INSEE
INSTITUT NATIONAL
DE LA STATISTIQUE
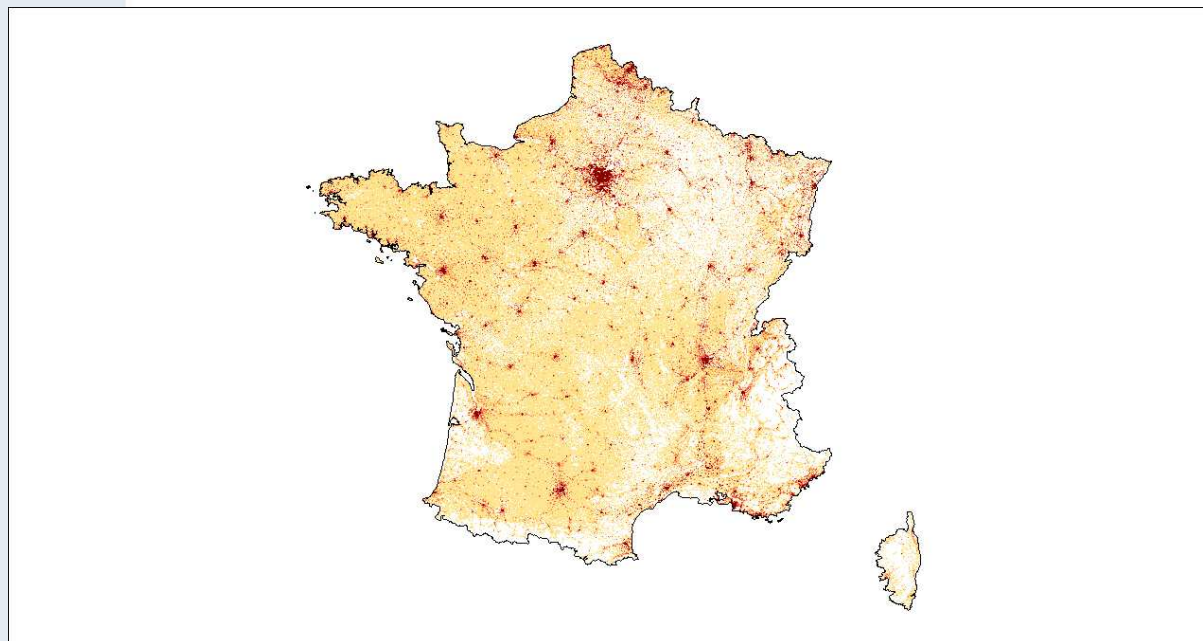ET DES ÉTUDES
ÉCONOMIQUES

Grided population data by Insee

Was it worth the effort?

Jean-Luc LIPATZ
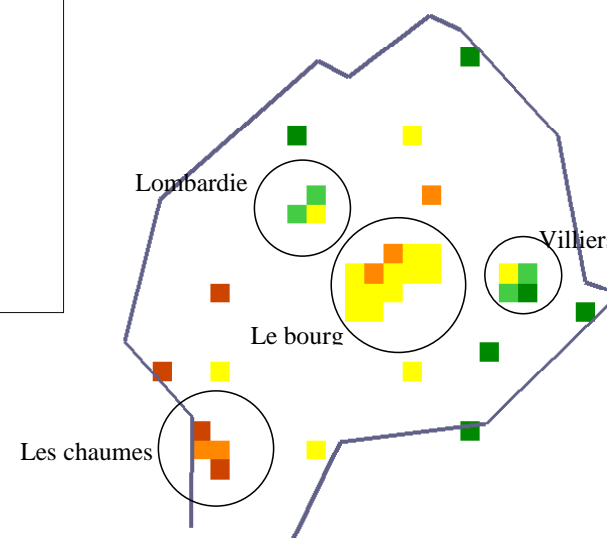**INSEE**

October 2011

The state of the art

http://www.efgs.info/data/france/grided-data-from-the-census

…for any municipality of France

…even for the smallest ones

No more disaggregation
Smaller grid cells
More variables

Census unemployment in 2009
In Curzay sur Vonne (465 inhabitants)

