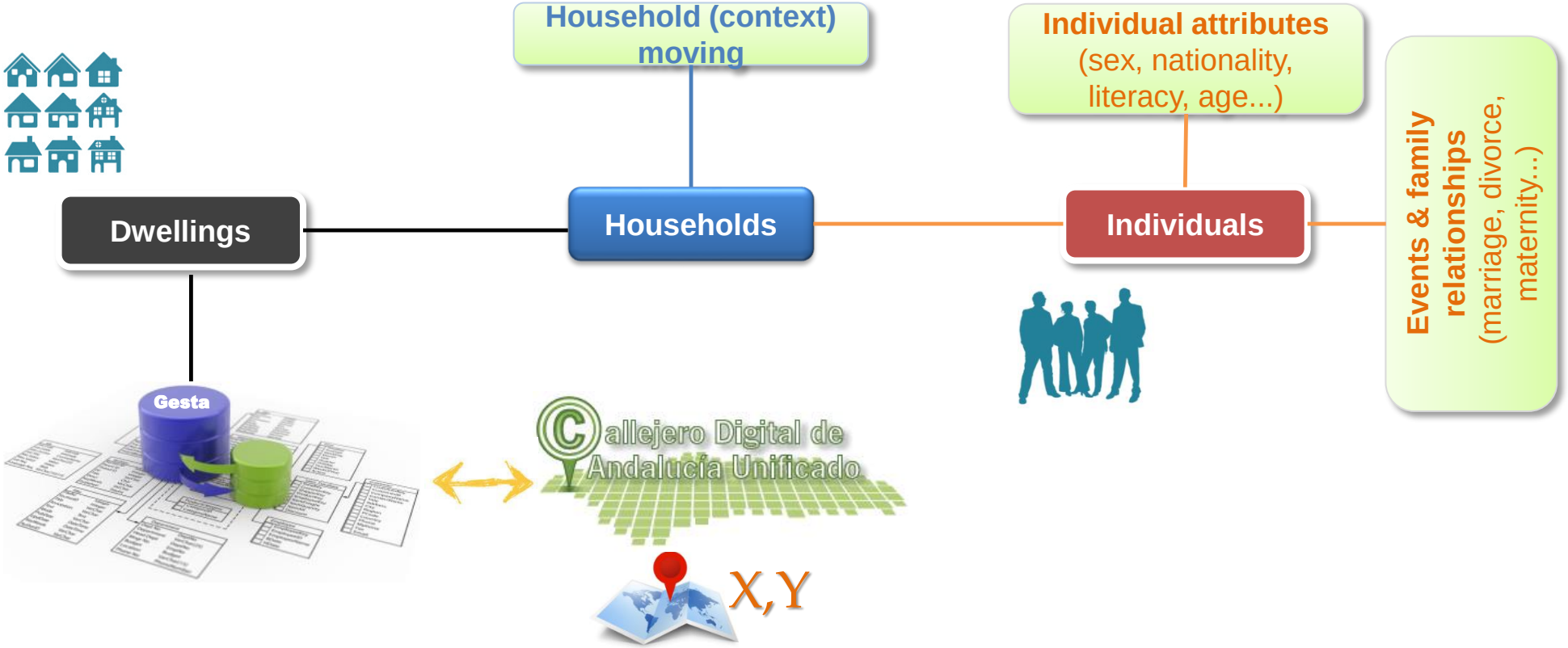
GRID PILOT PROJECT



GRID PILOT PROJECT



GRID PILOT PROJECT



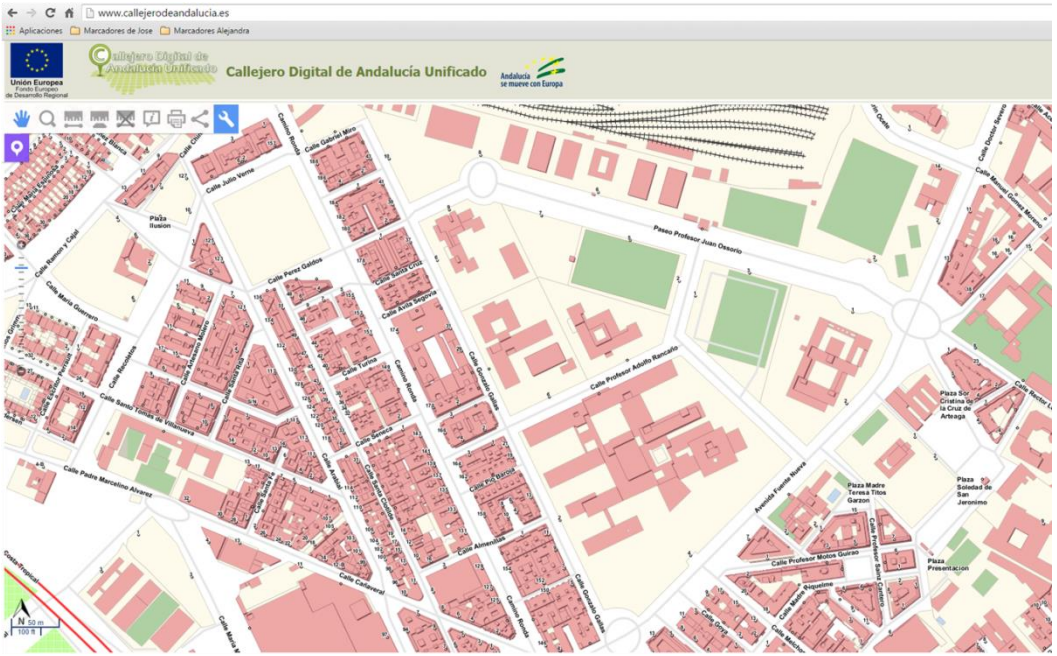
GRID PILOT PROJECT

Sources & Methods

Unified Digital Street Map of Andalusia (Spatial info)

Longitudinal population database of Andalusia

Alink in-home software solution: A semi supervised Statistical learning software for managing, linking and merging non structured data sets



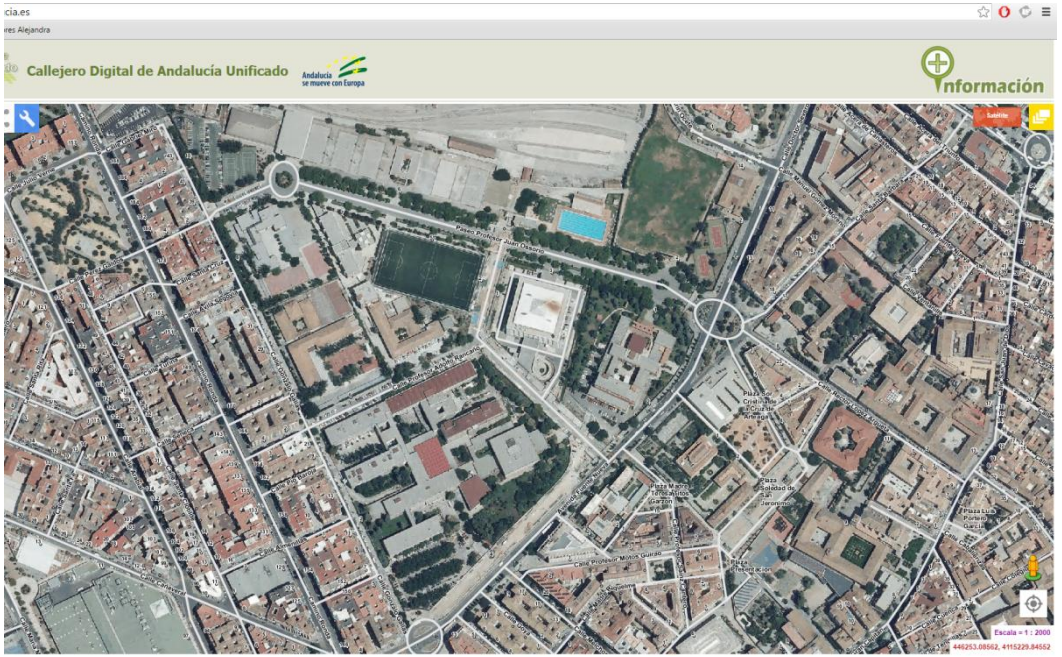
GRID PILOT PROJECT

Sources & Methods

Unified Digital Street Map of Andalusia (Spatial info)

Longitudinal population database of Andalusia

Alink in-home software solution: A semi supervised Statistical learning software for managing, linking and merging non structured data sets



