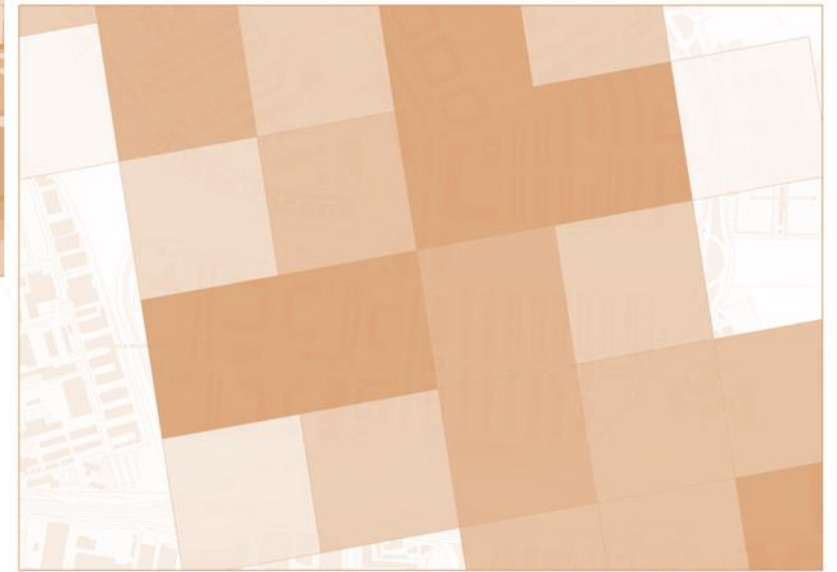


Official grid-based statistics a new approach to regional statistics in Andalusia



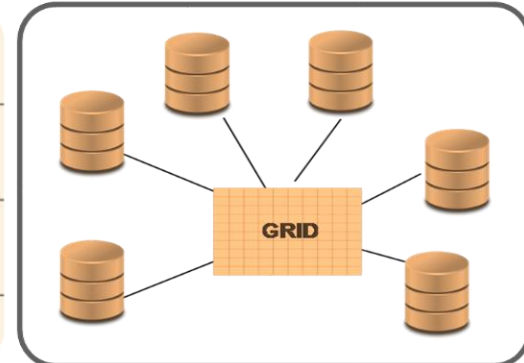
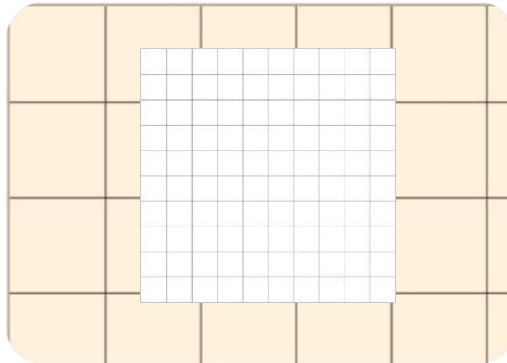
European Forum for Geography and Statistics, 15-17 November, Paris 2016

Instituto de Estadística y Cartografía de Andalucía

Iria Enrique Regueira

www.juntadeandalucia.es/institutodeestadisticaycartografia

  @IECA_Andalucia



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

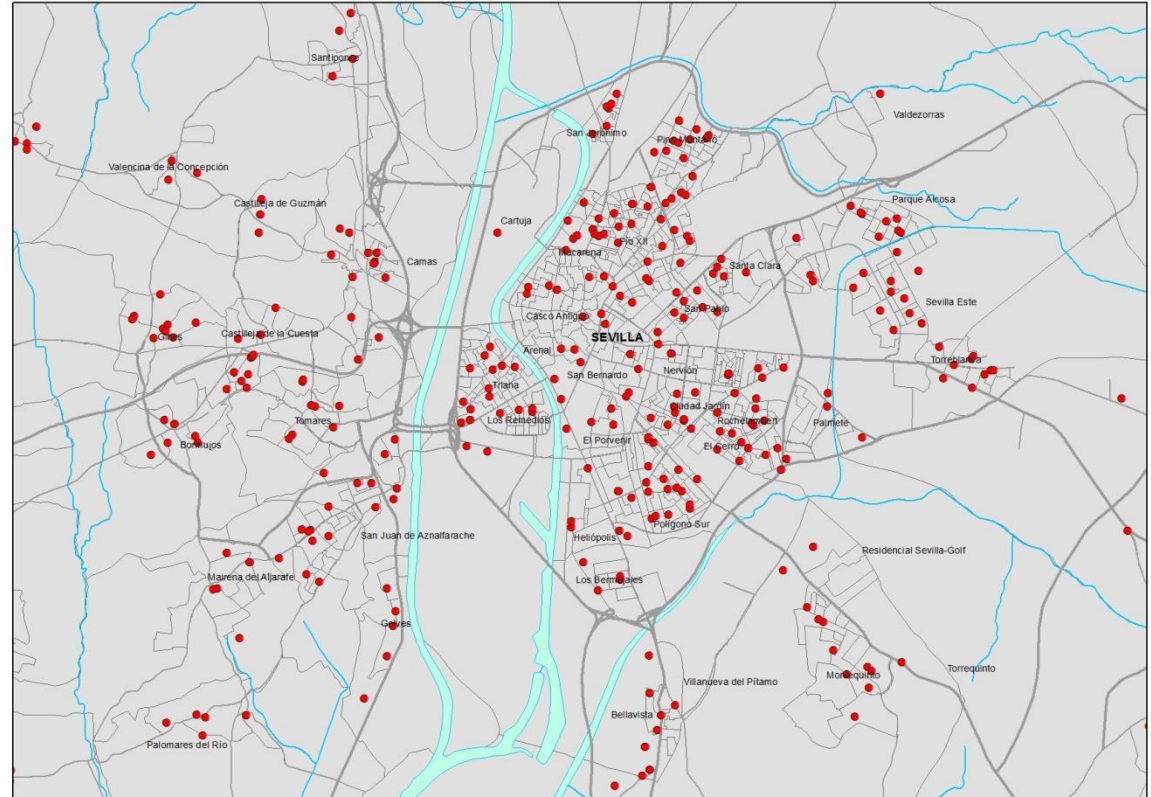
OVERVIEW

Regional official statistics are required to provide:

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

Useful information for events and phenomena analyses across the region

Public schools in Seville



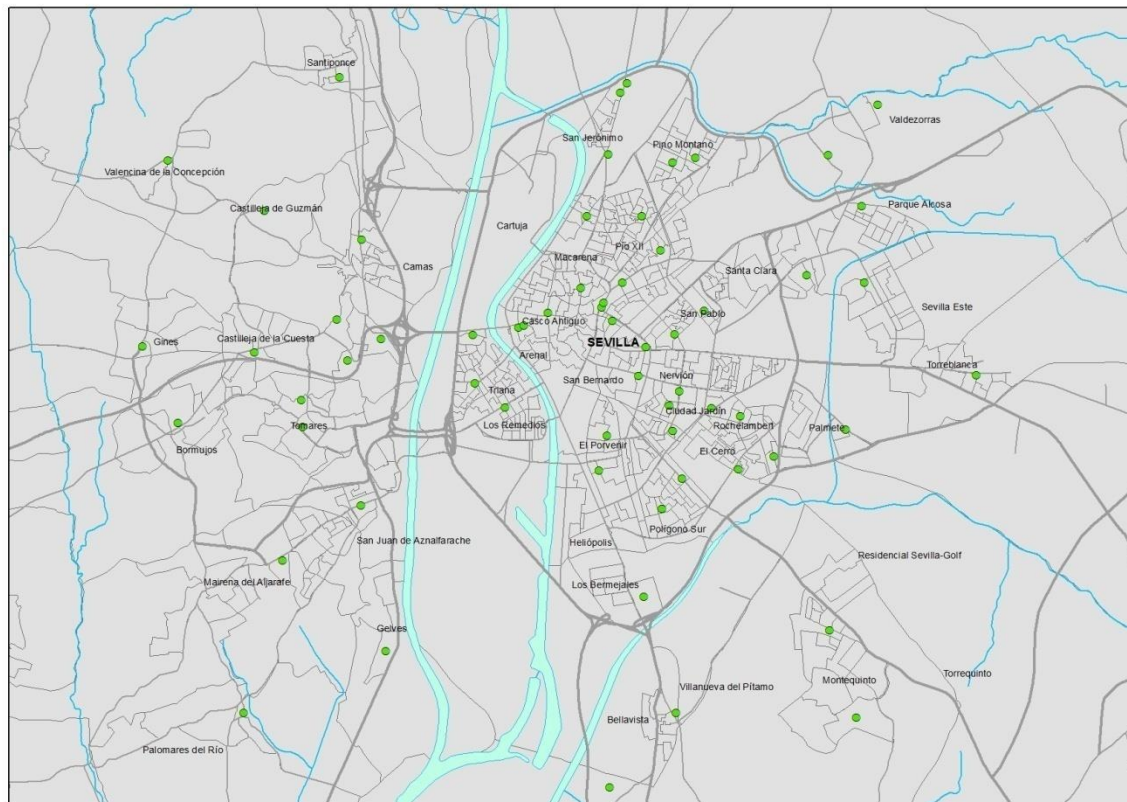
OVERVIEW

Regional official statistics are required to provide:

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

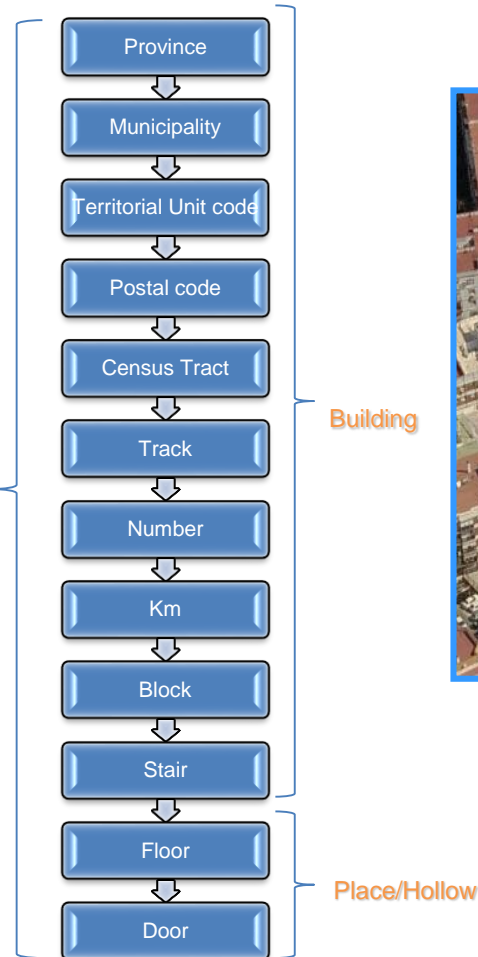
Data availability from multiple sources:

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

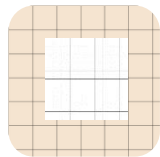
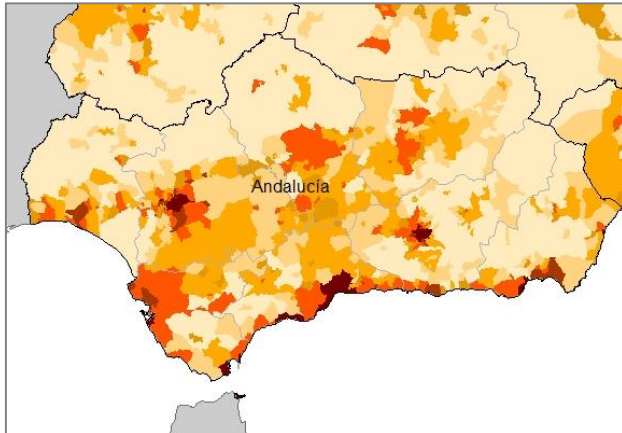
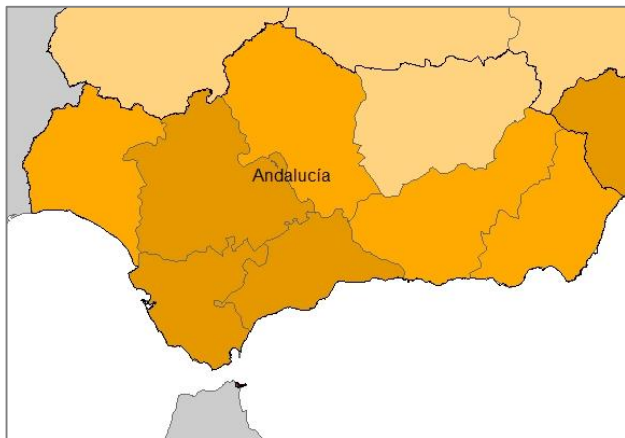
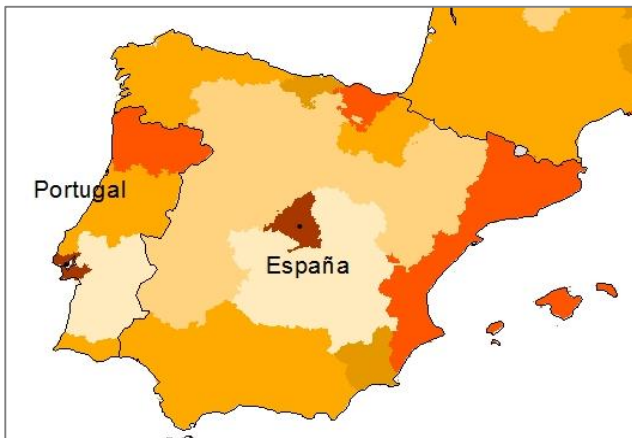
Official statistics usually reported according to a hierarchical system of administrative units

Modifiable areal unit

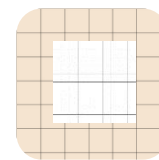
Nowadays regional planning and managing relies on municipal information or census tract data at best

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

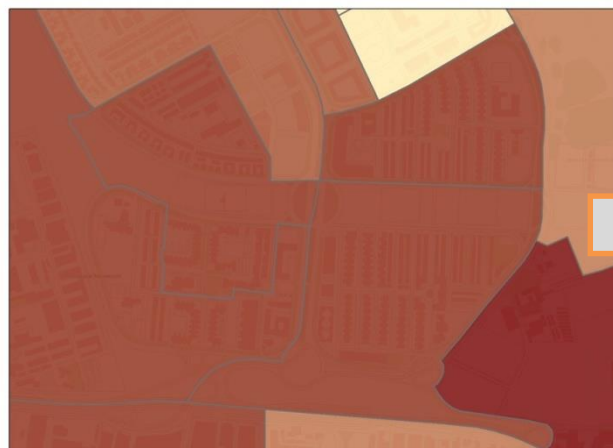
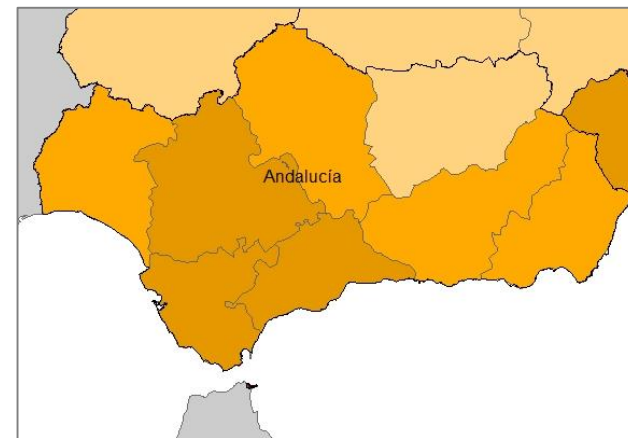
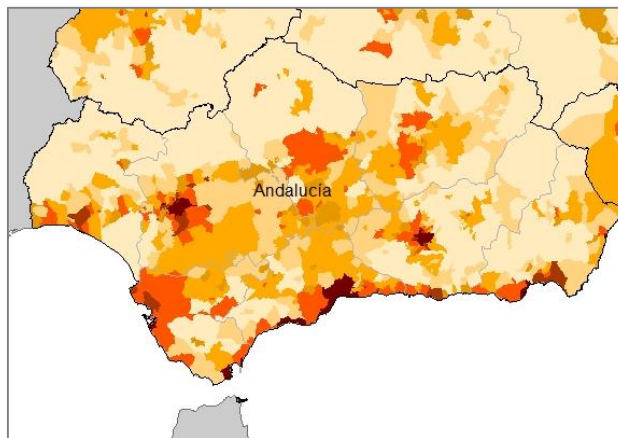
Equal size allowing for easy
comparison

Stability over time

Easy integration with other scientific
data

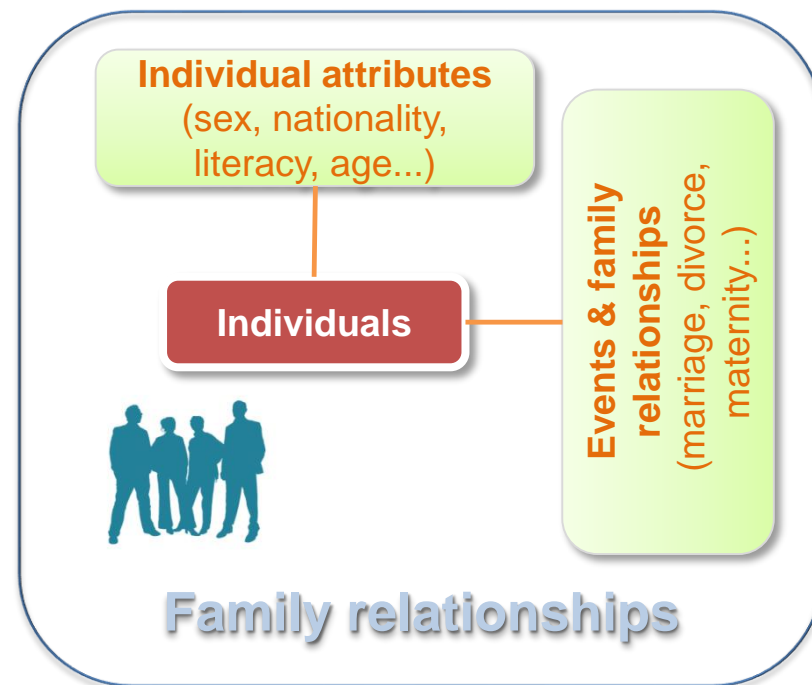
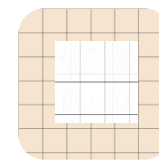
Grids can be built hierarchically in
terms of cell size, thus matching the
study area

Allow for spatial modelling techniques

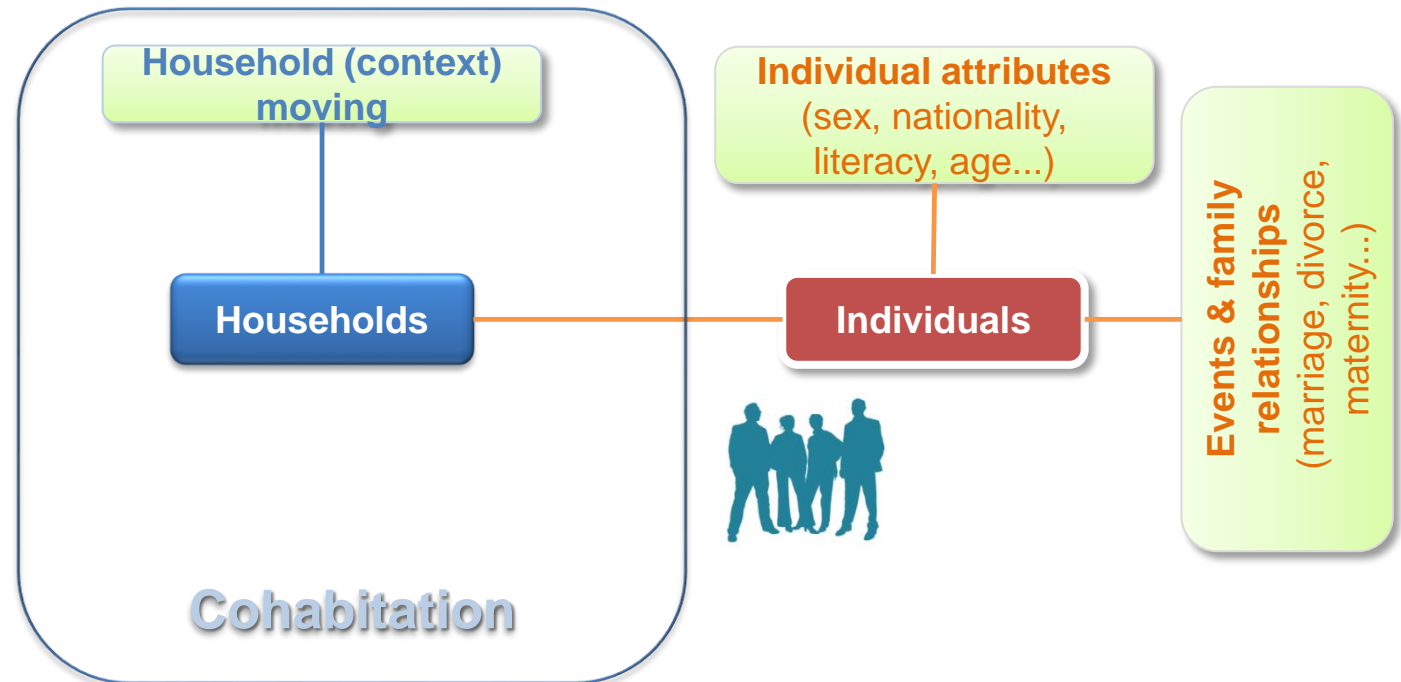
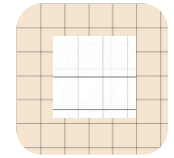


**High geographically detailed info, independent from administrative
boundaries crucial for regional and local administration**