That was impossible mission
spatial units at their maximum resolution

Under 10 000 inh.
Not breakable in smaller chunks

Over 10 000 inh.
Spatially referenced
But sample is not spatially representative

Year of
data collection

|	Year
🟥	2005
🟩	2006
🟦	2007
🟪	2008
🟨	2009

⬜	Not collected
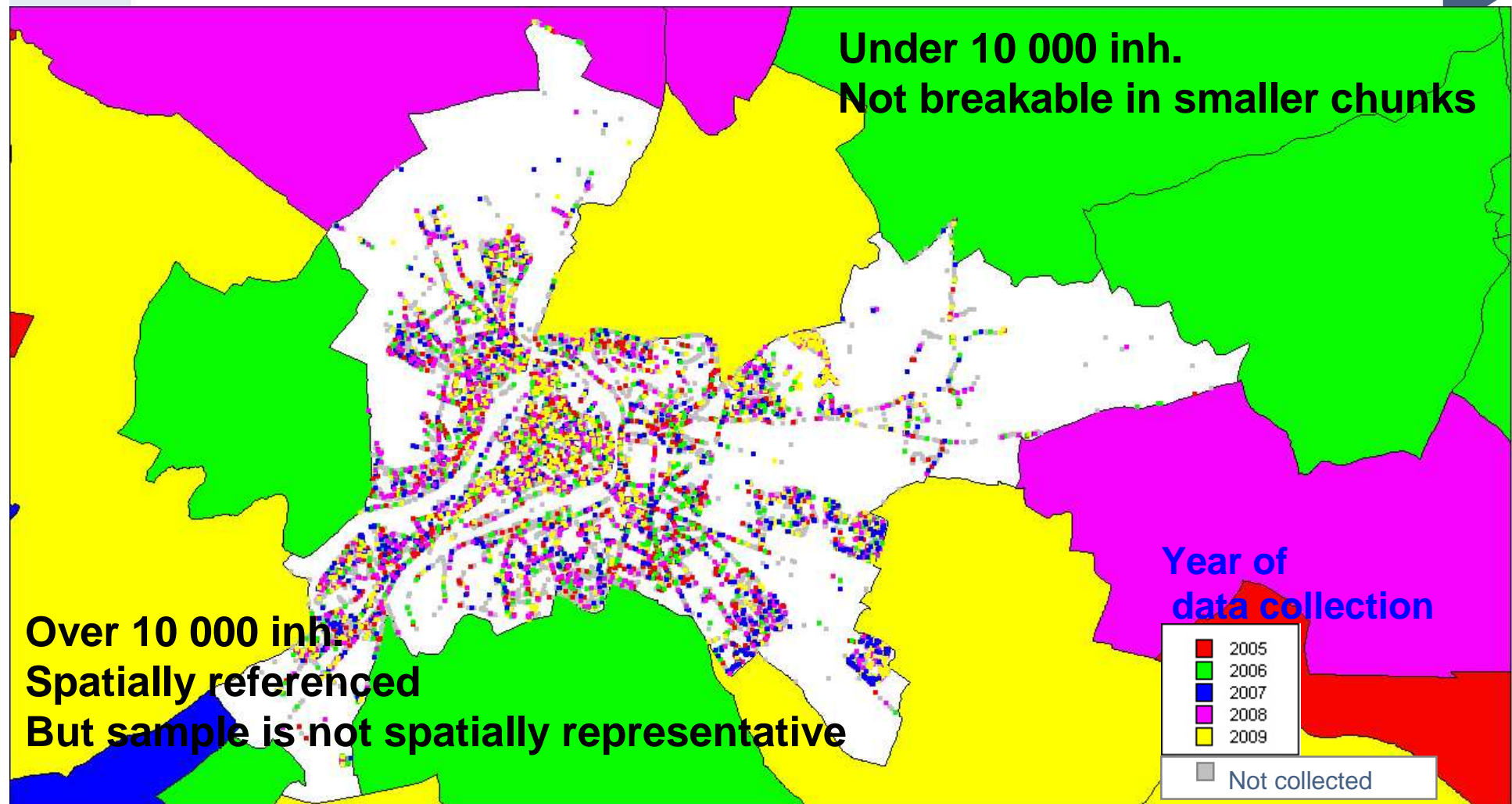
The milestones

› Coping with large municipalities :
EFGS Den Hague 2009

› Coping with small municipalities
part1 (administrative registers): EFGS Tallinn 2010