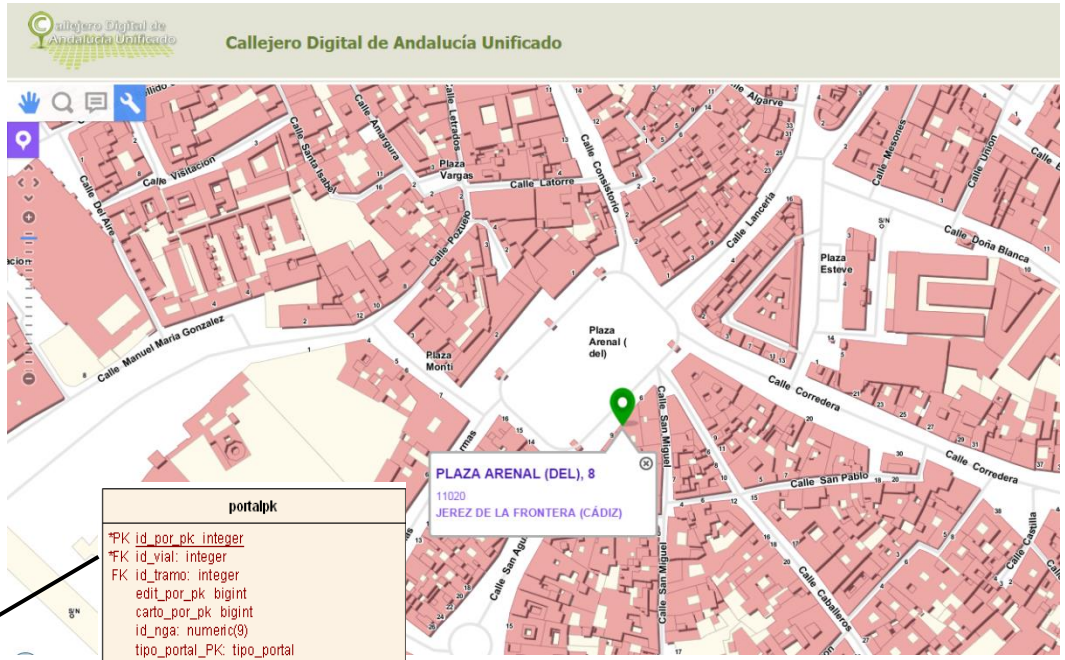
GRID PILOT PROJECT

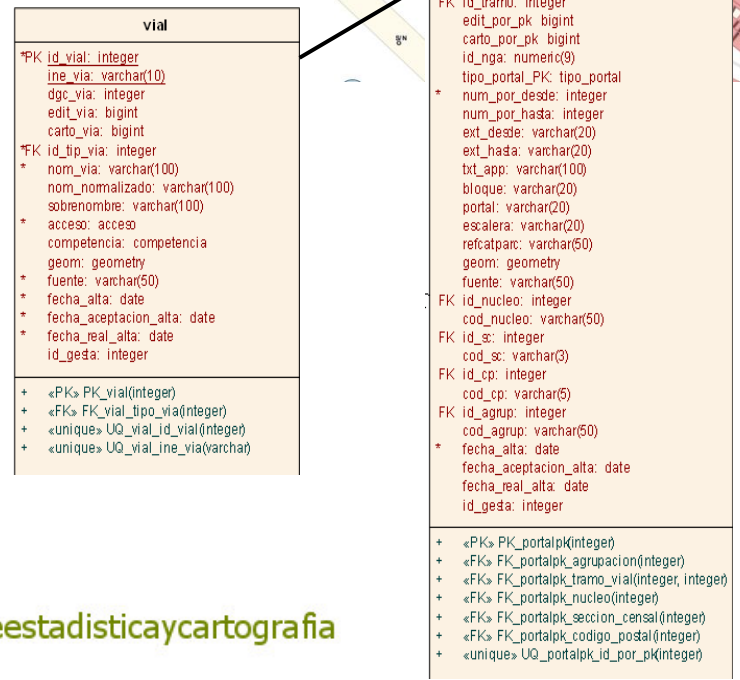
Sources & Methods

Unified Digital Street Map of Andalusia (Spatial info)

Longitudinal population database of Andalusia

Alink in-home software solution: A semi supervised Statistical learning software for managing, linking and merging non structured data sets





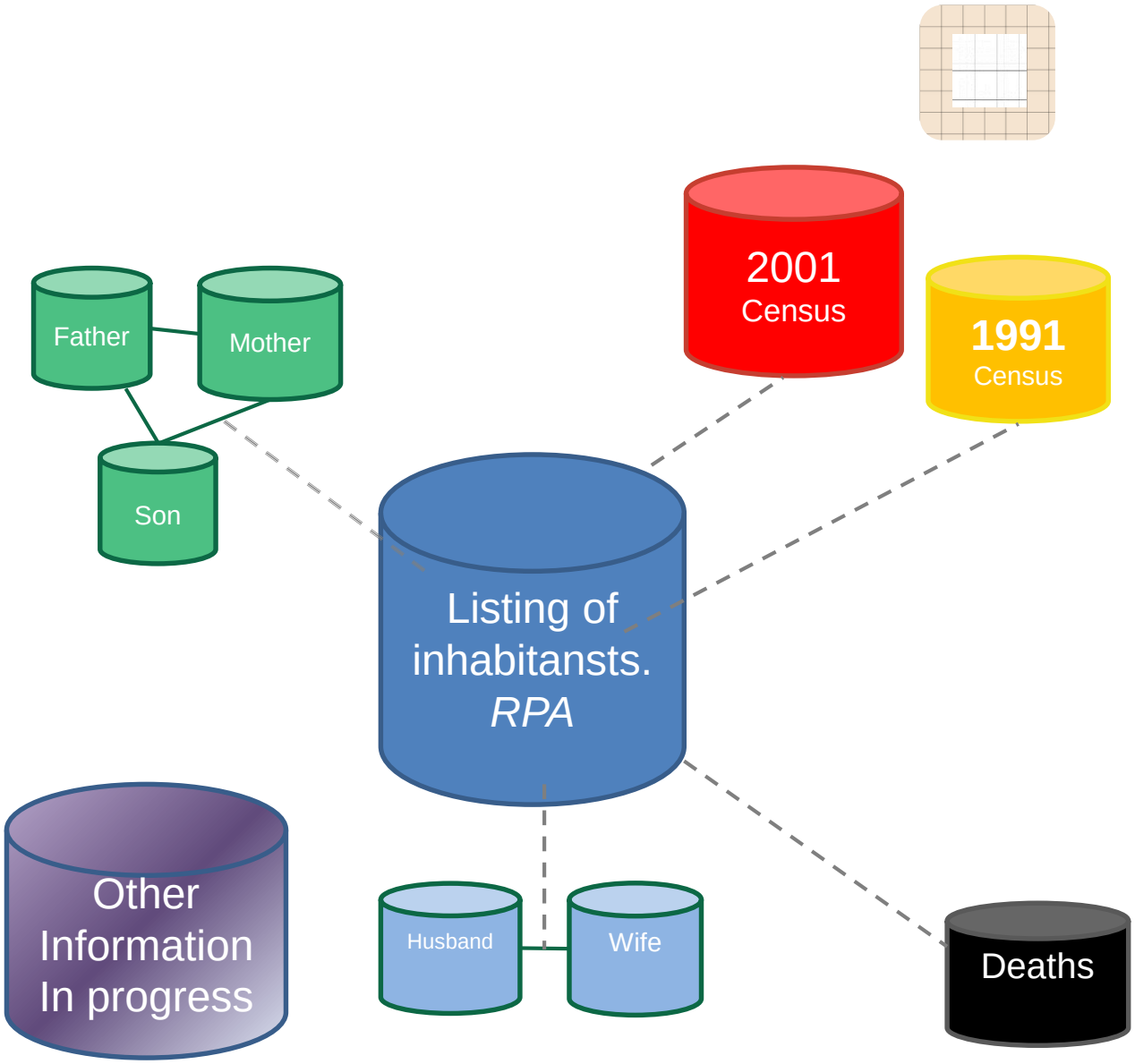
GRID PILOT PROJECT

Sources & Methods

Unified Digital Street Map of Andalusia (Spatial info)

Longitudinal population database of Andalusia (BDLPA)

Alink in-home software solution: A semi supervised Statistical learning software for managing, linking and merging non structured data sets