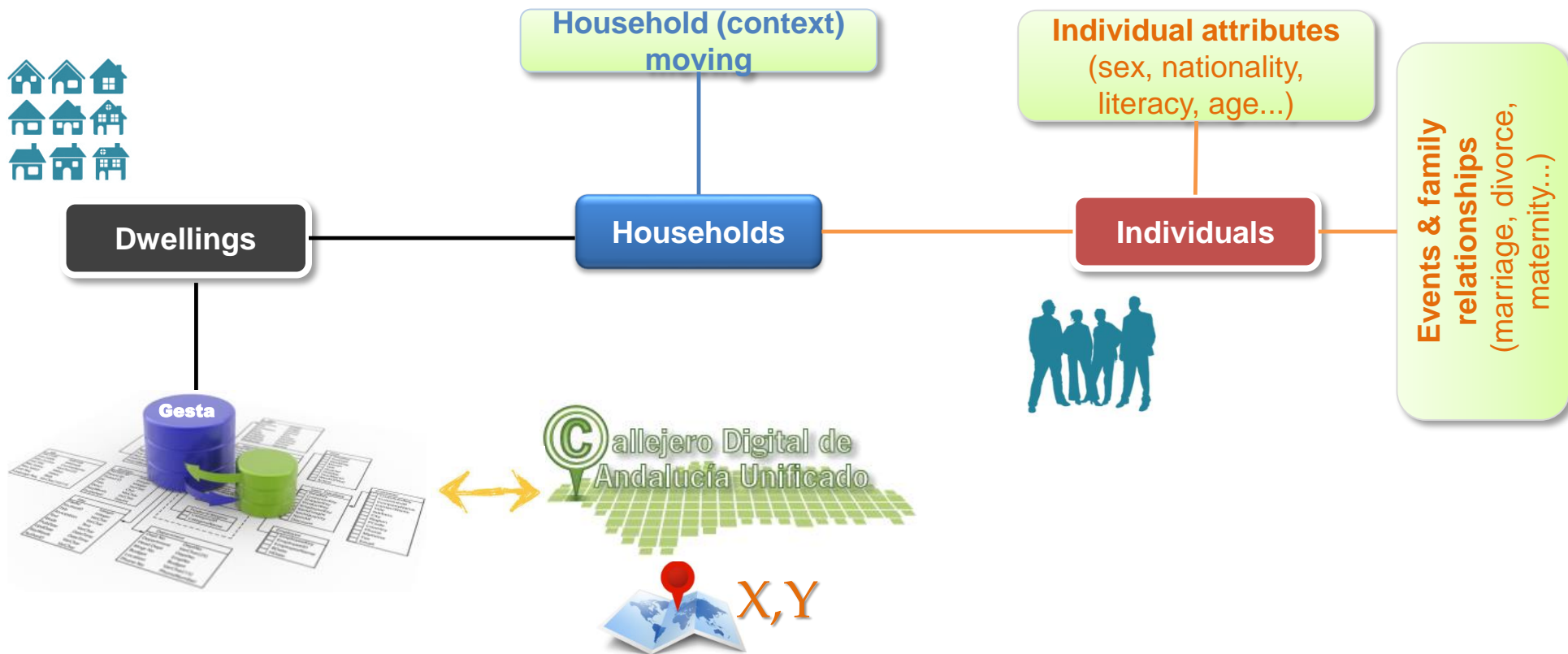
GRID PILOT PROJECT



GRID PILOT PROJECT



GRID PILOT PROJECT



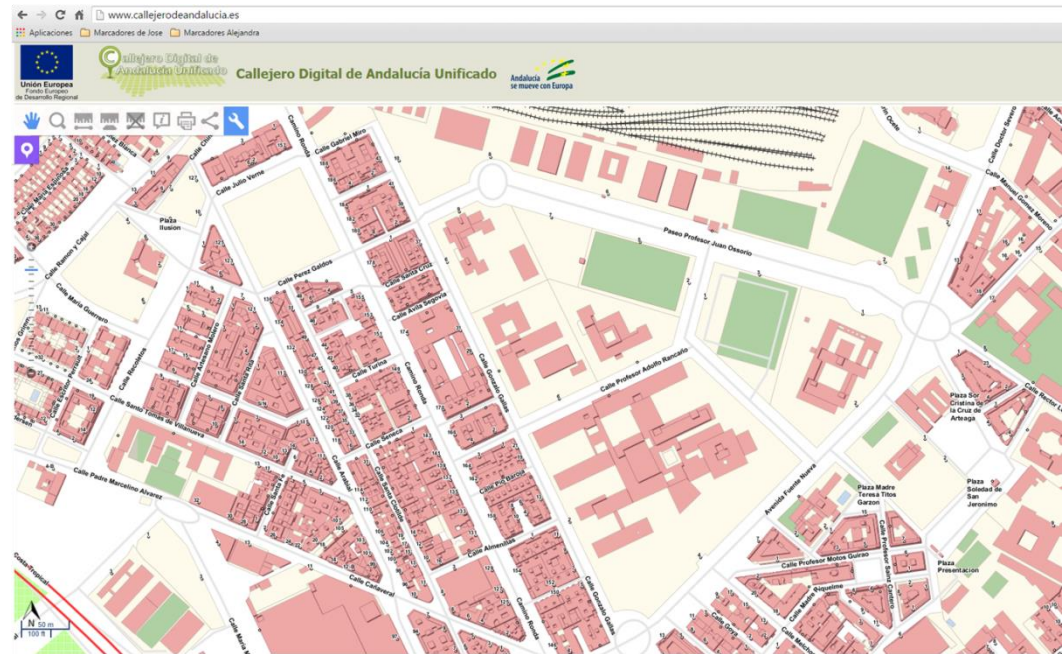
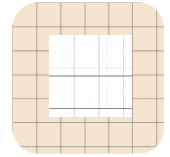
GRID PILOT PROJECT

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



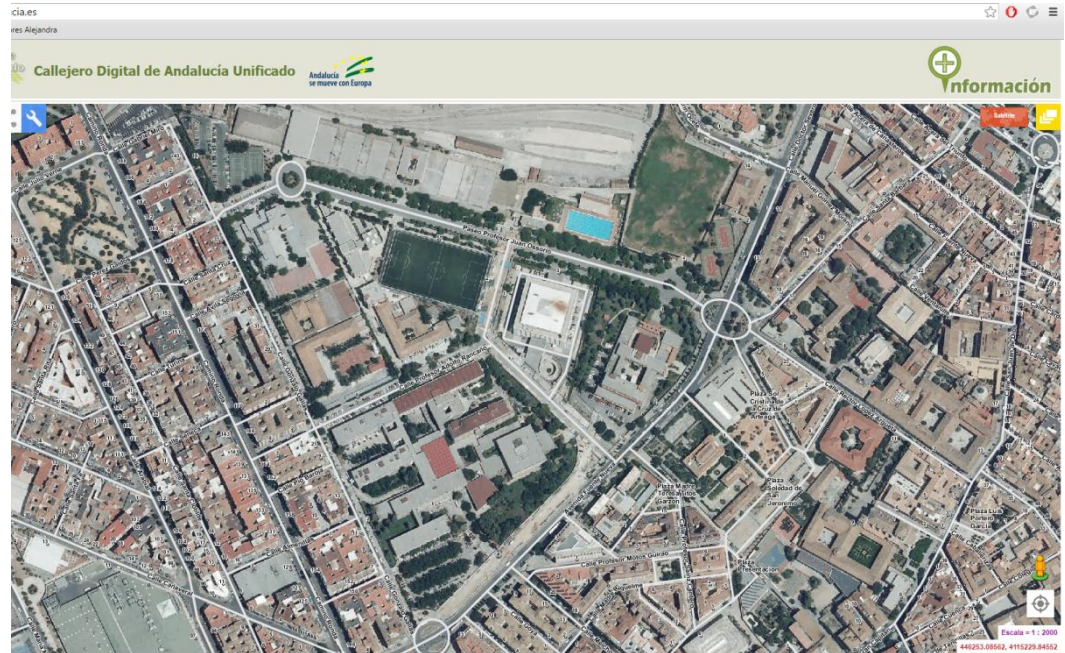
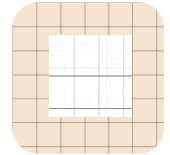
GRID PILOT PROJECT

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



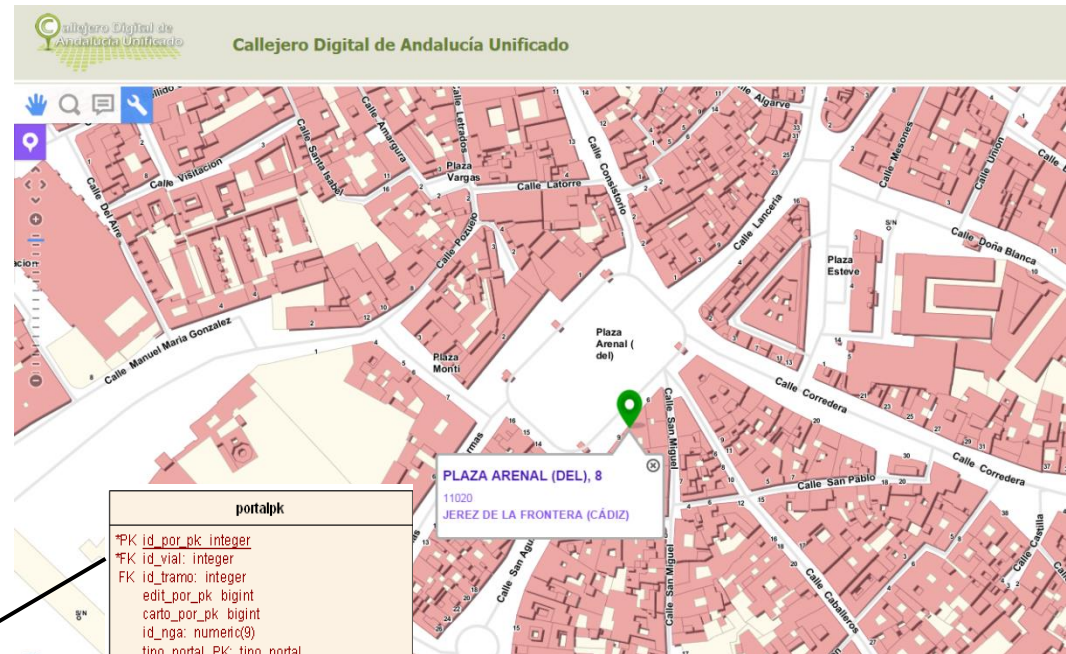
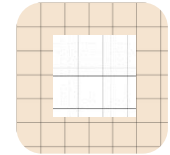
GRID PILOT PROJECT