› Coping with small municipalities
part2 (census) : GISCO Luxemburg 2011

Was is worth the effort?
The map from the JRC…

Between the two

500 Gbytes of Internet download

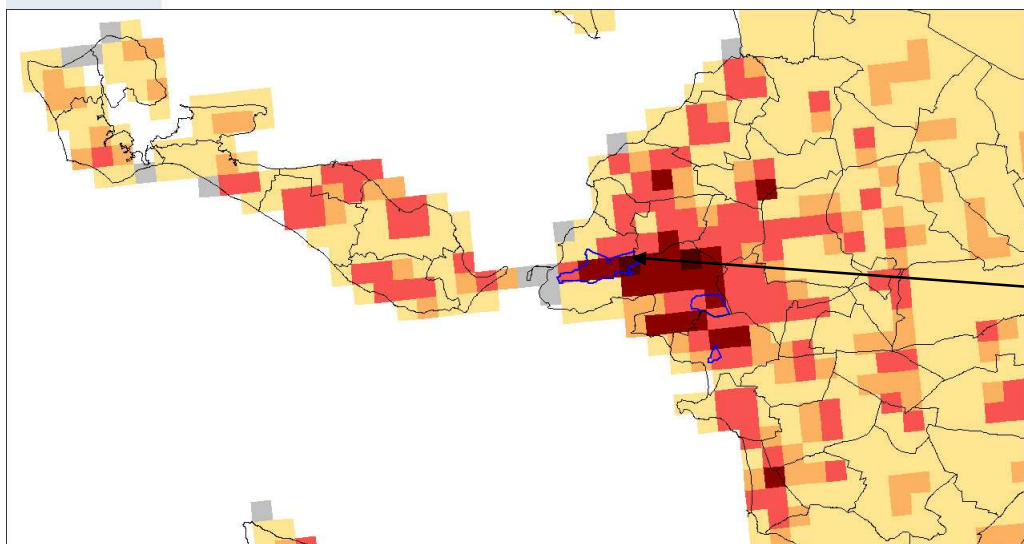700 000 files in the information system