GRID PILOT PROJECT

Sources & Methods

Unified Digital Street Map of Andalusia (Spatial info)

Longitudinal population database of Andalusia (BDLPA)

Alink in-home software solution: A semi supervised Statistical learning software for managing, linking and merging non structured data sets

http://www.juntadeandalucia.es/institutodeestadisticaycartografia/ieagen/otrosServidores/software/index-en.htm

GRID PILOT PROJECT

2013 Grid, hybrid model

Bottom-up approach

Successful georeferentiation, 87.4% of the population settled in Andalusia

28,806 inhabited cells at this stage of the process

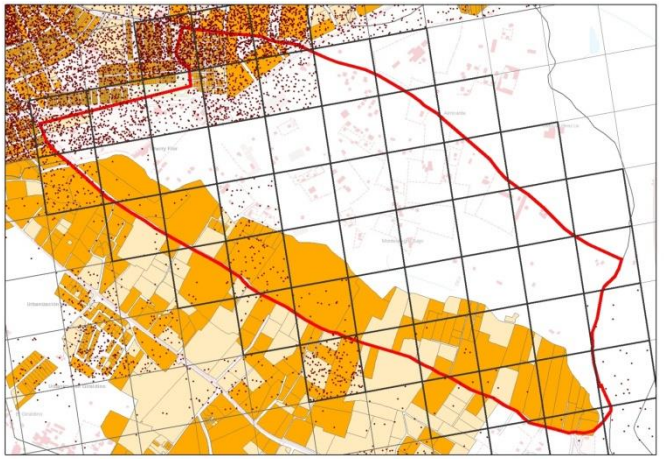
Top-down approach

Allocation of non georeferenced households based on statistical ancillary information from urban Cadastre and georeferenced population (BDLPA info)

12,6% of the population settled in Andalusia. 11,613 extra inhabited cells at this stage of the process

A grid of 1,416,093 cells 250mx250m covering Andalusia area

40,419 inhabited cells (2.9% of cells)



GRID PILOT PROJECT

2013 Grid, final results

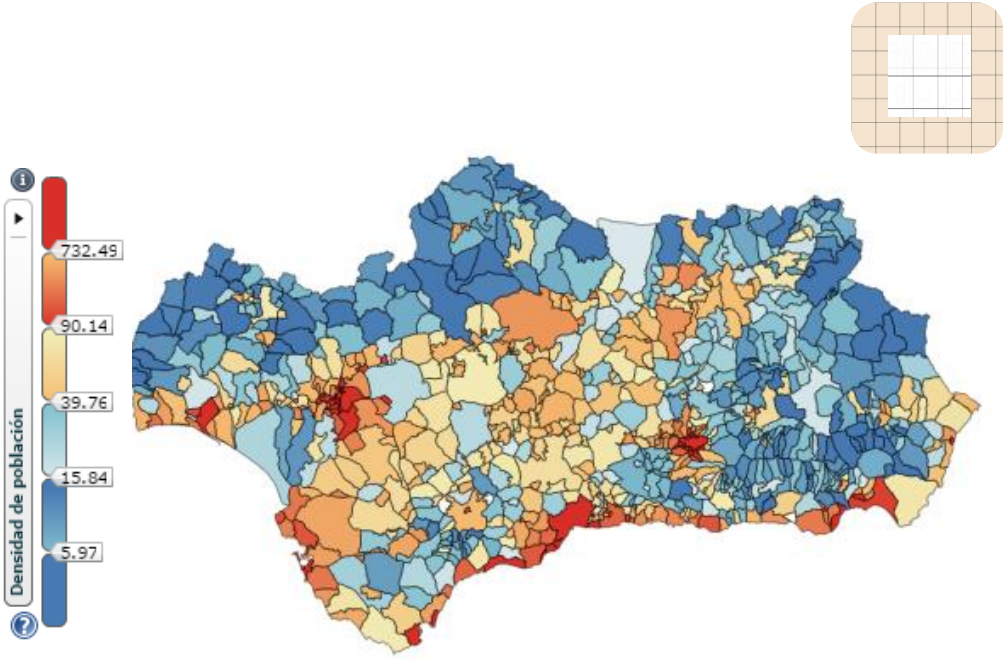
Bottom-up approach

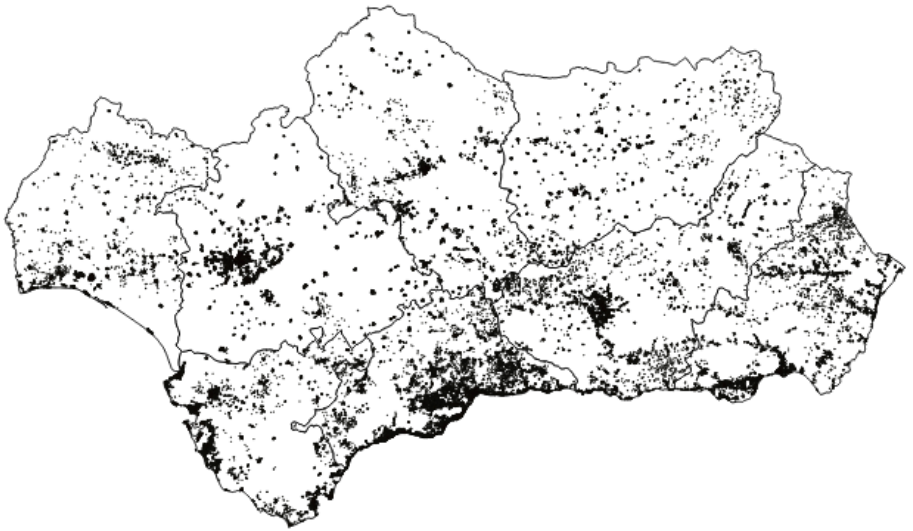
A second edition of 2013 Population grid was elaborated including 2011 Census building data.

Successful georeferentiation, 97% of the population settled in Andalusia.

A grid of 1,416,093 cells 250mx250m covering Andalucia area

48,187 inhabited cells (3.4% of cells)





