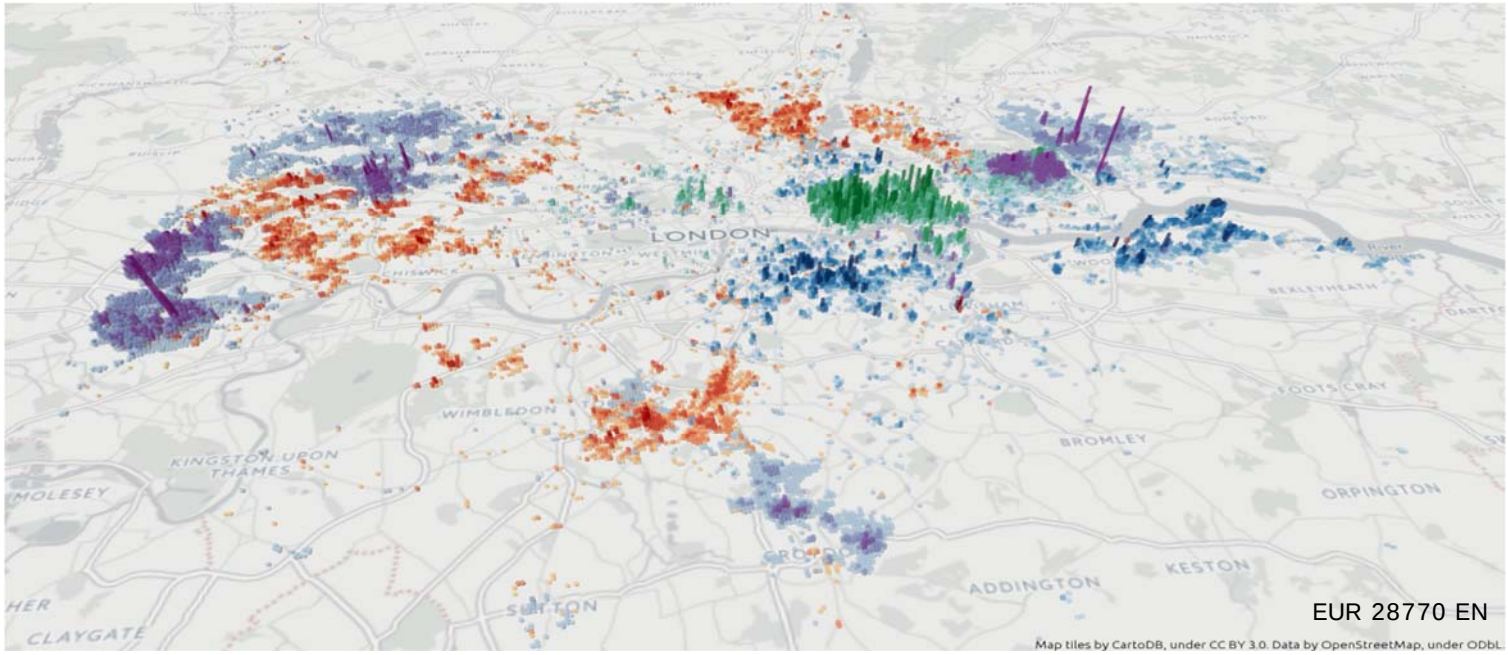
European Commission

JRC TECHNICAL REPORTS

High resolution map of migrants in the EU

Alessandrini A., Natale F., Sermi F., Vespe M.

2017



EUR 28770 EN

JRC Science Hub
https://ec.europa.eu/jrc

JRC 108175
EUR 28770 EN

PDF ISBN 978-92-79-73420-5 ISSN 1831-9424 doi: 10.2760/0199

Luxembourg: Publications Office of the European Union, 2017

How to cite this report: Alessandrini, A., Natale, F., Sermi, F. and Vespe, M., *High resolution map of migrants in the EU*, EUR 28770 EN, doi: 10.2760/0199.

Contents

Abstract

This report describes a data set generated from the harmonization and spatial processing of official census statistics collected from National Statistical Institutes in 8 EU Member States (France, Germany, Ireland, Italy, Netherlands, Portugal, Spain and UK).

The dataset provides a map at high spatial resolution of the population with migrant background in these Member States.

The uniqueness of the data set resides both the high level of spatial resolution (cells of 100 by 100 m) and the large geographical coverage that includes almost 45,000 local administrative units.