300 hours of computing time

But visual messages are not very different

…and the one from the Insee
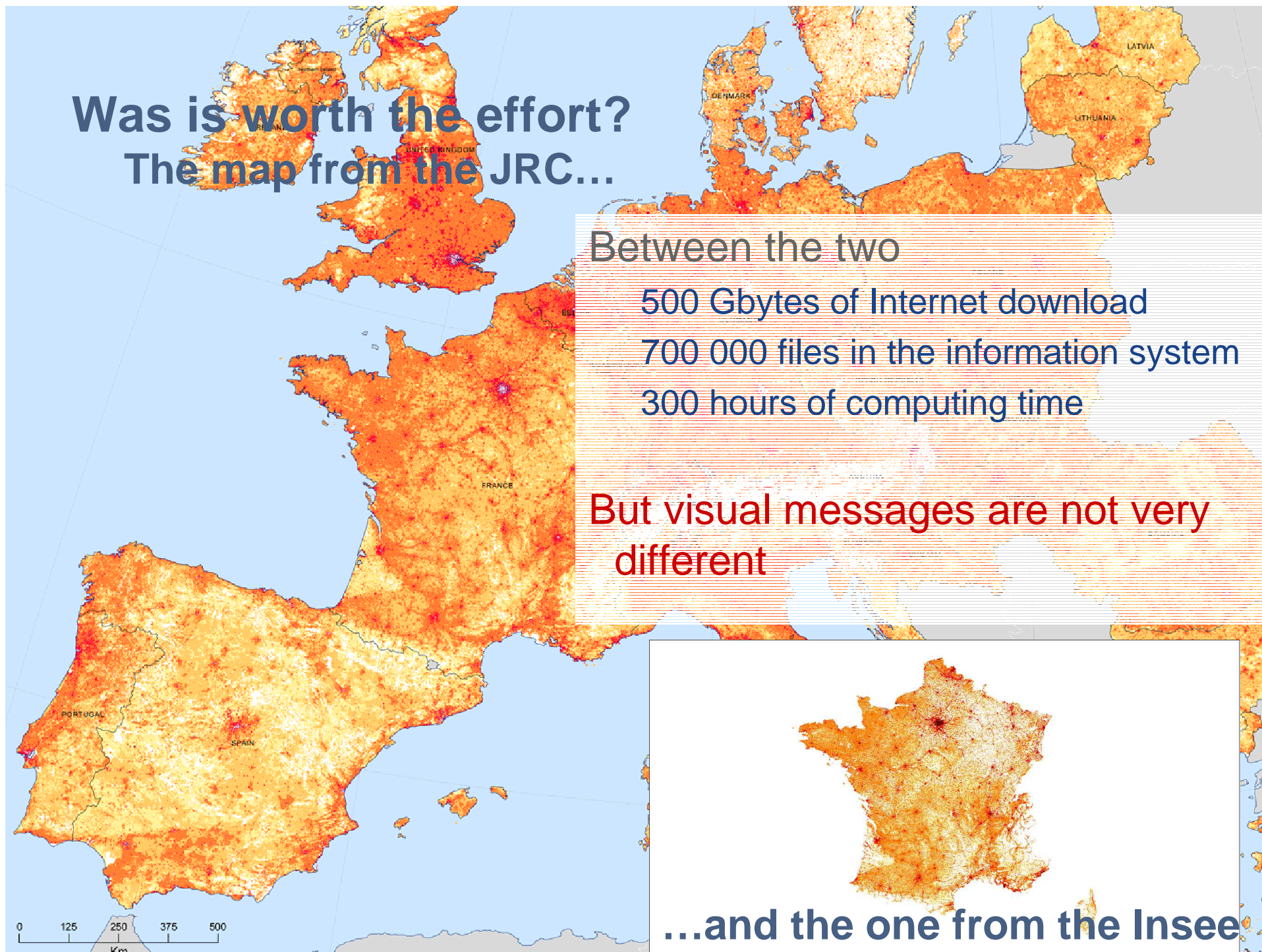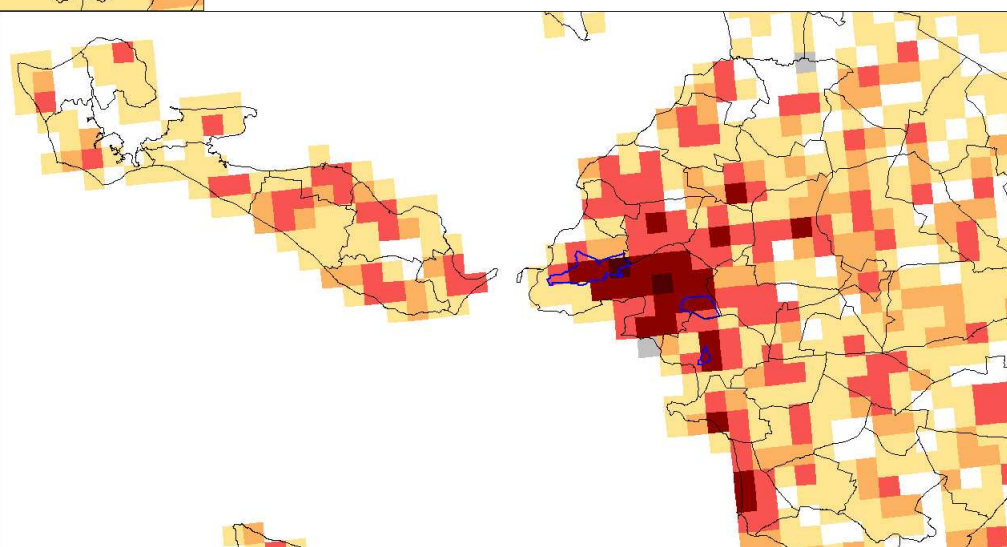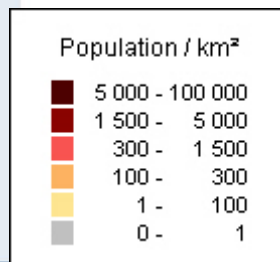
Zooming : similar pictures…