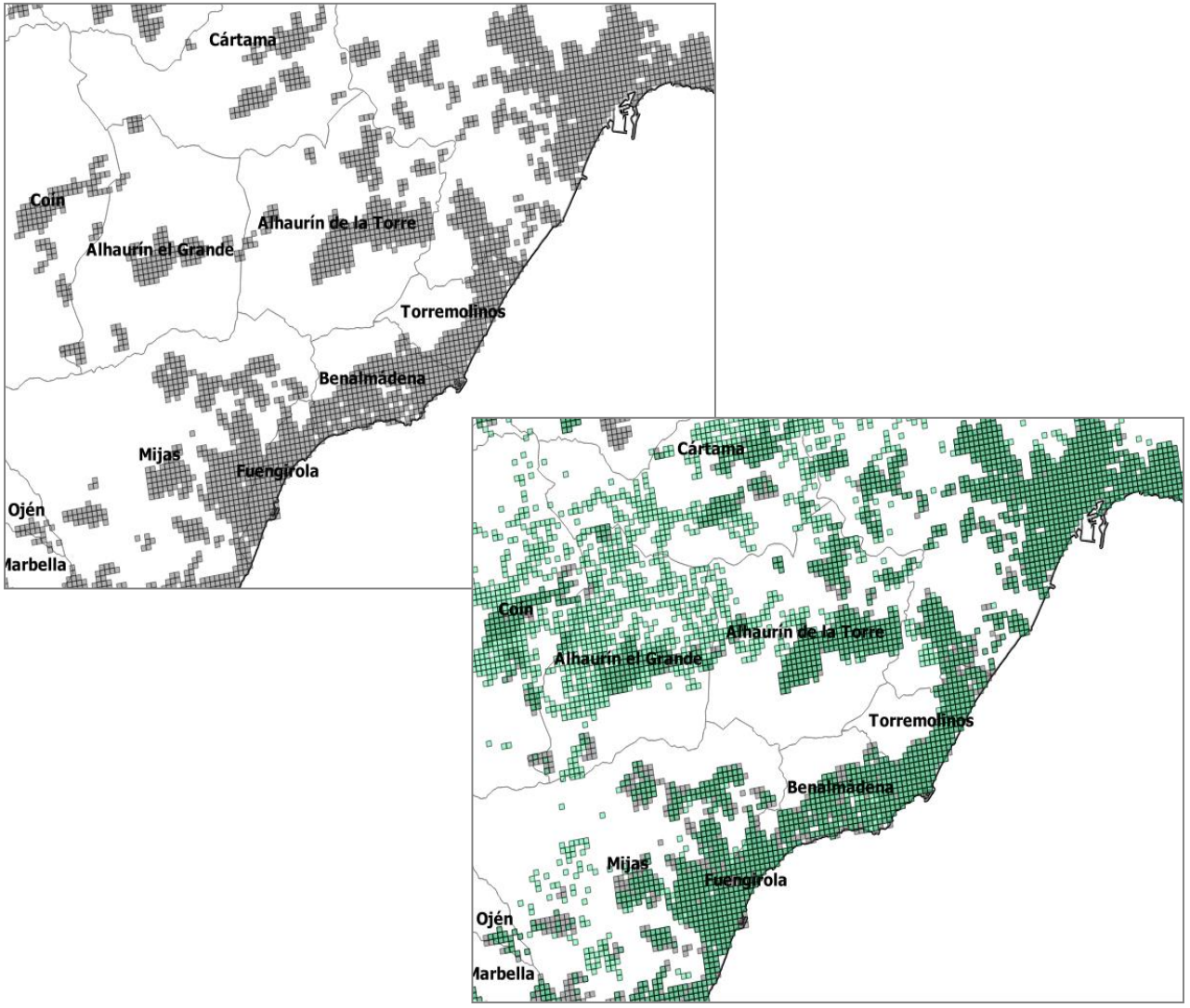
GRID PILOT PROJECT

Lessons learnt

Modelling is helpful but ancillary variables are crucial

Hybrid model relied on variables from urban cadastre and georeferenced population.

Our model tended to systematically underestimate thinly populated areas while overestimating densely populated urban areas to balance totals in census tracts.



GRID PILOT PROJECT

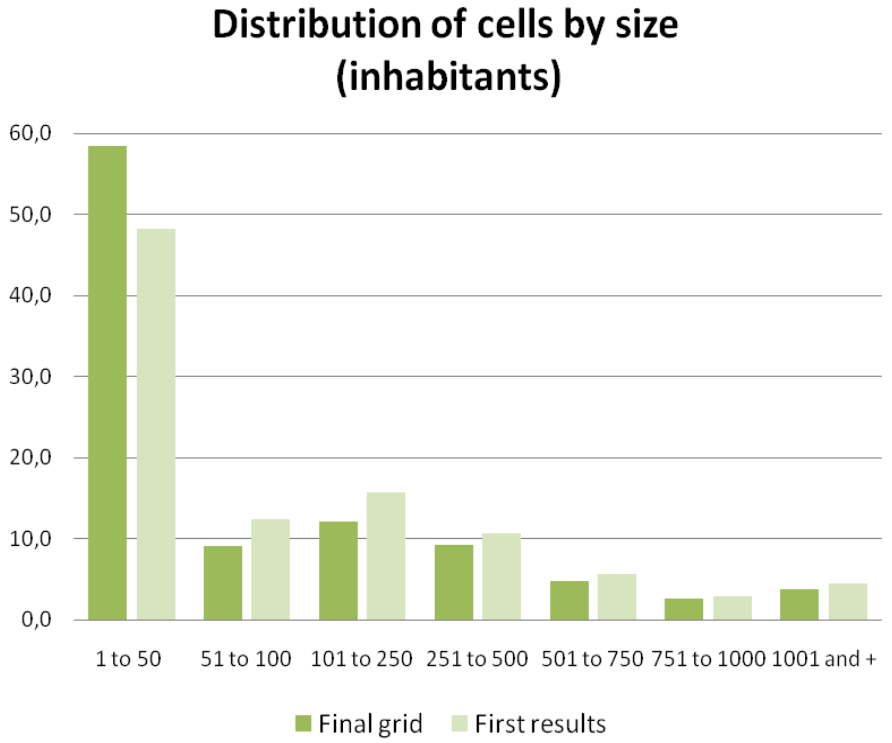
Lessons learnt

Modelling is helpful but ancillary variables are crucial

Hybrid model relied on variables from urban cadastre and georeferenced population.

Our model tended to systematically underestimate thinly populated areas while overestimating densely populated urban areas to balance totals in census tracts.

	Final grid	First results
Total number of cells	48187	40419
Median	27	55
Mode	1	4
Minimum	1	1
Maximum	4988	5339
Percentiles		
25	5	15
50	27	55
75	175	232



USABILITY & DATA DEMAND

Administrative planning & evaluation

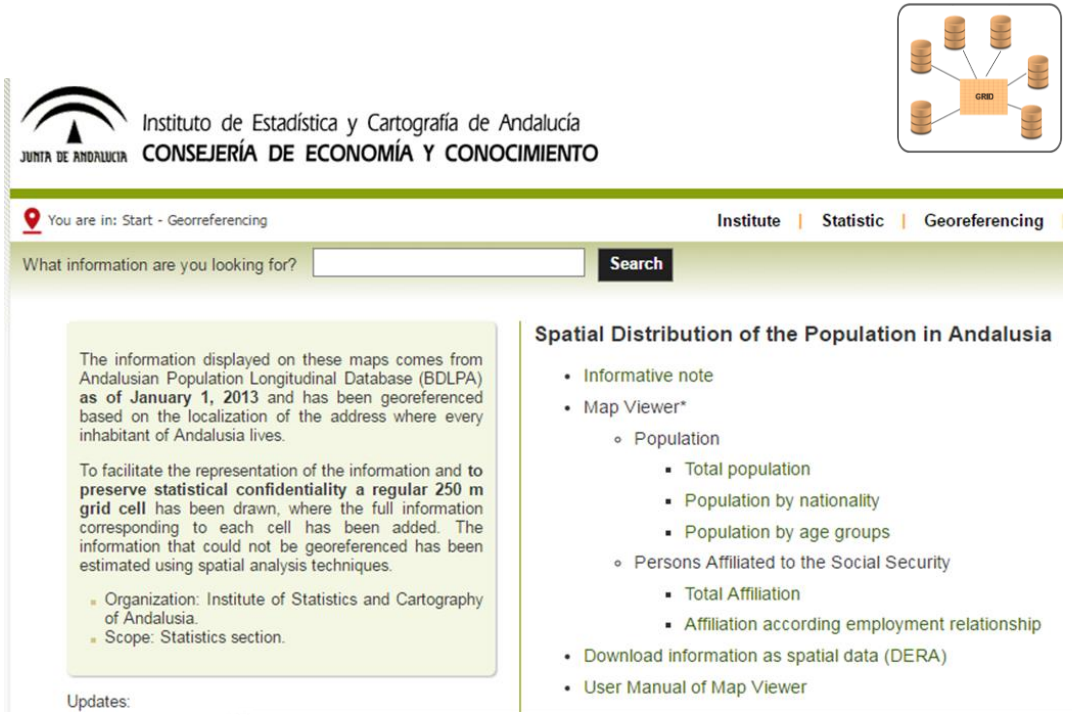
Education department uses grid data in order to forecast schooling demand

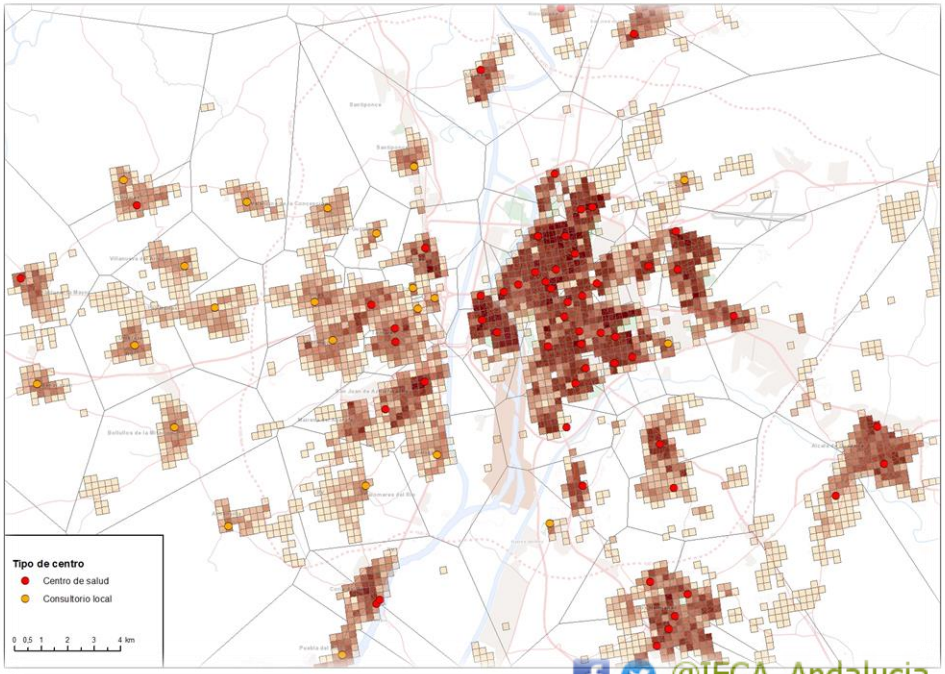
Infrastructures department used grid data in order to estimate potential demand for public transport