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



vial	
*PK	id_vial: integer
	ine_vial: varchar(10)
	dgc_vial: integer
	edit_vial: bigint
	carto_vial: bigint
*FK	id_tipo_vial: integer
*	nom_vial: varchar(100)
	nom_normalizado: varchar(100)
	sobrenombre: varchar(100)
*	acceso: acceso
	competencia: competencia
*	geom: geometry
*	fuelle: varchar(50)
*	fecha_alta: date
*	fecha_aceptacion_alta: date
*	fecha_real_alta: date
	id_geda: integer
+ «PK» PK_vial(integer)	
+ «FK» FK_vial_tipo_vial(integer)	
+ «unique» UQ_vial_id_vial(integer)	
+ «unique» UQ_vial_ine_vial(varchar)	

portalpk	
*PK	id_por_pk: integer
	id_vial: integer
FK	id_tramo: integer
	edit_por_pk: bigint
	carto_por_pk: bigint
	idnga: numeric(9)
	tipo_portal_PK: tipo_portal
*	num_por_desde: integer
	num_por_hasta: integer
	ext_desde: varchar(20)
	ext_hasta: varchar(20)
	bxt_app: varchar(100)
	bloque: varchar(20)
	portal: varchar(20)
	escalera: varchar(20)
	refcatparc: varchar(50)
	geom: geometry
	fuelle: varchar(50)
FK	id_nucleo: integer
	cod_nucleo: varchar(50)
FK	id_sc: integer
	cod_sc: varchar(3)
FK	id_cp: integer
	cod_cp: varchar(5)
FK	id_agrup: integer
	cod_agrup: varchar(50)
*	fecha_alta: date
	fecha_aceptacion_alta: date
	fecha_real_alta: date
	id_geda: integer
+ «PK» PK_portalpk(integer)	
+ «FK» FK_portalpk_agrupacion(integer)	
+ «FK» FK_portalpk_tramo_vial(integer, integer)	
+ «FK» FK_portalpk_nucleo(integer)	
+ «FK» FK_portalpk_seccion_censal(integer)	
+ «FK» FK_portalpk_codigo_postal(integer)	
+ «unique» UQ_portalpk_id_por_pk(integer)	

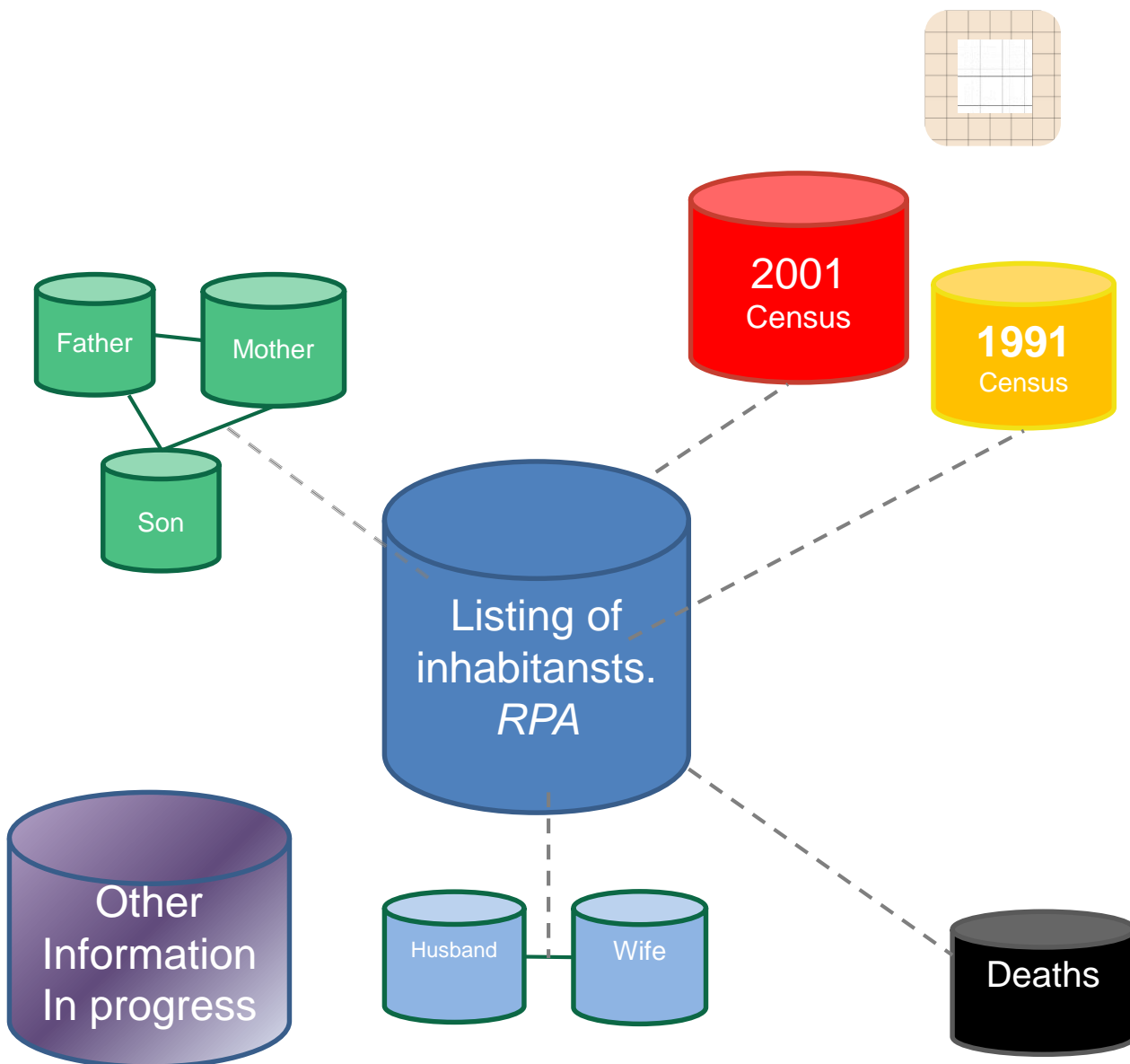
GRID PILOT PROJECT

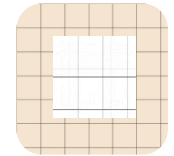
Sources & Methods

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

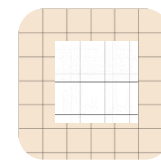
Unified Digital Street Map of Andalusia
(Spatial info)

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

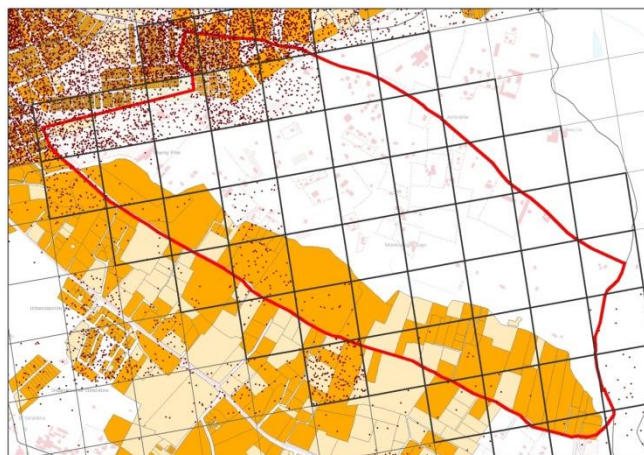
Top-down approach

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

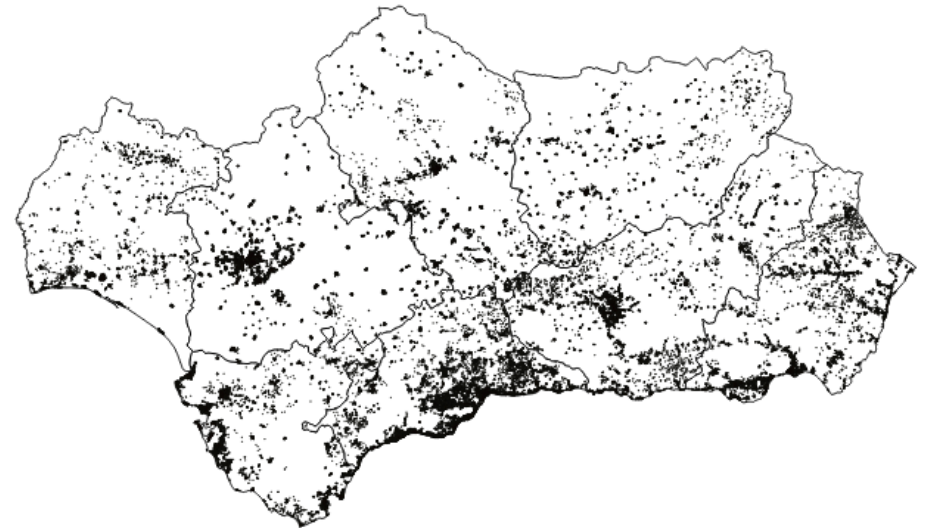
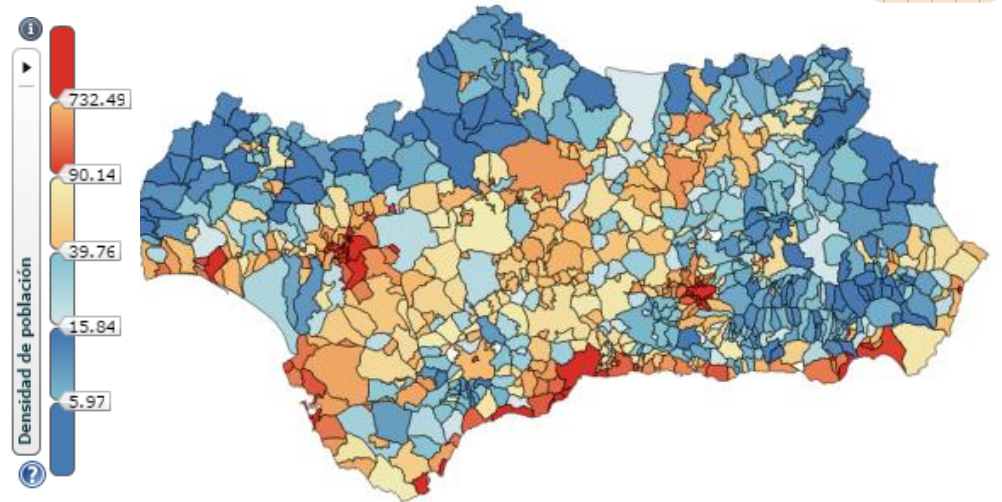
Bottom-up approach

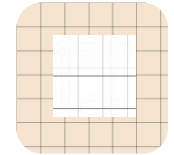
A second edition of 2013 Population grid was elaborated including 2011 Census building data.