JRC 2006 estimation

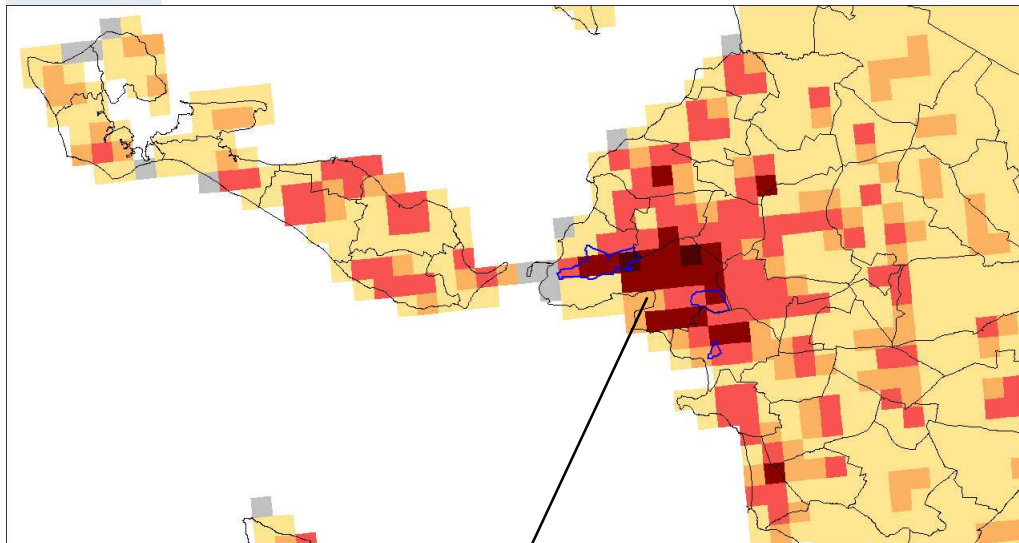**…but weak figures**

5 400 inh. estimated
9 800 inh. counted

Census 2007

Population / km²

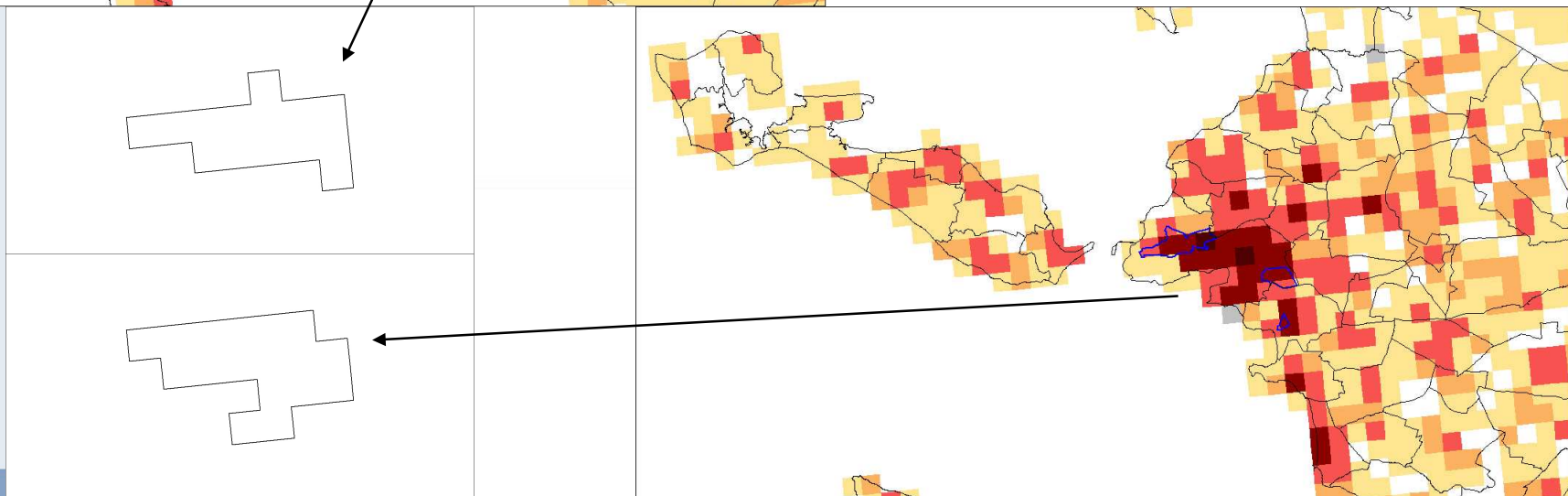|	
■	5 000 - 100 000
■	1 500 - 5 000
■	300 - 1 500
■	100 - 300
	1 - 100
■	0 - 1

Using the data : similar results…

JRC 2006 estimation

DG REGIO Urban centres :

**More than 50 000 inh.
in contiguous cells with
more than 1 500 inh. in each cell**

Census 2007

The effect of the grid projection system
Shifting the grid

	original LAEA 1km ²	Shifted LAEA 1km²
Poitiers	61766	62947
La Rochelle	68802	65840
Angoulème	35797	45292
Niort	43714	39060
Châtellerault	20685	20862
Saintes	15094	14219