Further analyses & grid indicators

Bicycle lane accessibility

Mortality ratio by grid cells





USABILITY & DATA DEMAND

Administrative planning & evaluation

Education department uses grid data in order to forecast schooling demand

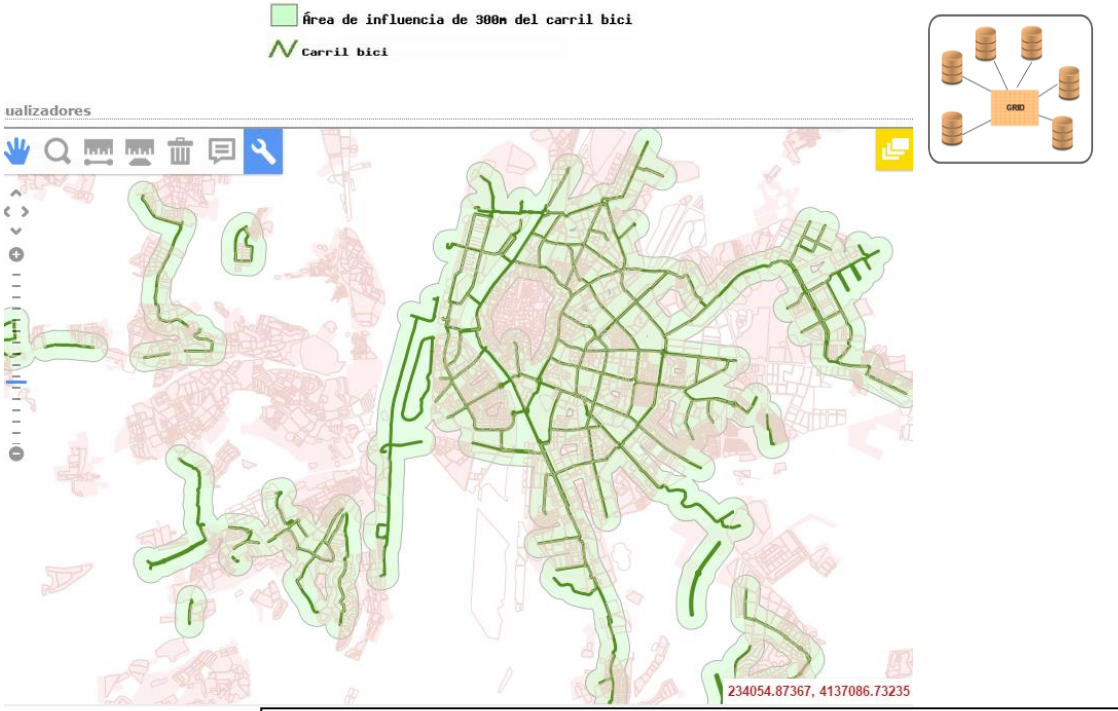
Infrastructures department used grid data in order to estimate potential demand for public transport

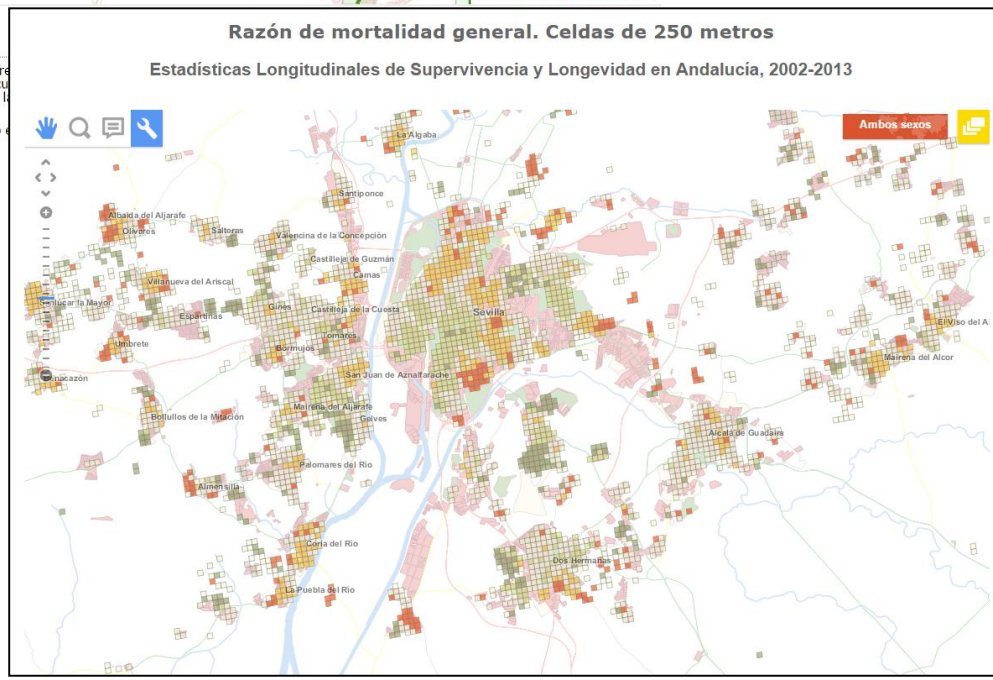
Further analyses & grid indicators

Bicycle lane accessibility

Mortality ratio by grid cells: Smoothed Standardized Mortality ratio by grid Cells (Survival & longevity statistics)

Grid mortality indicators (250m and 1 km)





GRID STANDARD AS REFERENCE

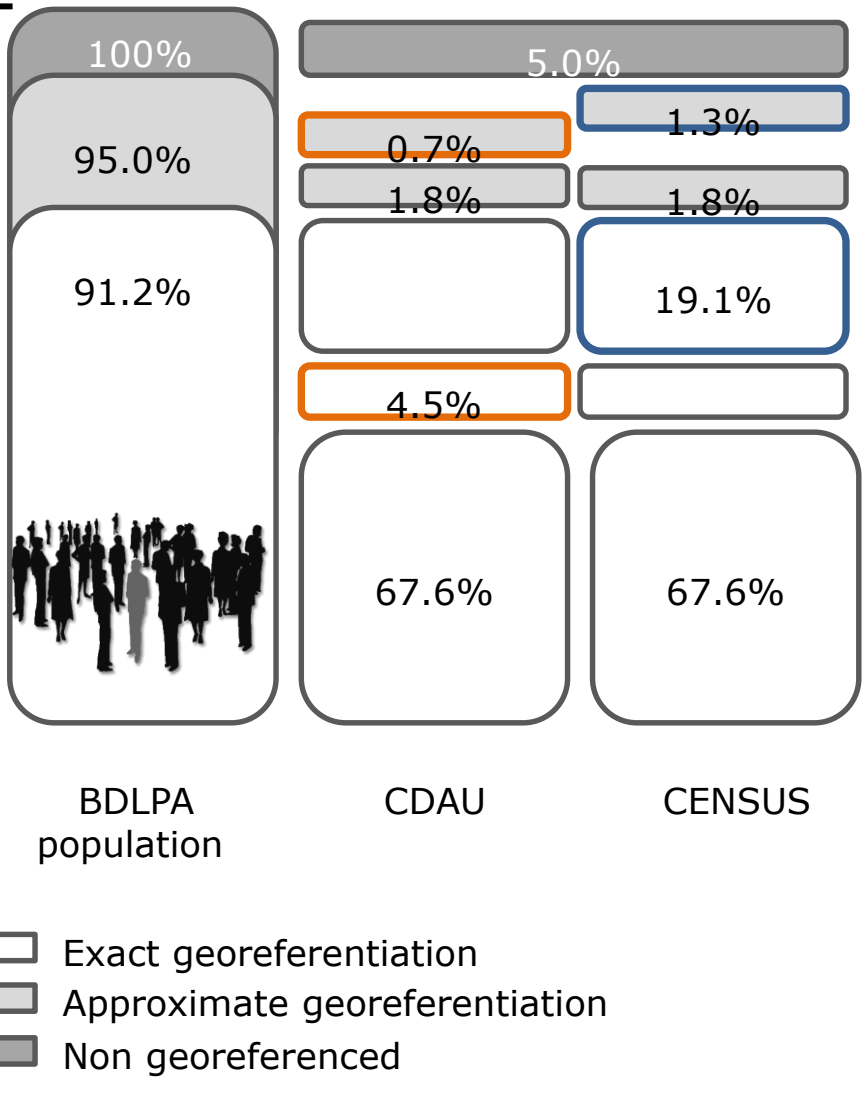
Current projects

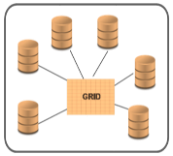
Georeferentiation of 2014 & 2015 population.