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

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

48,187 inhabited cells (3.4% of cells)





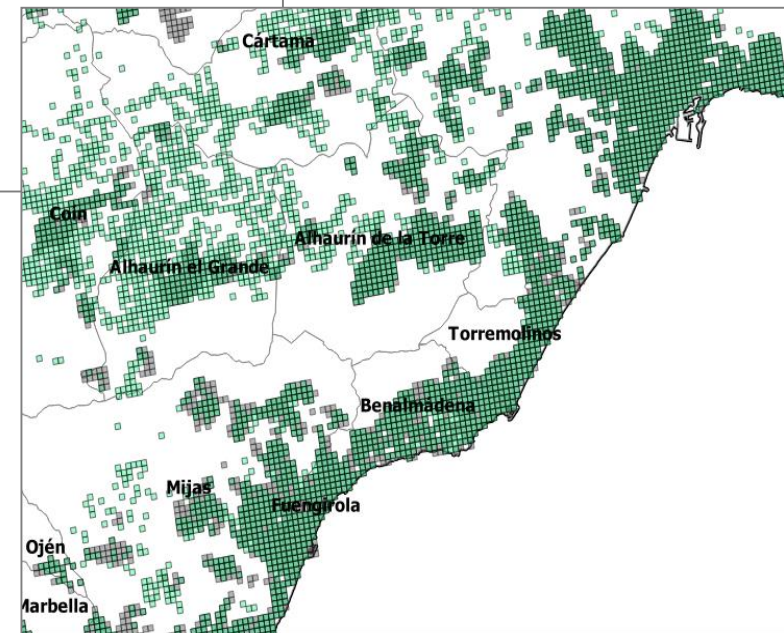
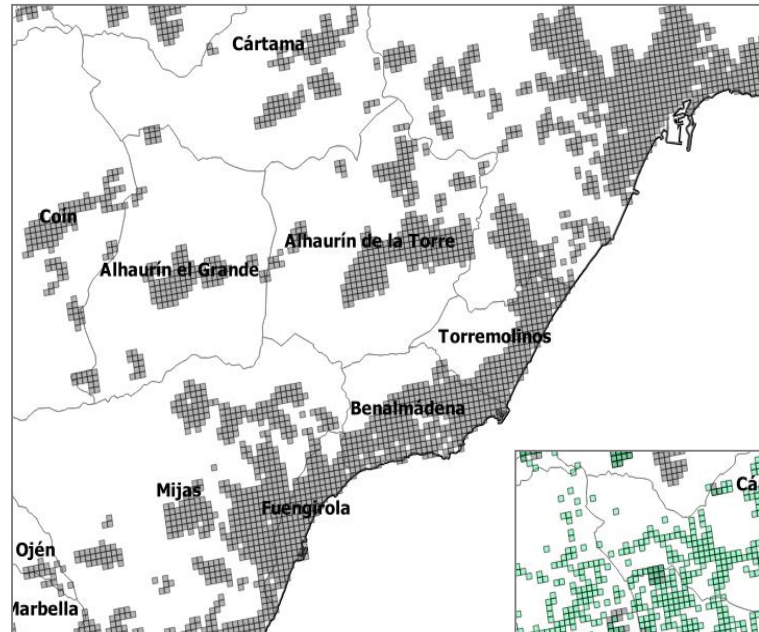
GRID PILOT PROJECT

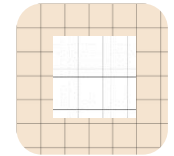
Lessons learnt

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

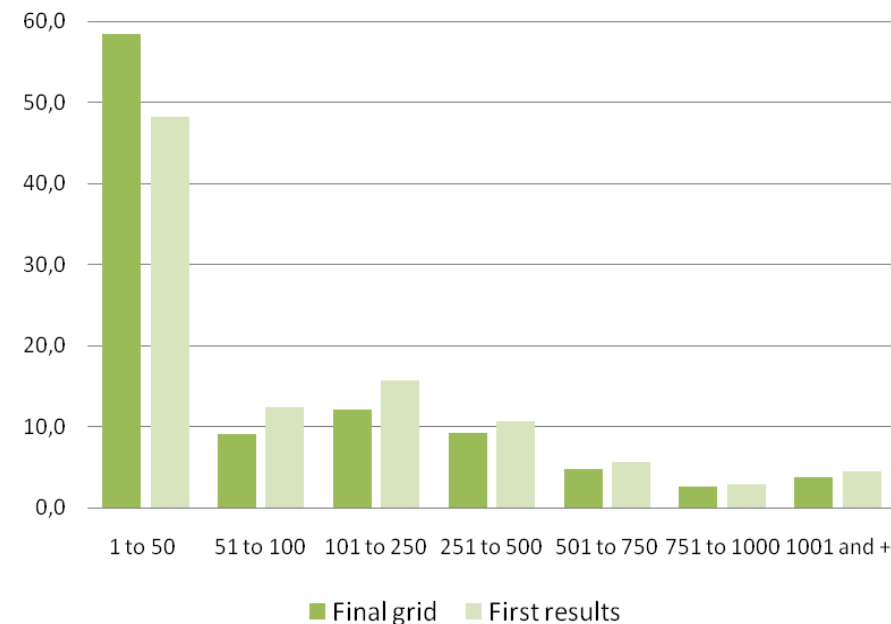
Modelling is helpful but ancillary variables are crucial

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

Distribution of cells by size (inhabitants)



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

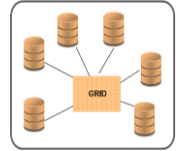
Further analyses & grid indicators

Bicycle lane accessibility

Mortality ratio by grid cells



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



You are in: Start - Georeferencing

Institute | Statistic | Georeferencing

What information are you looking for?

Search

The information displayed on these maps comes from Andalusian Population Longitudinal Database (BDLPA) as of January 1, 2013 and has been georeferenced based on the localization of the address where every inhabitant of Andalusia lives.

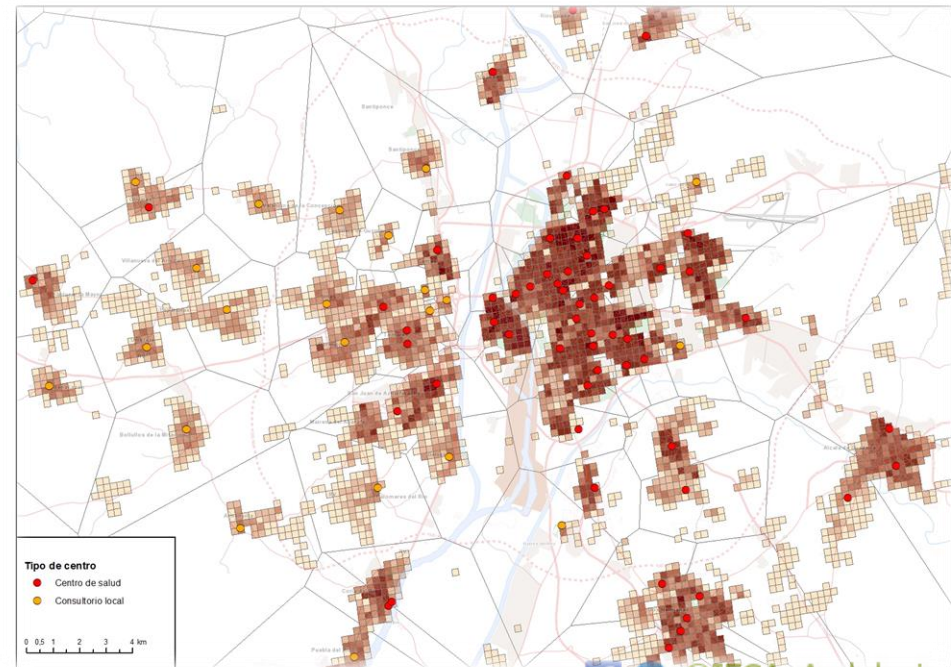
To facilitate the representation of the information and to preserve statistical confidentiality a regular 250 m grid cell has been drawn, where the full information corresponding to each cell has been added. The information that could not be georeferenced has been estimated using spatial analysis techniques.

- Organization: Institute of Statistics and Cartography of Andalusia.
- Scope: Statistics section.

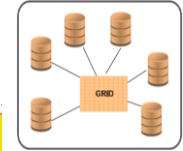
Updates:

Spatial Distribution of the Population in Andalusia

- Informative note
- Map Viewer*
 - Population
 - Total population
 - Population by nationality
 - Population by age groups
 - Persons Affiliated to the Social Security
 - Total Affiliation
 - Affiliation according employment relationship
- Download information as spatial data (DERA)
- User Manual of Map Viewer



Área de influencia de 300m del carril bici
Carril bici



USABILITY & DATA DEMAND

Administrative planning & evaluation

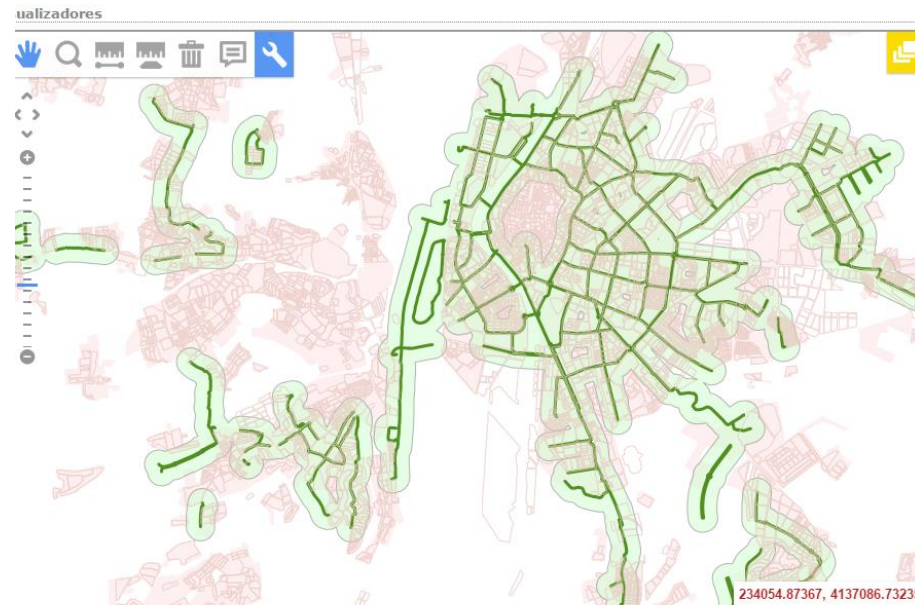
Education department uses grid data in order to forecast schooling demand

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

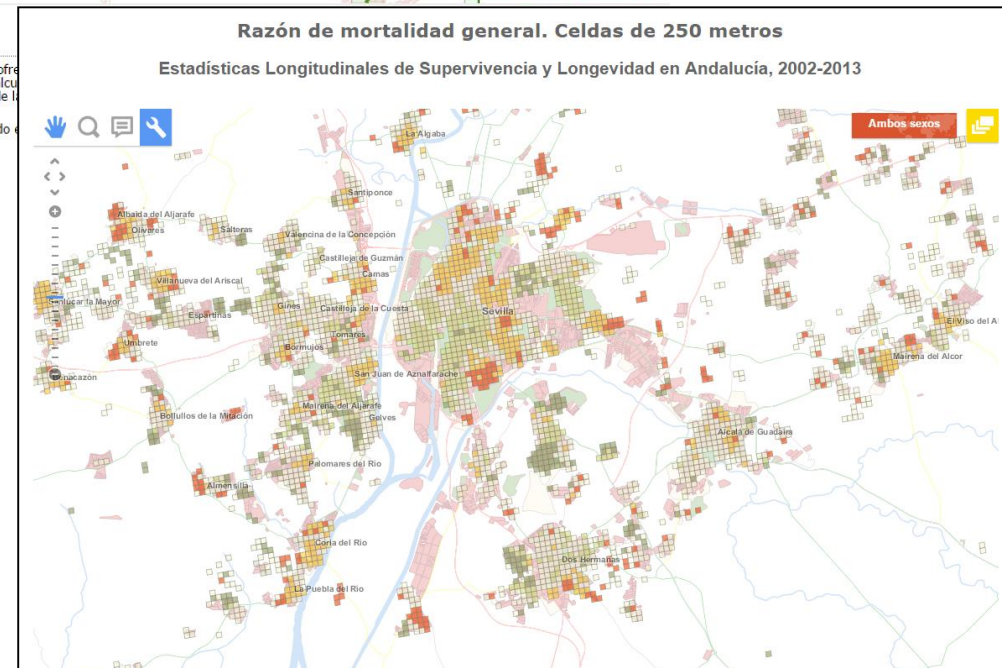
Further analyses & grid indicators

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)