From this data set it is possible to calculate indicators of concentration of migrants, diversity and spatial residential segregation which can support comparative research and policies on the local dimension of the integration of migrants in the EU.

Introduction

In the literature there are numerous studies regarding the concentration, diversity and segregation of migrants at local level.

These studies range from qualitative research and social surveys, looking at various aspects of urban diversity and socio-cultural integration of migrants' communities in specific neighbourhoods, to more quantitative analyses, considering spatial patterns of residential segregation.

Quantitative studies on spatial residential segregation typically rely on information collected from national censuses that provide aggregated statistics on the number of people by origin (country of birth and citizenship) and census areas.

Despite the advancement of methods and wealth of indexes and tools to measure spatial segregation [1] [2] there have been few empirical applications to compare the patterns of segregation across cities and countries [3] [4].

In most empirical applications the calculation of the segregation indexes at local level is limited to few large cities and is considering aggregate ethnic groups. The main reasons for the few comparative studies reside in: the difficulty of assembling data from several National Statistical Institutes (NSI); the lack of standardization in the aggregation levels, geometries, definitions; and confidentiality requirements.

We aim to address these challenges by proposing a data set that provides for the first time the possibility of mapping at high spatial resolution the migrants communities in several EU Member States.

The data set includes maps of the number of people by citizenship and/or country of birth organized in a uniform spatial grid with 100 by 100 m cells.

The uniqueness of the data set resides both in the high level of spatial resolution and the large geographical coverage which is includes almost 45,000 local administrative units in France, Germany, Ireland, Italy, Netherlands, Portugal, Spain and UK.

The data set is obtained by harmonizing and spatially disaggregating the statistics of the 2011 Census collected through ad hoc extractions from the NSI.

In this report we describe the methods applied for the processing of the original census data and the final characteristics and variables included in the dataset.

The proposed dataset allows the calculation of indicators ranging from the very simple ratio of migrants over the total population, to more structured indicators of spatial concentration, ethnic diversity and spatial residential segregation.

These indicators can be calculated for single cells, for administrative units at different levels (i.e. local administrative unit, functional urban area, region, province, country) or for ad hoc geometries.

Ideally, the indicators can be used in combination with spatially detailed information from micro census, social surveys or other sources of administrative data, to model relations with educational segregation, income segregation, housing conditions, electoral outcomes or in support of more qualitative case studies for specific neighbourhoods or communities.

At policy level, a systematic mapping of the concentration of migrant across and within cities offers to local and national administrators the possibility to fine-tune interventions in areas where the high concentration of migrant communities might pose challenges for local institutions such as hospitals and schools.

Since the database also includes high-resolution information on towns and villages, it enables researchers and policy-makers to investigate whether diversity and concentration of migrants are phenomena only affecting global cities or they are also relevant in the case of specific urban and rural peripheries. Such local perspective is critical to address, on the one hand, issues of migrants' integration and, on the other, public perceptions around migration which risk to go unnoticed in public discussion and research based on migration data aggregated at country or regional level.

Data sources

There are two main data sources used for the preparation of the maps: 1) the national census statistics that report the number of persons by country of birth and citizenships living in each census area; 2) the CORINE land cover (CLC) together with the European Settlement Map (ESM) layers [5–7], which provide ancillary information to disaggregate data from the census areas into a uniform grid.

Census statistics

In most cases, data was obtained by asking to the NSI to perform ad hoc extractions. For the UK and Ireland, data was downloaded from publicly accessible, data dissemination platforms[1].

The definition of migrants in the census data is based on the country of origin and the citizenship criteria. To produce the maps we used data by citizenship in the case of Italy and France, and by country of birth in the other cases (Table 1).

Since each NSI uses different coding systems and classifications to represent the origin of migrants, after collecting the data a reclassification and harmonization procedure was implemented, including the conversion of the different national classifications in standardized ISO country codes (ISO 3166-1 alpha-3 code). For simplicity ISO country codes are applied both to the citizenships and to country of birth even though the citizenship may not necessarily correspond to countries in some cases.

The geographical resolution and geometries used to define the census sampling units are extremely variable across EU Member States[2] (Figure 1). In the case of Netherlands, data was provided using very detailed sampling areas represented by the postal code level (groups of buildings including around 25 households and with an average surface of around 0.001 square km). Other countries reported data at higher resolution (from 0.01 to 1.7 square km) using census sampling areas with a regular grid (Germany) or polygons with variable shapes and sizes, designed to represent statistical units with similar numbers households or people (France, Ireland, Italy, Portugal, Spain and UK).