Similar approach to Grid 2013.

BDLPA completed and upgraded a inhabited building address' repository.

Linkage results with CDAU and Census, georeferentiation of 95%





GRID STANDARD AS REFERENCE

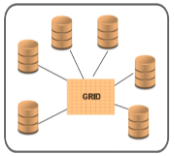
Current projects

Georeferentiation of 2014 & 2015 population.

Revision of allocation model of non georeferenced: Identification of uninhabited/vacant residential buildings as potential location for non georeferenced building entrances from BDLPA.

Revision of non-disclosure rules for hierarchical areas and nested tables supression, quadtrees.

ENLARGE GRID DOMAIN



Household (context) moving

Individual attributes (sex, nationality, literacy, age...)

Events & family relationships (marriage, divorce, maternity...)

Dwellings

Households

Individuals

Callejero Digital de Andalucía Unificado

X,Y

Labour (social security, tax info...)

- Unemployment database
- Education Database
- Health Database
- Social Assistance Database

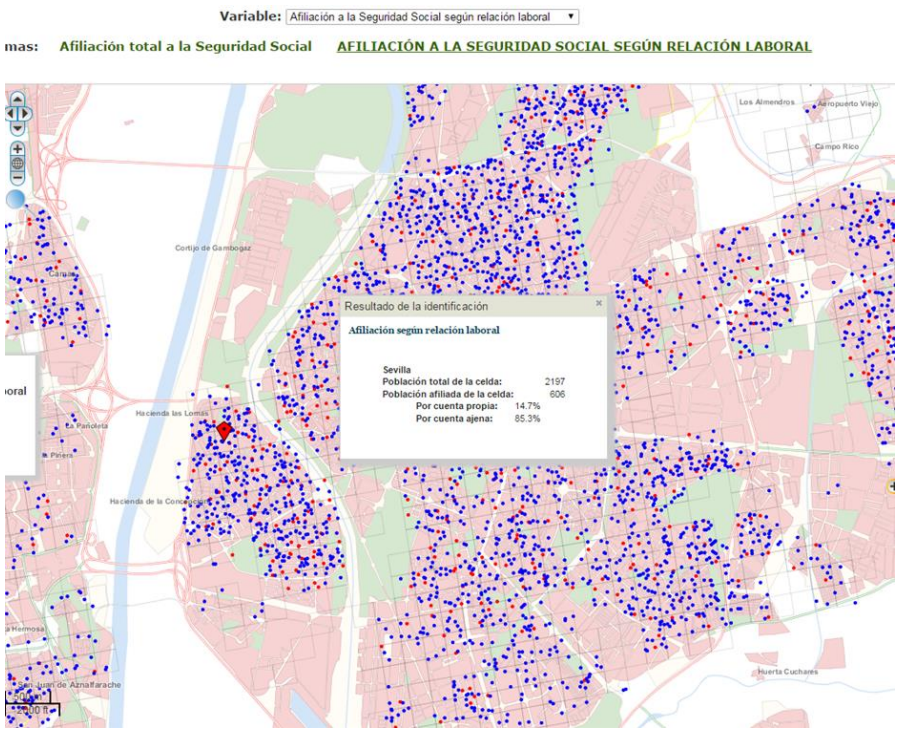
ENLARGE GRID DOMAIN

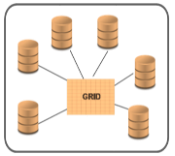
Current projects

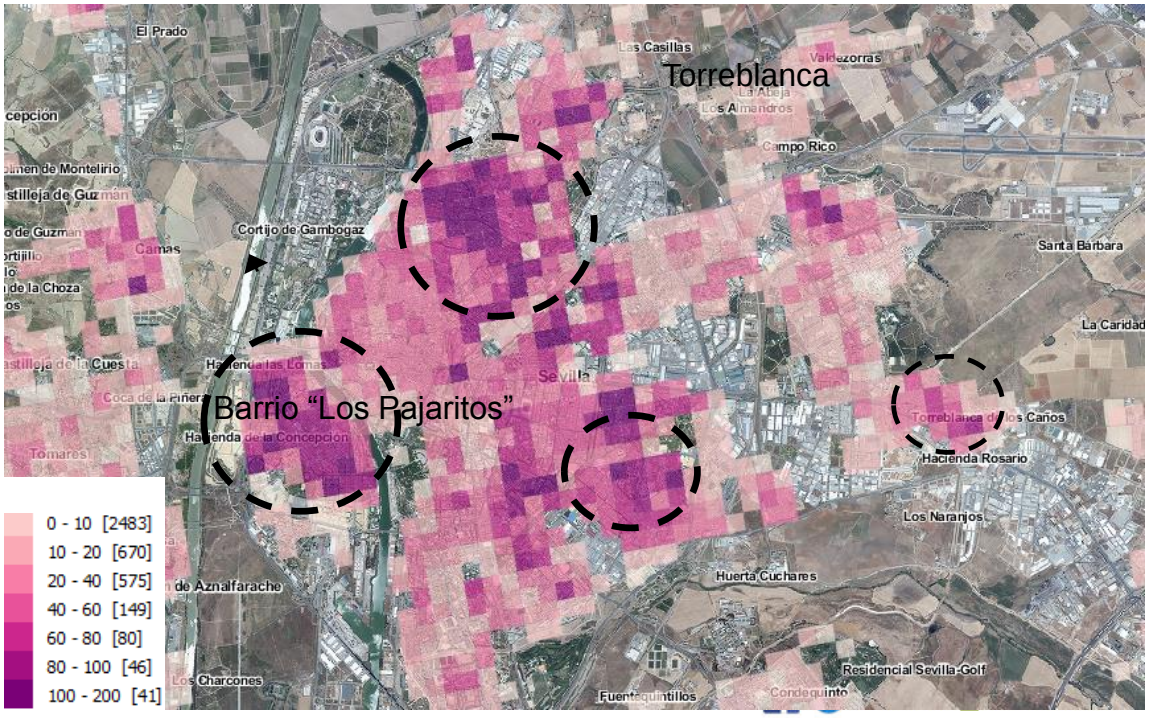
Enlarge grid domain by linkage of georeferenced population and administrative registers (education, health…)

Labour status, salaried, self-employed from social security records linked to 2013 population

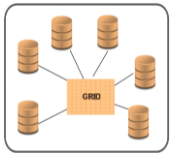
Pensioners by type of assistance will be linked from 2014 population onwards







ENLARGE GRID DOMAIN



Household (context) moving

Individual attributes (sex, nationality, literacy, age...)

Events & family relationships (marriage, divorce, maternity...)

Dwellings

Households

Individuals

Callejero Digital de Andalucía Unificado

X,Y

Establishments

Business & Establishments frame

Labour (social security, tax info...)

- Unemployment database
- Education Database
- Health Database
- Social Assistance Database

ENLARGE GRID DOMAIN

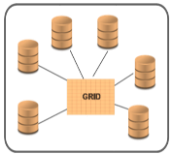
Current projects

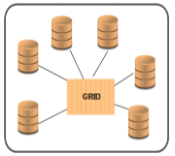
Georeferentiation of establishments and employment in establishment in Andalusia.