Metodología
El artículo de la Cartografía de carriles bici ofrece la metodología de la IECA, se ha calculado el ratio de las áreas de influencia, a partir de la información, se ha calculado

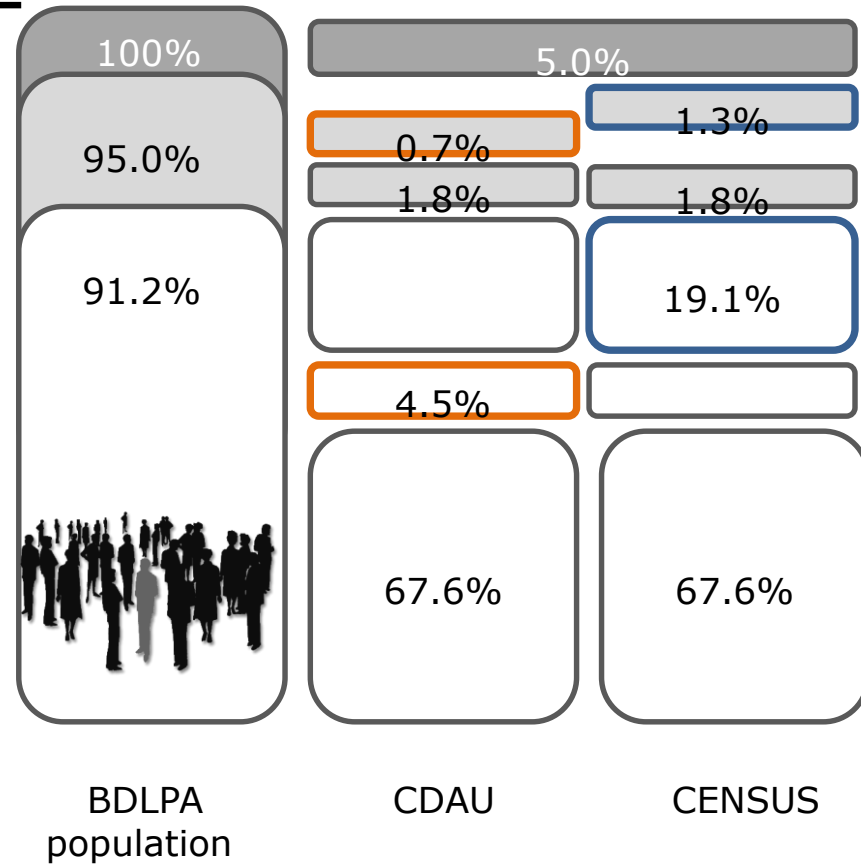
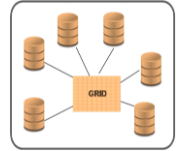


GRID STANDARD AS REFERENCE

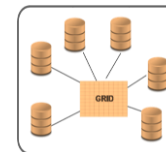
Current projects

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%



- ☐ Exact georeferentiation
- ☐ Approximate georeferentiation
- ☐ Non georeferenced



GRID STANDARD AS REFERENCE

Current projects

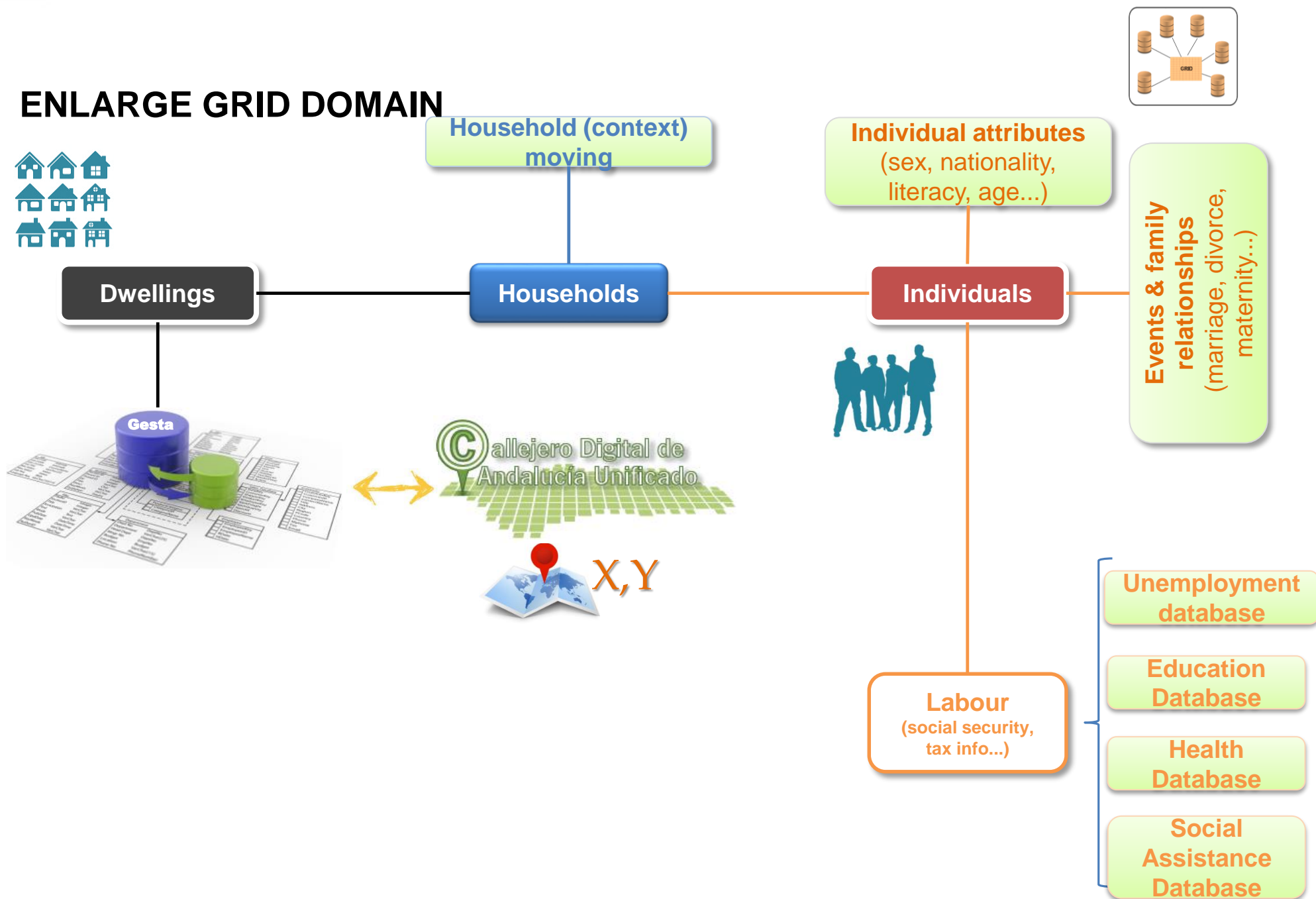
Georeferentiation of 2014 & 2015 population.

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



ENLARGE GRID DOMAIN

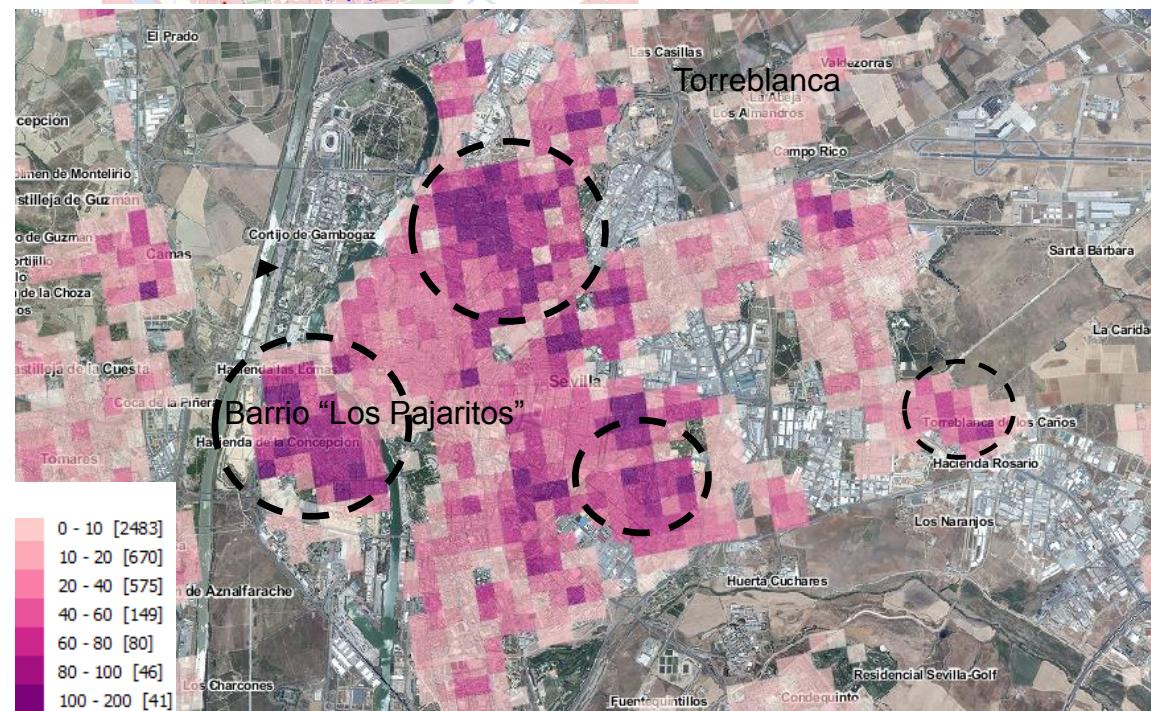
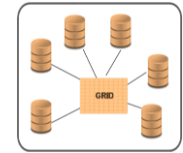
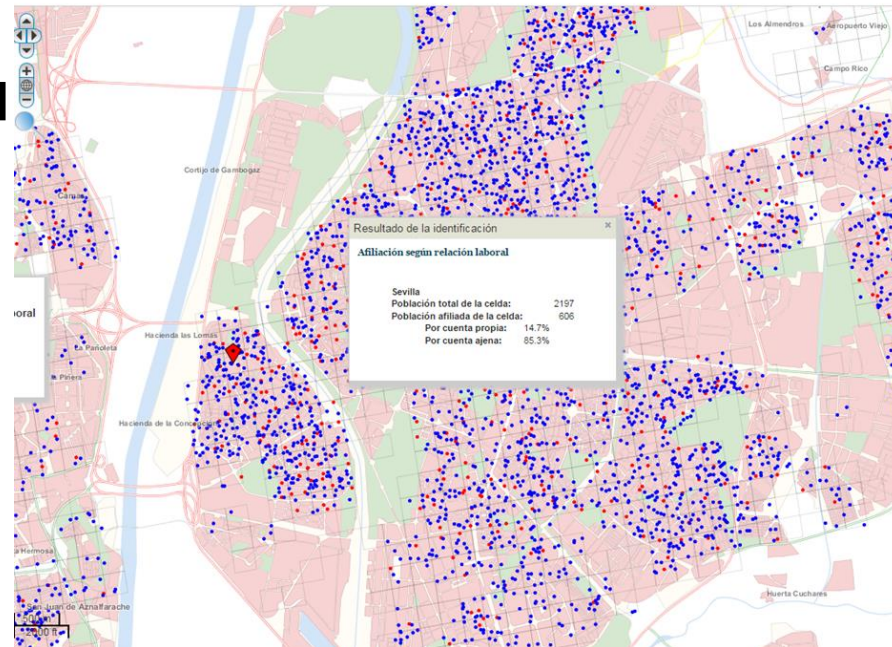
Current projects