Populations of the urban centres
(simulation with fiscal population 2009)
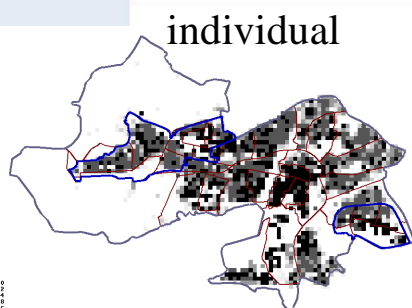
Ordering is grid dependent
Accuracy is not guaranted
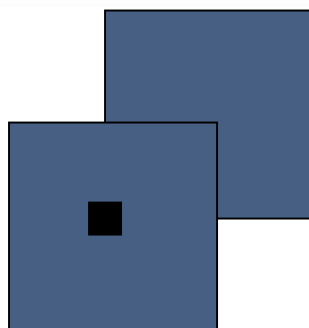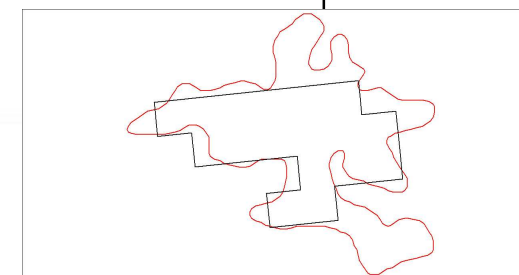
The effect of the grid size
Using individual data

	LAEA 1km²	LAEA 1km² shifted	in : Individual out: L2E (200m)²
Poitiers	61766	62947	72900
La Rochelle	68802	65840	86200
Angoulème	35797	45292	50600
Niort	43714	39060	44600
Châtellerault	20685	20862	23300
Saintes	15094	14219	16500

Now over the 50 000 inh. threshold

individual

200m wide grid cells

$$\hat{f}(x) = \sum_{i=1}^{N} \frac{1}{M\mathbf{h}} K\left(\frac{x - X_i}{\mathbf{h}}\right)$$
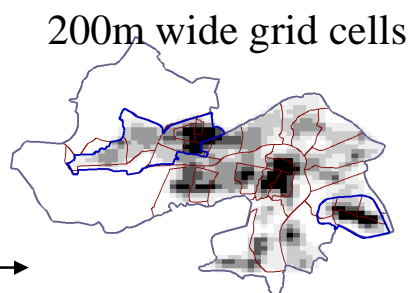
Grid cells are pixels
The picture is important…