Similar approach to population grid. Georeferentiation, 90% of the establishments settled in Andalusia 2014

High rate of approximate georreferentiation 28% georeferentiation through neighbouring area or street center is highly disturbing in establishment location

Establishment characteristics & location are “visible”





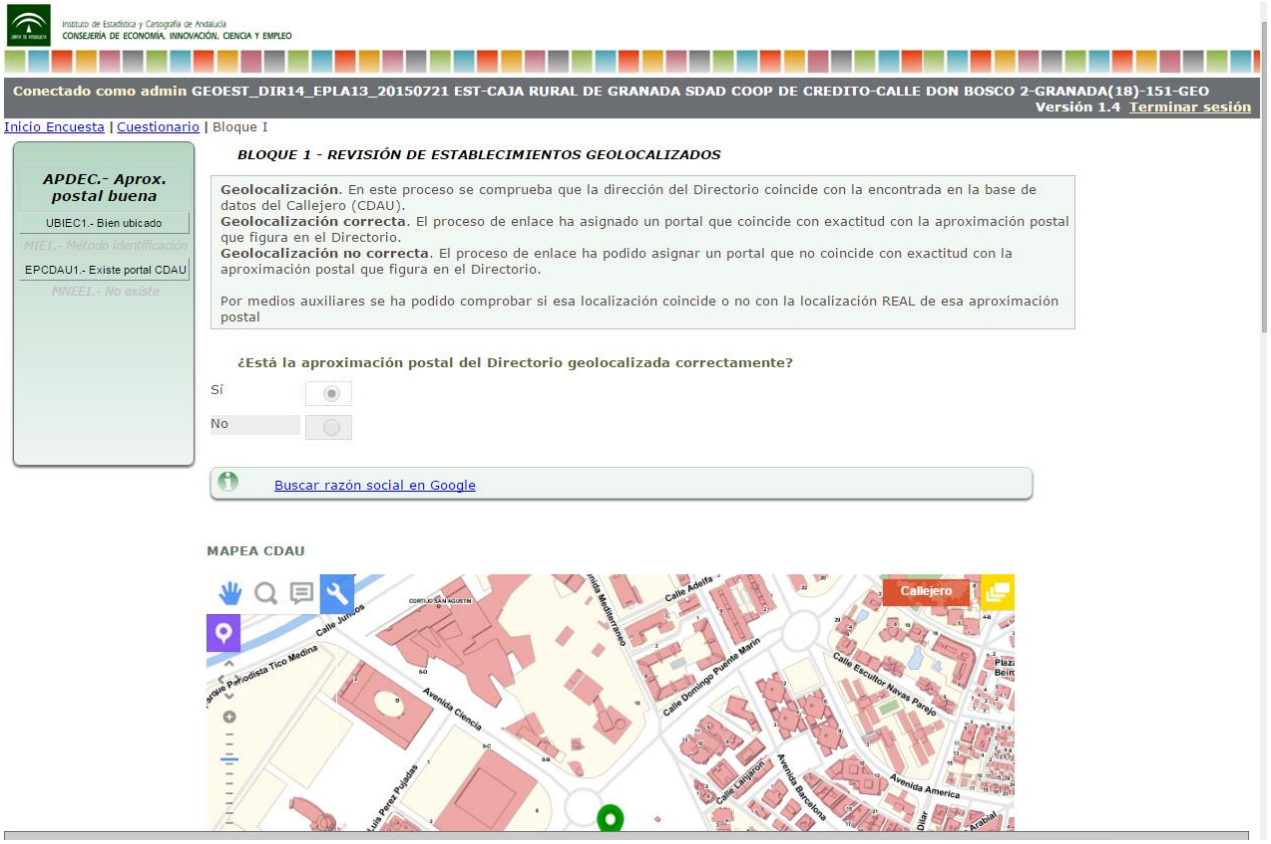
ENLARGE GRID DOMAIN

Current projects

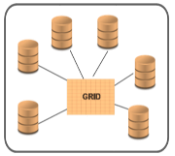
Georeferentiation of establishments and its employment in Andalusia.

Establishments over 50 employees assisted revision of Alink geocodification and Establishment frame address.

A guided form accessing CDAU, internet and street view info allows statistical agents to refine Establishment frame address info and/or assigned geocodification



ENLARGE GRID DOMAIN



Household (context) moving

Individual attributes (sex, nationality, literacy, age...)

Events & family relationships (marriage, divorce, maternity...)

Dwellings

Households

Individuals

Callejero Digital de Andalucía Unificado

X,Y

Establishments

Business & Establishments frame

Labour (social security, tax info...)