[1] UK data Archive and IE National archives

[2] According to the Regulation on Inspire there is an obligation to provide census data to Eurostat using a uniform EU grid with cells of 1 by 1 km but this requirement will come into force only for the 2021 census round and it will still not be applicable to all variables included in the census.

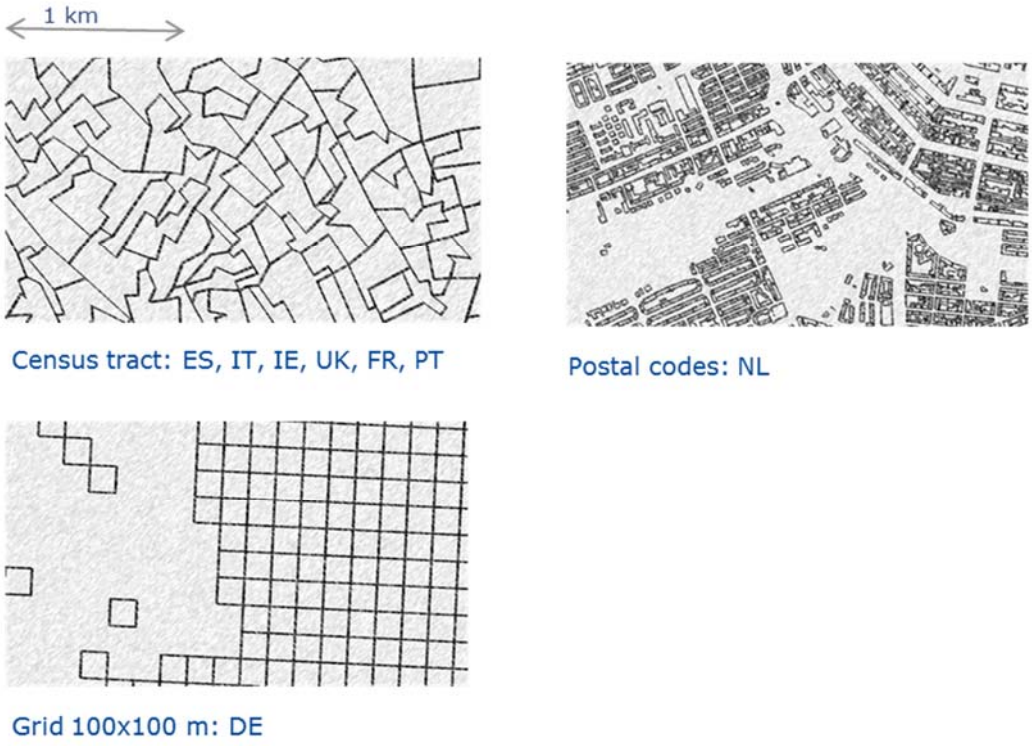


Figure 1 Different geometries and resolution in the original Census data provided by the NSI.

These differences in geometries and resolution were harmonized through the dasymmetric mapping method described more in detail in the next chapter.

Another significant aspect in the national census data, is related to the data protection policy applied to hide confidential information. The three main methods used by the NSI to preserve confidentiality consisted in: hiding values below a certain threshold (e.g. census areas and origins with less than five persons and zero), applying a perturbation methodology to swap the data between census areas, and aggregating the data by the origin dimension.

The aggregation along the origin dimension (by continents or EU versus Third country origin) allowed to preserve the geographical detail. This aggregation resulted in three different data sets that present the population of migrants for the same geographical units at increasing aggregation levels for the origins. The more aggregated datasets contain less empty cells (i.e. made empty for confidentiality reasons) and therefore their sum matches more closely the figures published in official statistics at country level.

In the case of Germany, we adopted in collaboration with the Federal Statistical Office an ad hoc procedure to perform the extraction of the data based on flexible geometries. The procedure consisted in merging cells through an agglomerative algorithm (max-p algorithm

- PySAL) which was constrained to an optimal population size per region minimizing the loss of information due to the confidentiality rules.

In addition to the confidentiality thresholds applied by the NSI before processing the data we further removed records (combinations of census area and origin) having a population count of less than 5 persons[3]. This additional filter was introduced to bring at the same level of confidentiality data sets which were provided using less restrictive thresholds and to ensure that no confidential data would be contained in the final data set.