200m
Municipality

100m
Neighborhood

1km
NUTS3

but should have enough pixels

INSEE
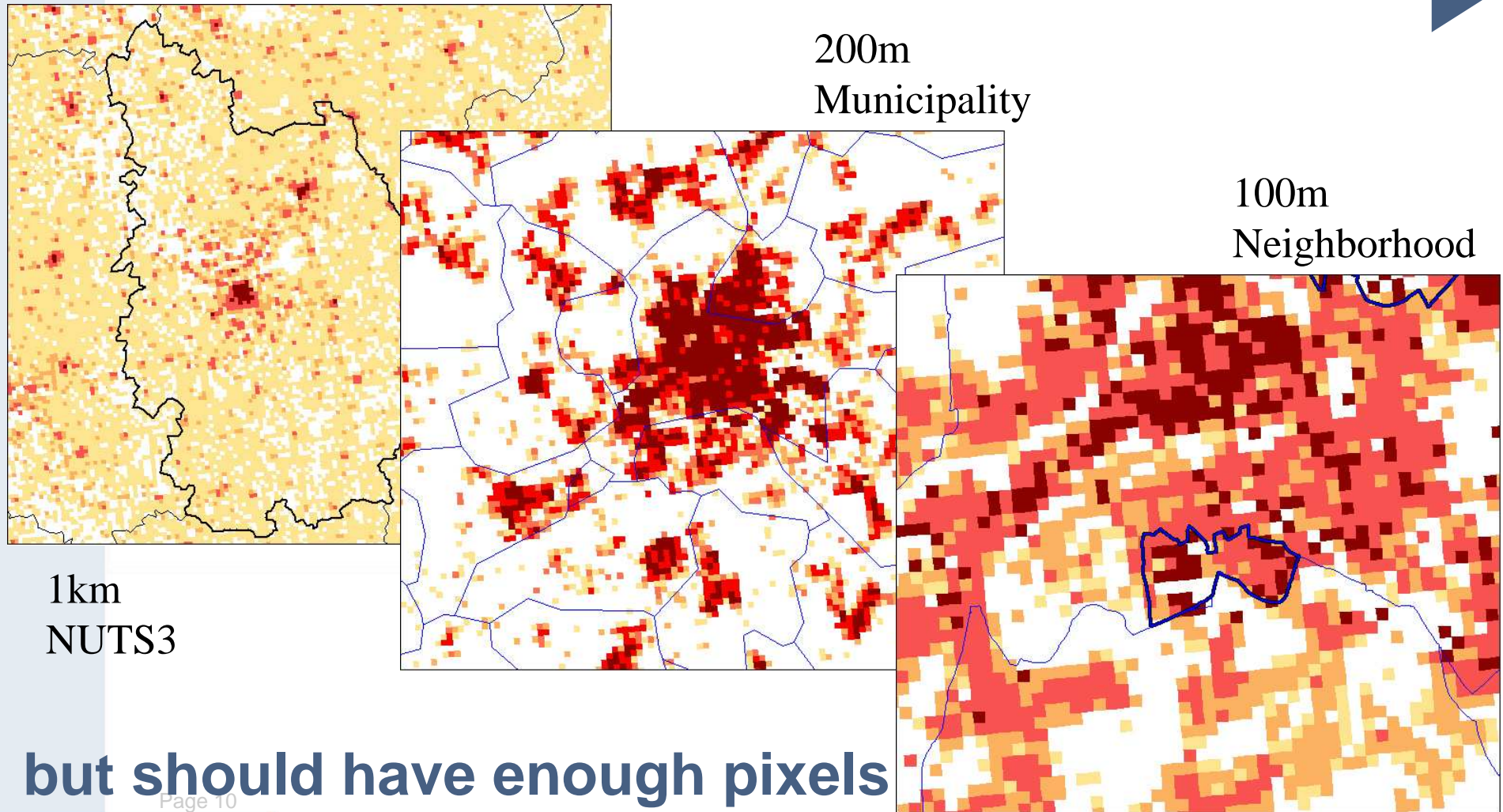
Consequences for dissemination

› We shall need more variable than just population counts

› Target cell size must be 200m or lower : most cells will encounter confidentiality problem

   – Masking data will destroy anything

› It is the global picture that is important : Real data for each cell is not necessary

   ➔ geographical swapping and imputation

   ➔ the way to dissemination of files of individual records is open

INSEE

Dissemination
Scrambling the data

id	n	SEXE_1	SEXE_2	AGE_1	AGE_2	AGE_3
00134	1	1	0	0	0	1
00135	1	1	0	0	1	0
00222	20	8	12	3	13	4
00308	11	8	3	3	8	0
00309	12	7	5	0	11	1
00394	3	2	1	0	1	2
00395	3	1	2	1	0	2
00473	1	0	1	0	1	0
00474	3	1	2	0	3	0

After treatment

id	n	SEXE_1	SEXE_2	AGE_1	AGE_2	AGE_3
00134	1	0	1	0	0	1
00135	1	0	1	0	1	0
00222	20	11	9	4	12	4
00308	11	6	5	2	9	0
00309	12	6	6	0	11	1
00394	3	3	0	0	1	2
00395	3	1	2	0	1	2
00473	1	0	1	0	1	0
00474	3	2	1	0	3	0

Scrambled
but still usable

original
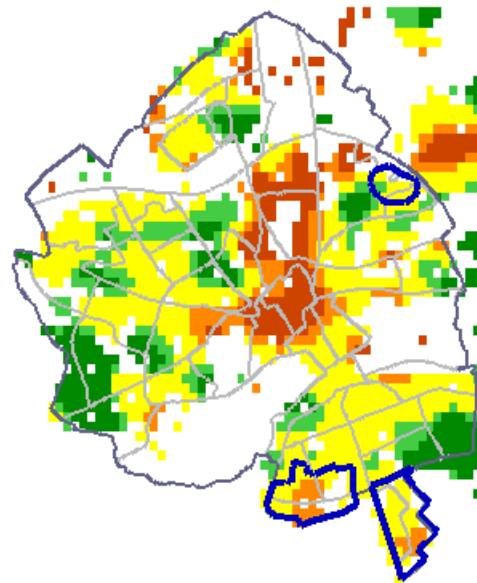
Municipality of Caen
Young employees residence place
(kernel density analysis)