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

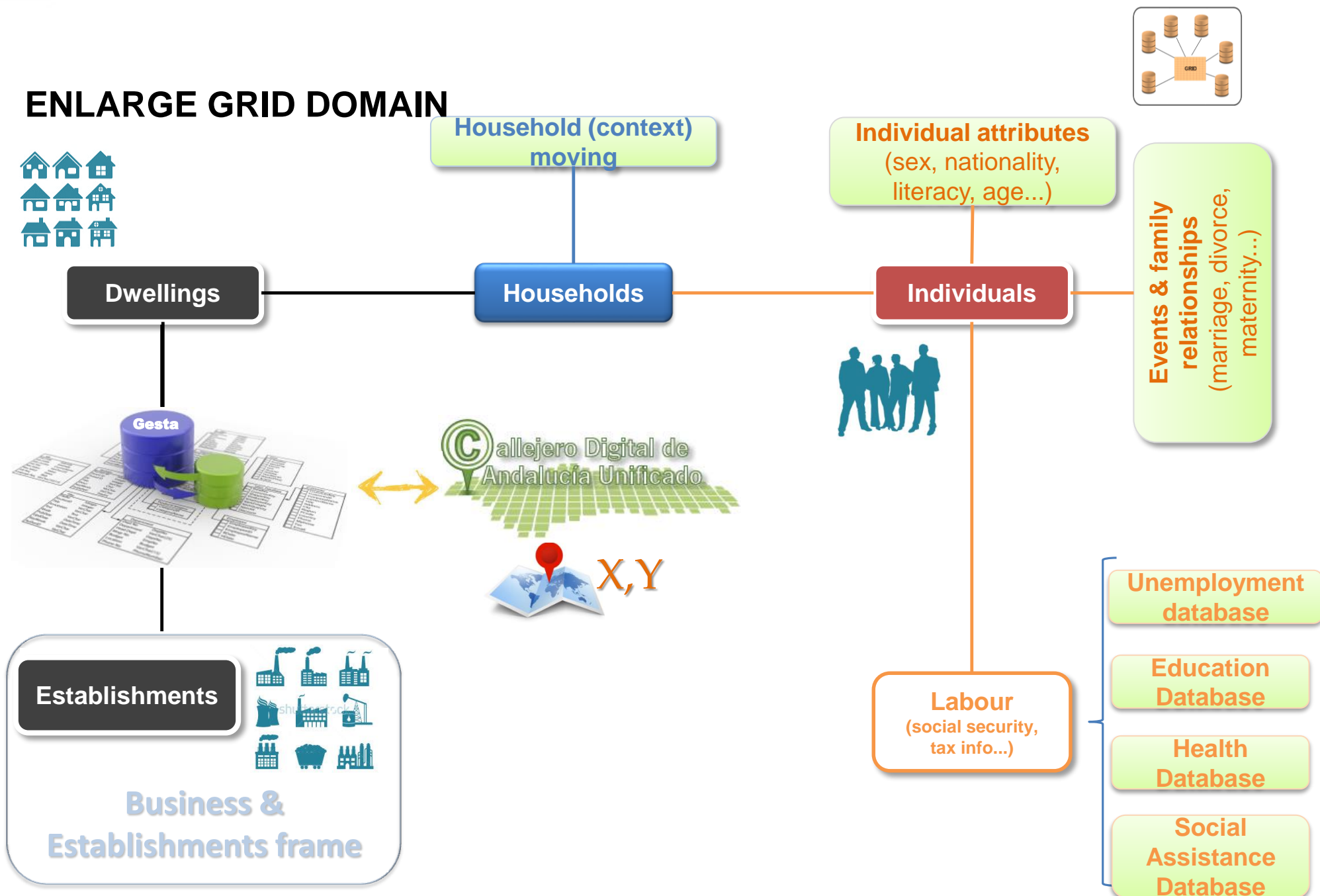
Labour status, salaried, self-employed from social security records linked to 2013 population

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

Variable: Afiliación a la Seguridad Social según relación laboral
mas: Afiliación total a la Seguridad Social **AFILIACIÓN A LA SEGURIDAD SOCIAL SEGÚN RELACIÓN LABORAL**



ENLARGE GRID DOMAIN



ENLARGE GRID DOMAIN

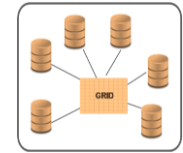
Current projects

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 georeferentiation 28%
georeferentiation through neighbouring area or street center is highly disturbing in establishment location

Establishment characteristics & location are “visible”



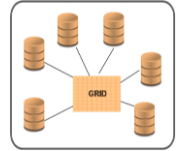
ENLARGE GRID DOMAIN


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




Instituto de Estadística y Cartografía de Andalucía
CONSEJERÍA DE ECONOMÍA, INNOVACIÓN, CIENCIA Y EMPLEO

Conectado como admin GEOEST_DIR14_EPLA13_20150721 EST-CAJA RURAL DE GRANADA SDAD COOP DE CREDITO-CALLE DON BOSCO 2-GRANADA(18)-151-GEO Versión 1.4 Terminar sesión

[Inicio Encuesta](#) | [Cuestionario](#) | Bloque I

BLOQUE 1 - REVISIÓN DE ESTABLECIMIENTOS GEOLOCALIZADOS

APDEC.- Aprox. postal buena

UBIEC1.- Bien ubicado

MIE1.- Método Identificación

EPCDAU1.- Existe portal CDAU

MNEE1.- No existe

Geolocalización. En este proceso se comprueba que la dirección del Directorio coincide con la encontrada en la base de datos del Callejero (CDAU).

Geolocalización correcta. El proceso de enlace ha asignado un portal que coincide con exactitud con la aproximación postal que figura en el Directorio.

Geolocalización no correcta. El proceso de enlace ha podido asignar un portal que no coincide con exactitud con la aproximación postal que figura en el Directorio.

Por medios auxiliares se ha podido comprobar si esa localización coincide o no con la localización REAL de esa aproximación postal


¿Está la aproximación postal del Directorio geolocalizada correctamente?

Sí ☒

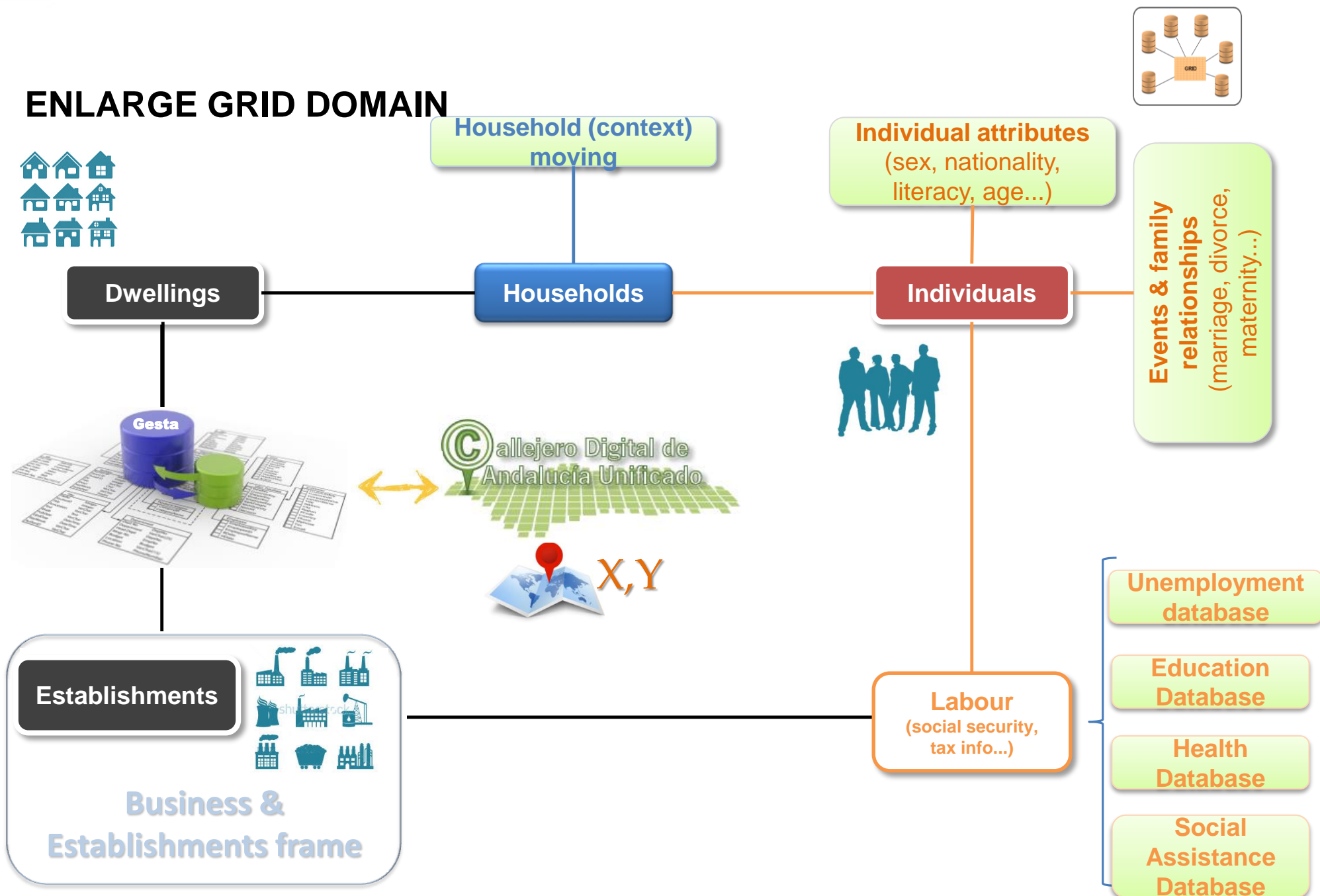
No ☐

 [Buscar razón social en Google](#)