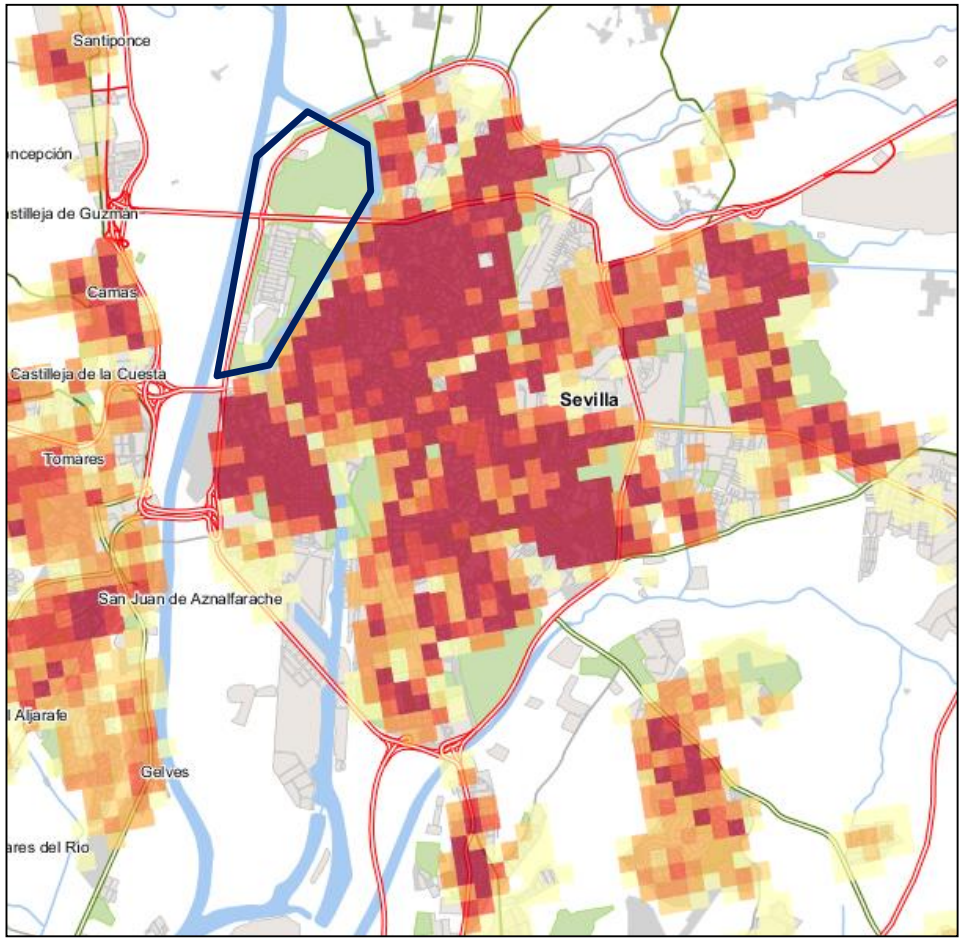
Unemployment database

Education Database

Health Database

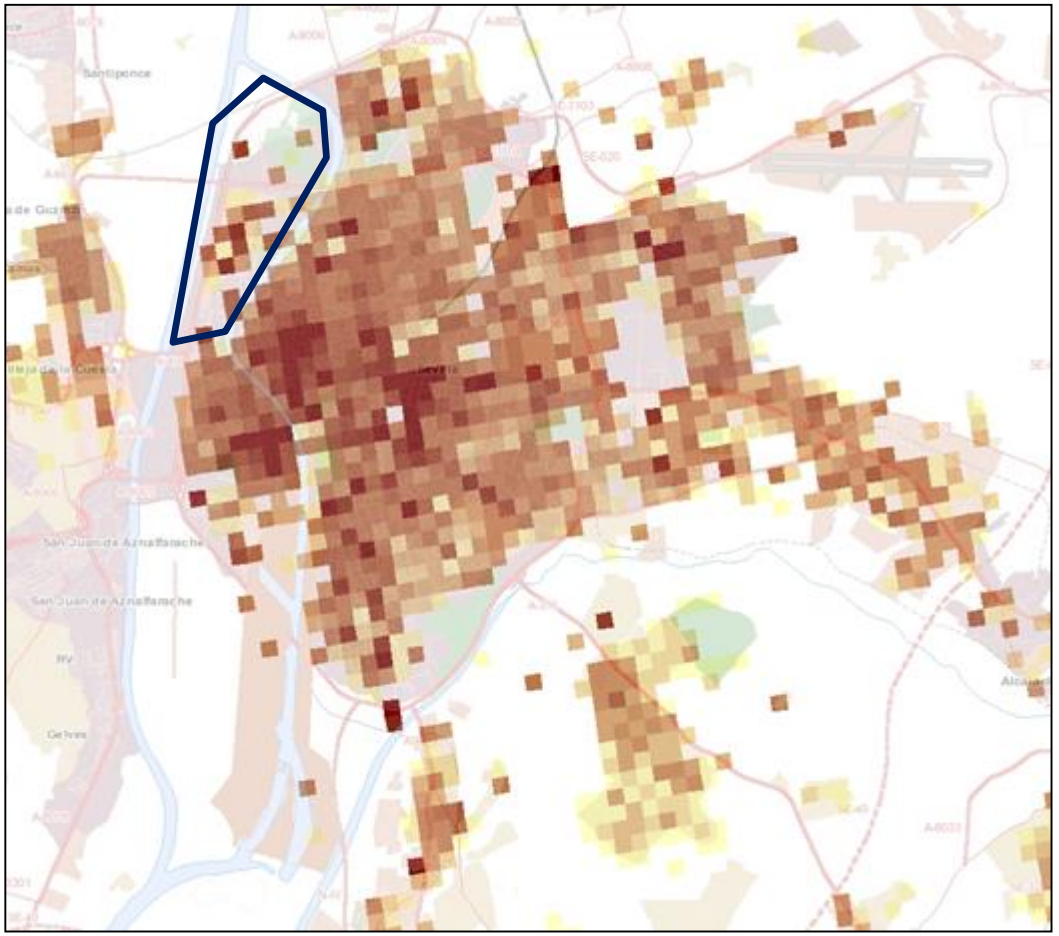
Social Assistance Database

Population grid
(“night grid”)



Número de habitantes
- Menos de 50
- De 50 a 100
- De 100 a 250
- De 250 a 500
- De 500 a 750
- De 750 a 1000
- Más de 1000

Establishments grid
(“day grid”)

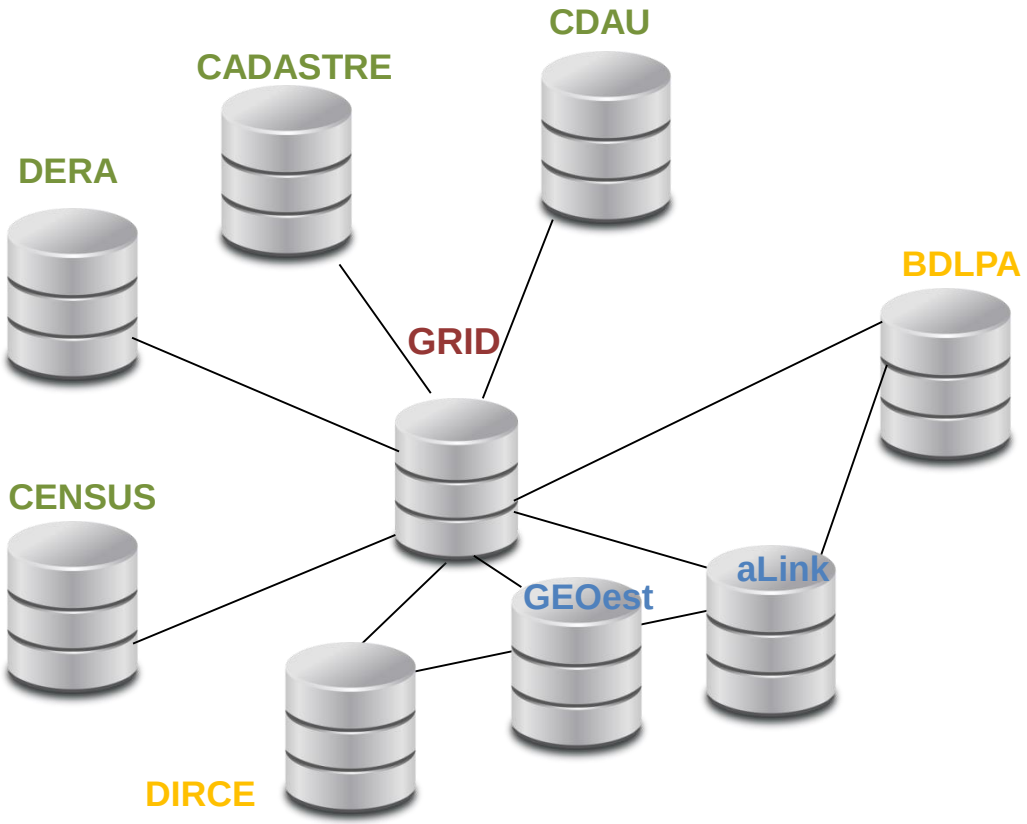


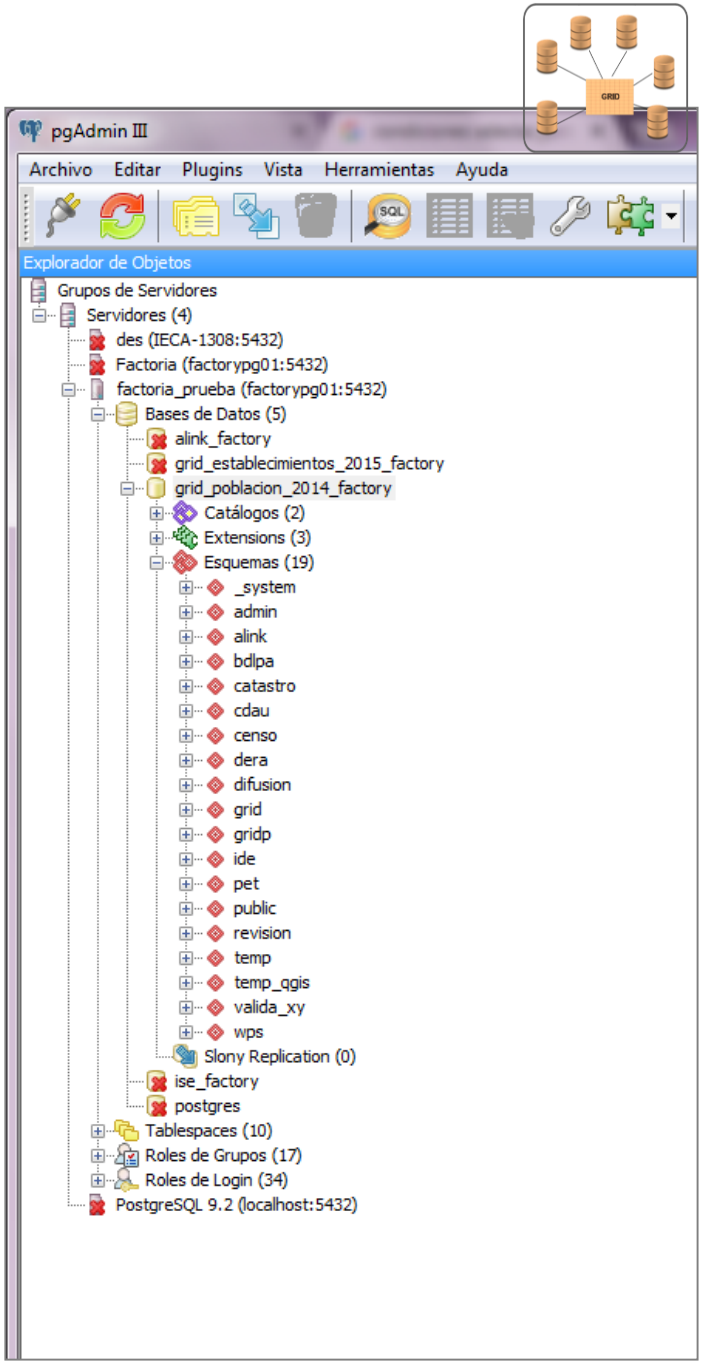
Empleo total
- Menos de 10
- 11 - 25
- 26 - 50
- 51 - 100
- 101 - 250
- 251 - 500
- 501 - 750
- 751 - 1000
- Más de 1.000

IT INFRASTRUCTURE

Current projects

Need for Gis architecture & infrastructure to integrate grid on structural statistics production





THANKS FOR YOUR ATTENTION!

http://www.juntadeandalucia.es/institutodeestadisticaycartografia/index-en.html

Iria.enrique@juntadeandalucia.es