a - Surreprésentation age_1/n   0.81 << >> 1.25

scrambled

a - Surreprésentation age_1/n   0.80 << >> 1.23

Jean-Luc LIPATZ   Lisbon   October 2011

INSEE
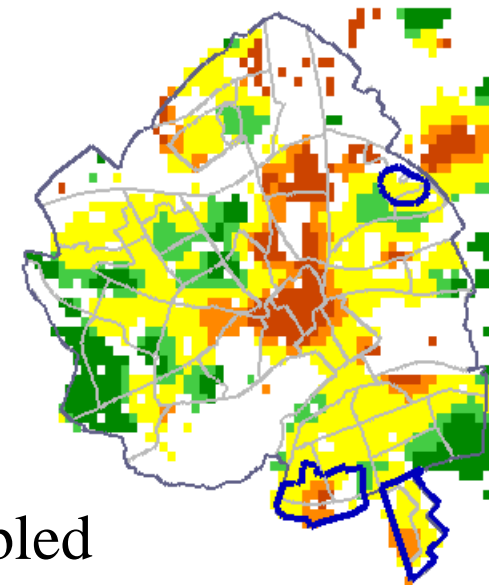
Conclusions

› If you want, you can

› Having smoothed data is also an advantage

› The Geostat projects are not ambitious enough.

- Agriculture
- Commerce
- Comptes nationaux - Finances publiques
- Conditions de vie - Société
- Conjoncture
- Économie
- Enseignement - Éducation
- Entreprises
- Industrie - IAA - Construction
- Population
- Revenus - Salaires
- Santé
- Services - Tourisme - Transports
- Territoire
  - Pays, villes d'Europe et du monde
  - Régions, départements et villes de France
  - Environnement
- Travail - Emploi

Poitou-Charentes
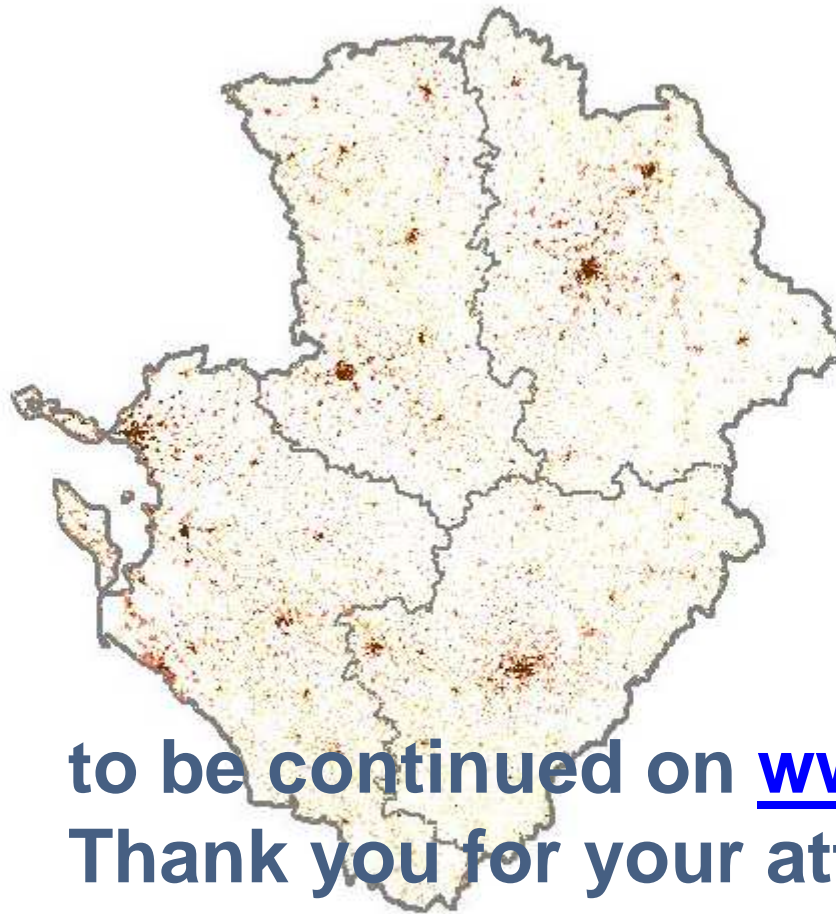
:: Légende

Densité de population
Nombre d'habitants estimés
par carreaux de 200 m

- 0 - 4
- 5 - 9
- 10 - 19
- 20 - 59
- >= 60
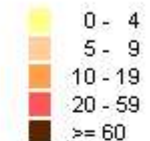
:: Pour en savoir plus

- Documentation sur les données carroyées

:: Liens

- Changer de région

to be continued on www.insee.fr…
Thank you for your attention

Source : Insee, RFL 2009.

Carte : carreaux de 200m sur 200m

:: Téléchargement des données

Chaque fichier compacté (format zip) contient les données sous deux formes :

- Un fichier en format dBase (DBF),
- Un fond de carte en format de transport MapInfo (MIF/MID).