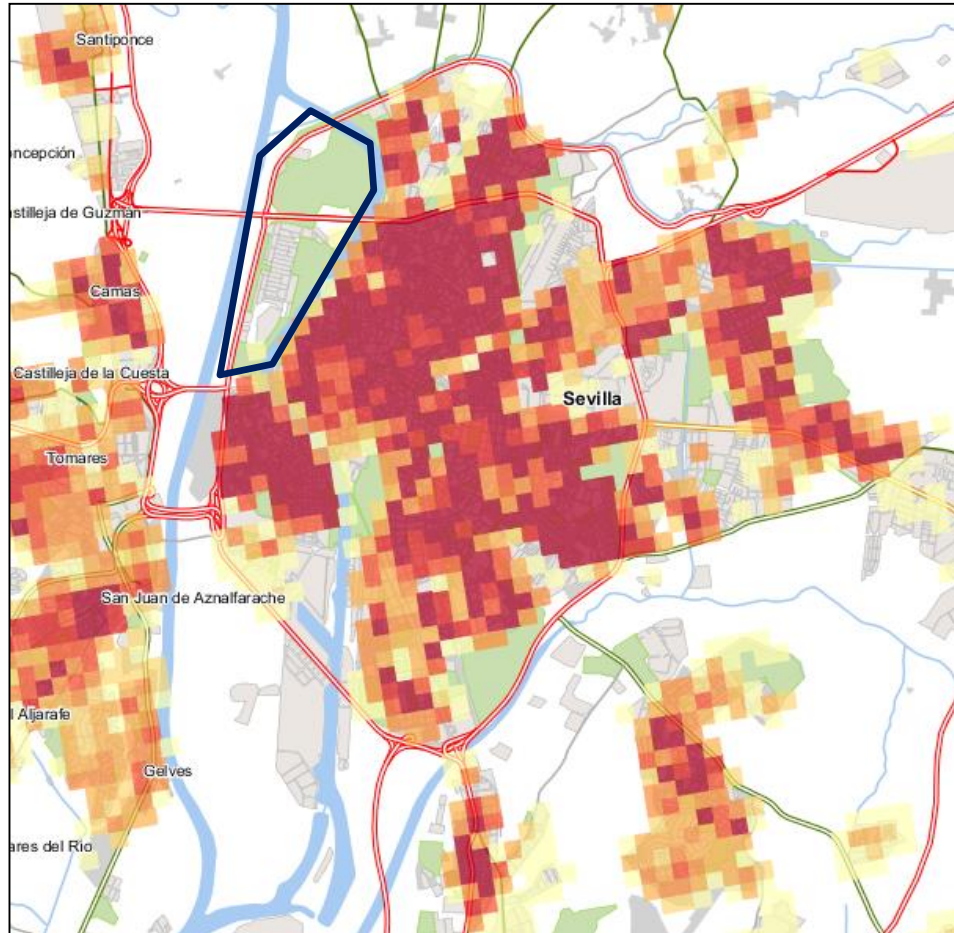
MAPEA CDAU



ENLARGE GRID DOMAIN



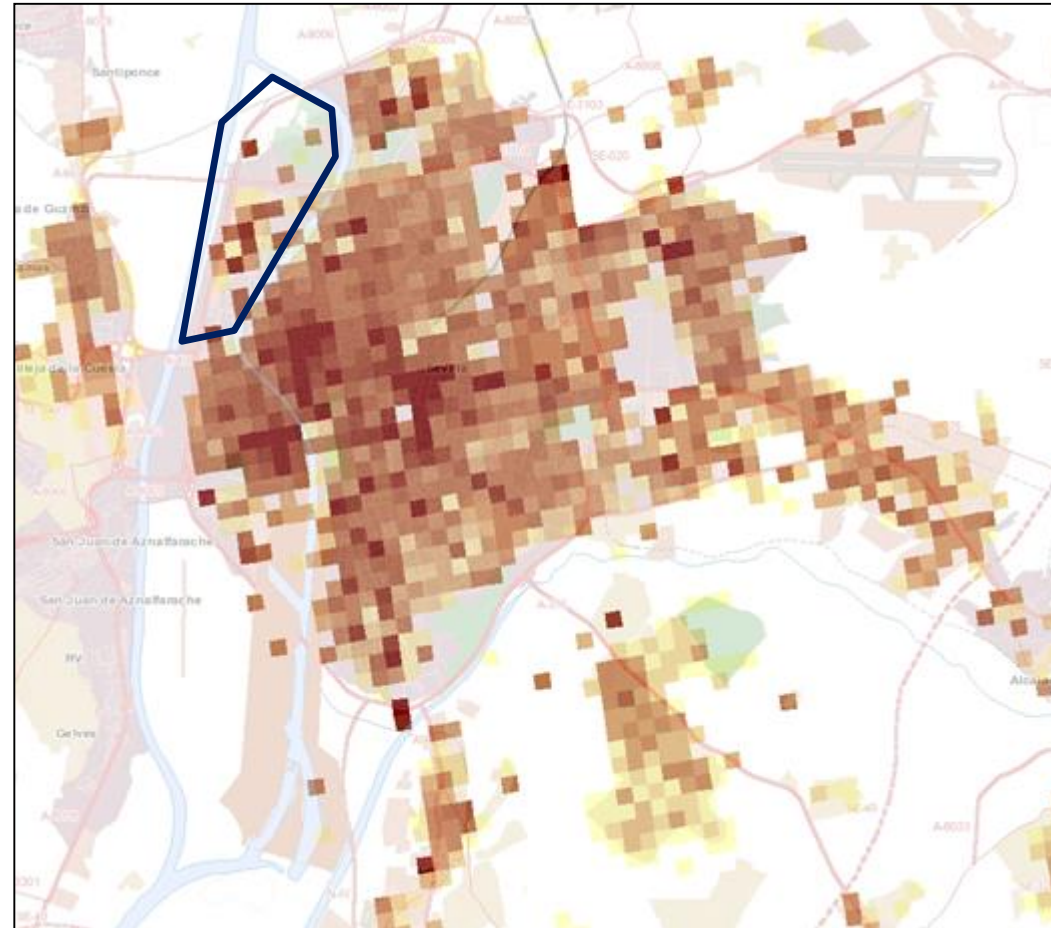
Population grid ("night grid")



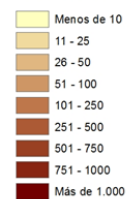
Número de habitantes



Establishments grid ("day grid")



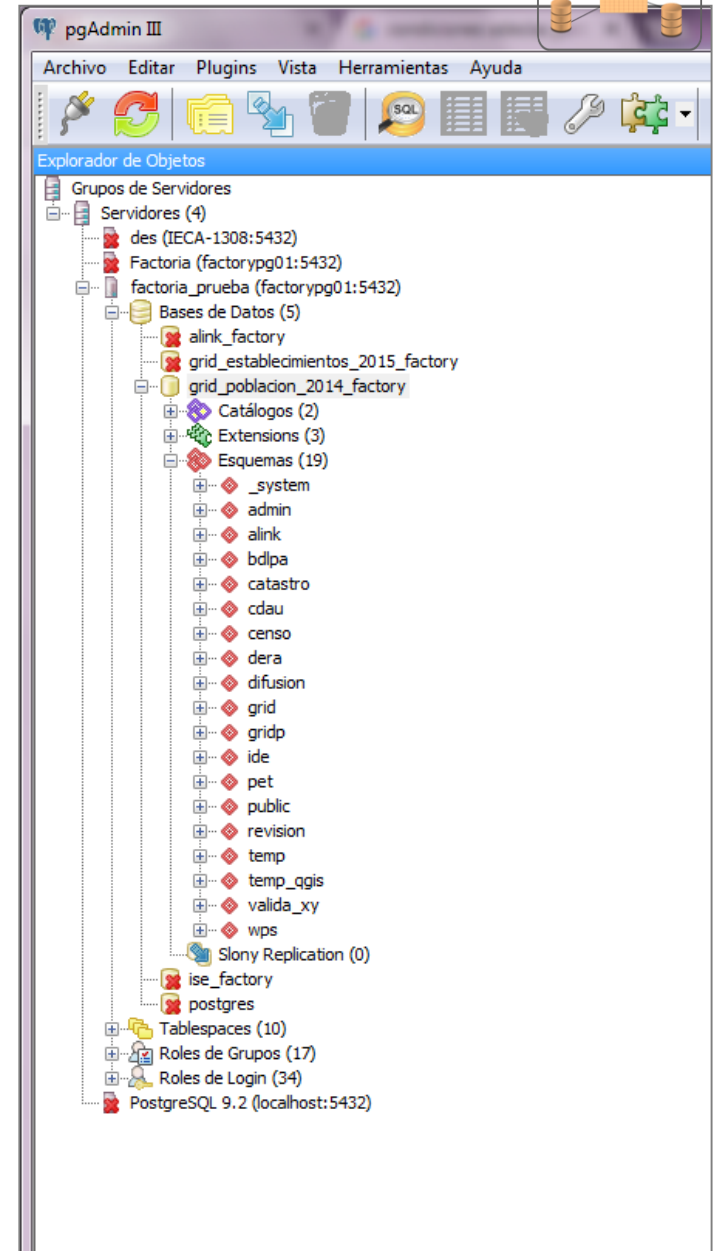
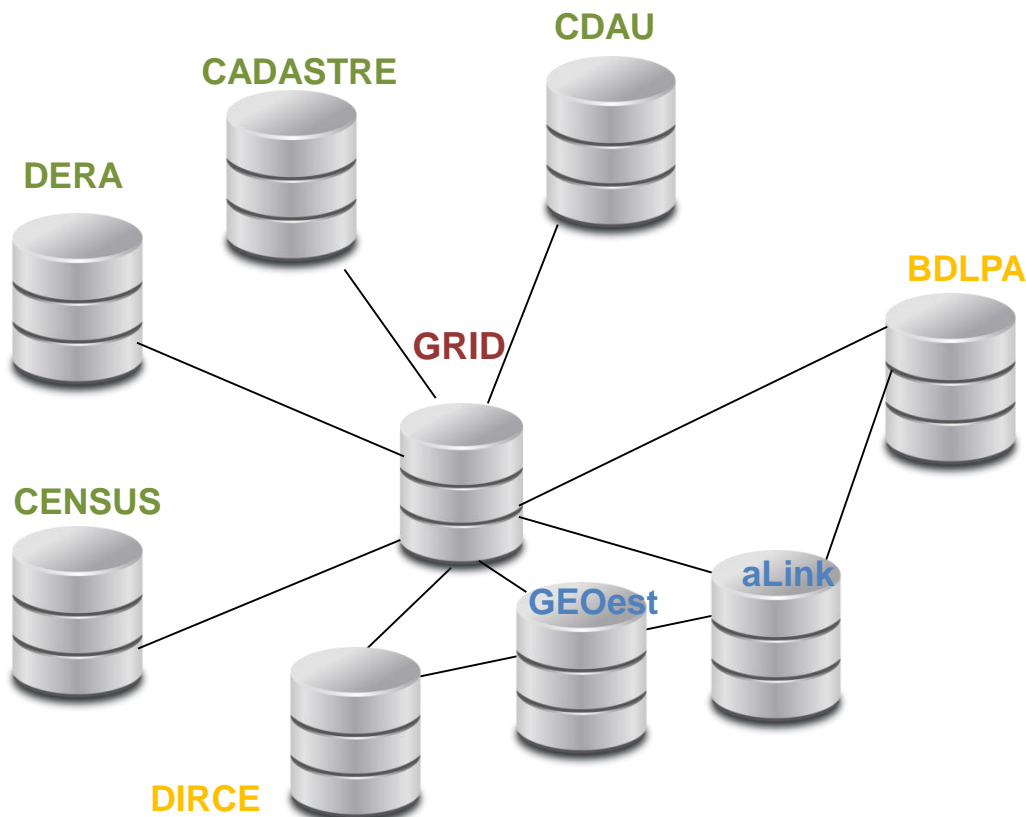
Empleo total



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