European City Statistics: Prospects for the Future

Teodóra Brandmüller
Eurostat, Regional statistics and geographic information unit

Cities4People: Smart Cities and Data Analytics
’s-Hertogenbosch, 18 April 2018
Most densely populated...

Metropolitan region:  
*Southampton*  
5000 inhabitant/km²

NUTS 3 region:  
*Paris*  
21 000 inhabitant/km²

Local Administrative Unit:  
11ᵗʰ arr. of Paris  
42 000 inhabitant/km²
... the most dense 1 km\(^2\) grid cell
Hospitalet de Llobregat
Population density based on the GEOSTAT population grid, 2011

(number of inhabitants / 10 x 10 km)
A system of urban/rural typologies

Grid cells

A classification of raster sets at 1 km² using metrics of population density RIS 2012/13.

Where available, the population distribution is derived from registers. Elsewhere, it is downscaled from base LAU2 population figures.

LAU2 units

Degree of urbanisation

A classification of zone administrative units LAU2 based on the share of rural population living in urban clusters and in urban centres.

NUTS 3 regions

Urban/rural typology

A typology of NUTS 3 regions based on the share of rural population living in rural regions and in urban clusters.

Three levels of urban/rural classification based on population distribution

Rural grid cells

Urban clusters > 5,000

Urban centres > 50,000

Rural areas

Towns and suburbs

Cities

Thinly populated

Intermediate density

Densely populated

Rural regions

Intermediate regions

Urban regions

For more information...


Regional and Urban Policy
Characteristics of the definition

- Population based definition
- Starts from the population grid
  - Avoids distortions caused by large variations in the area of administrative territorial units
- Uses three categories at three spatial levels
- Enables the collection, compilation and dissemination of harmonised statistics
- Allows better targeted policy-making at EU level
Commuter flows of London

- Inner London — West (UKI3)
- Inner London — East (UKI4)
- Outer London — East and North East (UKI5)
- Outer London — South (UKI6)
- Outer London — West and North West (UKI7)
- Berkshire, Buckinghamshire and Oxfordshire (UKJ1)
- Surrey, East and West Sussex (UKJ2)
- Hampshire and Isle of Wight (UKJ3)
- Kent (UKJ4)
- East Anglia (UKH1)
- Bedfordshire and Hertfordshire (UKH2)
- Essex (UKH3)
Characteristics of the definition

- Population based definition
- Starts from the population grid
  - Avoids distortions caused by large variations in the area of administrative territorial units
- Uses three categories at three spatial levels
- Enables the collection, compilation and dissemination of harmonised statistics
- Allows better targeted policy-making at EU level
- Administrative boundaries do not necessarily mirror the current social and economic reality
Prospects for the future

• Complement existing territorial classifications with more flexible, more functional statistical geography
Commuter flows of Milano and Rome

Source: Presentation of L. Franconi at the Workshop on Labour Market Areas, ISTAT
Day and night-time population
Ljubljana

Administrative data calibrated by patterns observed in mobile phone network data
Source: Statistical Office of the Republic of Slovenia
Where are people during a typical weekday?
Belgium

Source: Joint project of Statistics Belgium, Proximus and Eurostat
Classification of the territory: residential, commuting and working areas profiles

Belgium

Note: experimental statistics based on a joint Eurostat, Statistics Belgium and Proximus project

Source: Eurostat
Prospects for the future

• Exploit novel data sources: open data, big data, etc.
• Promote geocoding of statistical and administrative data sources
Change in usually resident population (%) 2000–2017

Source: Central Statistical Bureau of Latvia
Changes in resident population

Riga
2000–2017

Source: Central Statistical Bureau of Latvia
Prospects for the future

- Use technology embedded in smart systems aiming at transforming data into knowledge presented in the form of statistics: Trusted Smart Statistics
Trusted statistics might not be good enough

- Ensuring validity and accuracy of the outputs
- Respecting data subjects' privacy and protecting confidentiality
- Confirm the principles covering the institutional environment, the statistical production processes and the output of statistics.
- Privacy by design, end-to-end security, auditable data life-cycle, transparency, satisfy consent and purpose conditions, assessment boards, ...
Exercise

Find out whether Bob weights more than Carl

- **Method 1**: by scale
  - measure Bob’s weight $x_1$
  - measure Carl’s weight $x_2$
  - compare $x_1$ vs $x_2$

  $$y(x_1, x_2) = \begin{cases} 
  1, & \text{if } x_1 > x_2 \\
  0, & \text{otherwise}
\end{cases}$$

- **Method 2**: by see-saws
  - The result is obtained without learning any other information!

**Privacy preserving computation**
Cross-domain usage of private data: problem statement

PROBLEM: let result party get output $y$, without disclosing private input $x_i$ to any other party (not to result party, not to other input party $x_j$)
Cross-domain usage of private data: via Secure Multiparty Computation

Input data parties (one or more)

Computing parties

Result parties (one or more)

$p_{n,k}$

non-invertible function $p_n = g(x_n)$

private $\rightarrow$ public
Cross-domain usage of private data: via Secure Multiparty Computation

Input data parties (one or more)

Computing parties

Result parties (one or more)

\[ p_{n,k} \]

export public data that cannot be inverted back to private data (sharing)
Cross-domain usage of private data: via Secure Multiparty Computation

Input data parties (one or more)

Computing parties

Result parties (one or more)

"use" data without "sharing" data

$p_{n,k}$
A real pilot case-study in Estonia

Linking Education records from Ministry of Education with Employment records from Tax Authority

From Internet of Things to ...

A set of sensors, actuators, smart objects, data communications and interface technologies that

- allow information to be collected, tracked and processed across local and global network infrastructures,
- enabling the future hyper-connected society
... Smart statistics

... data capturing, processing and analysis will be embedded in the system itself ...

... putting intelligence to all stages of the data life-cycle

... cognitive processes will be fundamental on how smart statistics will be created
"Smart statistics" in the business case 2018-2020

Citizen science data and smart cities

- Use of citizen science data – Use of smart devices, wearable sensor technology (e.g. fitness trackers) related to wellbeing and physical activities and/or health-related information (e.g. pulses, blood pressure)
- **Statistical themes:** Quality of life, environmental protection, smart traffic, etc. focusing on the use of new technologies and sensors in an urban environment
"Smart statistics" in the business case 2018-2020

Smart cities and connected vehicles

- Use of smart vehicles, smart parking, meteorological stations
- **Statistical themes:** Urban mobility, road safety, optimised transportation resources, autonomous vehicles
Prospects for the future

• Complement existing territorial classifications with more flexible, more functional statistical geography

• Exploit novel data sources: open data, big data, etc.

• Promote geocoding of statistical and administrative data sources

• Use technology embedded in smart systems aiming at transforming data into knowledge presented in the form of statistics: Trusted Smart Statistics
Thank you for the attention!

http://ec.europa.eu/eurostat/web/regions-and-cities/overview

Questions:
ESTAT-REGIO@ec.europa.eu