The following table summarizes the main characteristics of the original data Census data used in this study.

Table 1 Characteristics of the Census data used in the study

Country	Census areas	Criteria for the definition of migrants	Aggregation level for the origin	Data protection policy
Italy	920 K medium resolution polygons 'Sezioni censimento' (approx. 0.06 square km and representing ~250 households)	Citizenship	Nation	Threshold value
Germany	3.1 M cells in a regular 100x100 m grid (0.01 square km)	Birth and citizenship	Nation	Threshold value
France[4]	19 K low resolution polygons 'TRIRIS' and 'IRIS' (approx. 1.7 square km representing ~1800-5000 residents)	Citizenship	Nation	Spatial aggregation (low resolution areas)
Spain	35 K low resolution polygons 'Secciones censales' (approx. 0.2 square km representing ~1000-2500 residents)	Birth and citizenship	Nation (only 20)	First 20 communities
UK	232 K medium resolution polygons 'Output areas' (approx. 0.07 square km representing clusters of adjacent postal codes)	Birth	Nation (only 20)	Statistical perturbation: random record swapping
Portugal	low resolution polygons (approx. 0.6 square km)	Birth	Continent	Aggregation (continent)
Netherlands	417 K high resolution polygons 'postal codes' (approx. 0.001 square km representing ~25 households)	Birth and citizenship	Country	

[3] The final data set produced after the spatial disaggregation shows also cells with values below five persons and with decimals but these values are estimated and do not represent confidential values.
[4] Data for France include only municipalities with more than 5000 inhabitants (around 2000 out of 36000).

Ireland	18 K low resolution polygons 'Small areas' (approx. 0.7 square km representing 50-200 households)	Birth and citizenship	Nation (only 5)	Only major communities

Ancillary data (ESM and CLC)

The ESM is a raster dataset that map the distribution of the human settlements in Europe. It is derived from SPOT satellite imagery and additional ancillary dataset (CORINE, GEOSTAT population data). The dataset is produced using different machine learning algorithm by extracting image features to detect the built-up areas. The ESM reports at 10 m resolution the percentage of the built-up areas in the cell. This feature was used as main input to distribute the original residential population from the census areas.

The CLC, is a raster and vector dataset, reporting a classification of the land use (44 classes). The CLC is derived using satellite imagery (Landsat, SPOT, IRS). We used the CLC vector dataset as additional weighting factor applied to the built-up area from the ESM data set.

Spatial disaggregation (dasymetric mapping)

The main part of data processing was a dasymetric mapping. This is a well-established method used to disaggregate spatial data and produce high resolution gridded population maps.

With this method we redistributed the population by origins from the original census areas to a regular grid at 100 m resolution using as reference two ancillary datasets: the ESM built-up distribution and the CLC, both at 10 m resolution.

In a nutshell the method allocates higher shares of the total population to cells characterized by a higher surface occupied by buildings and with an urban land cover classification, as compared to cells occupied, for example, by green areas or with an industrial or agricultural land cover classification.

The application of the dasymetric mapping methodology served two main purposes. Firstly, it allowed to harmonize the level of spatial aggregation across the different datasets to an intermediate level of cells of 100 by 100 m. The high-resolution dataset (e.g. Netherlands) were up-scaled while the medium and low resolution dataset (e.g. Italy, Spain, France) were down-scaled. Secondly, it allowed to harmonize the different geometries to a uniform and spatially consistent grid covering the entire EU.

The dasymetric mapping consisted in the following steps (Figure 2):

1. Creation of uniform grid layer for the entire EU at 100 m resolution. This empty layer is the target were the original population count is redistributed. The target layer was re-projected to the European Terrestrial Reference System (ETRS-LAEA; EPSG 3035), an equal-area projection used for statistical mapping across the European area.
2. A weighting factor for the redistribution was calculated from the ESM and corrected on the basis of classes reported in the CLC layer.
3. The census area population relative to each citizenship, was proportional distributed to the destination grid based on the building area footprint using the following formula:

$$P_{grid} = P_{census} \frac{BF_{grid}}{BF_{census}}$$

Pgrid : *migrant communities' population in the destination grid cell*
Pcensus : *migrant communities' population in the source census polygon*
BFgrid : *building area footprint in the destination grid cell*
BFcensus : *building area footprint in the source census polygon*

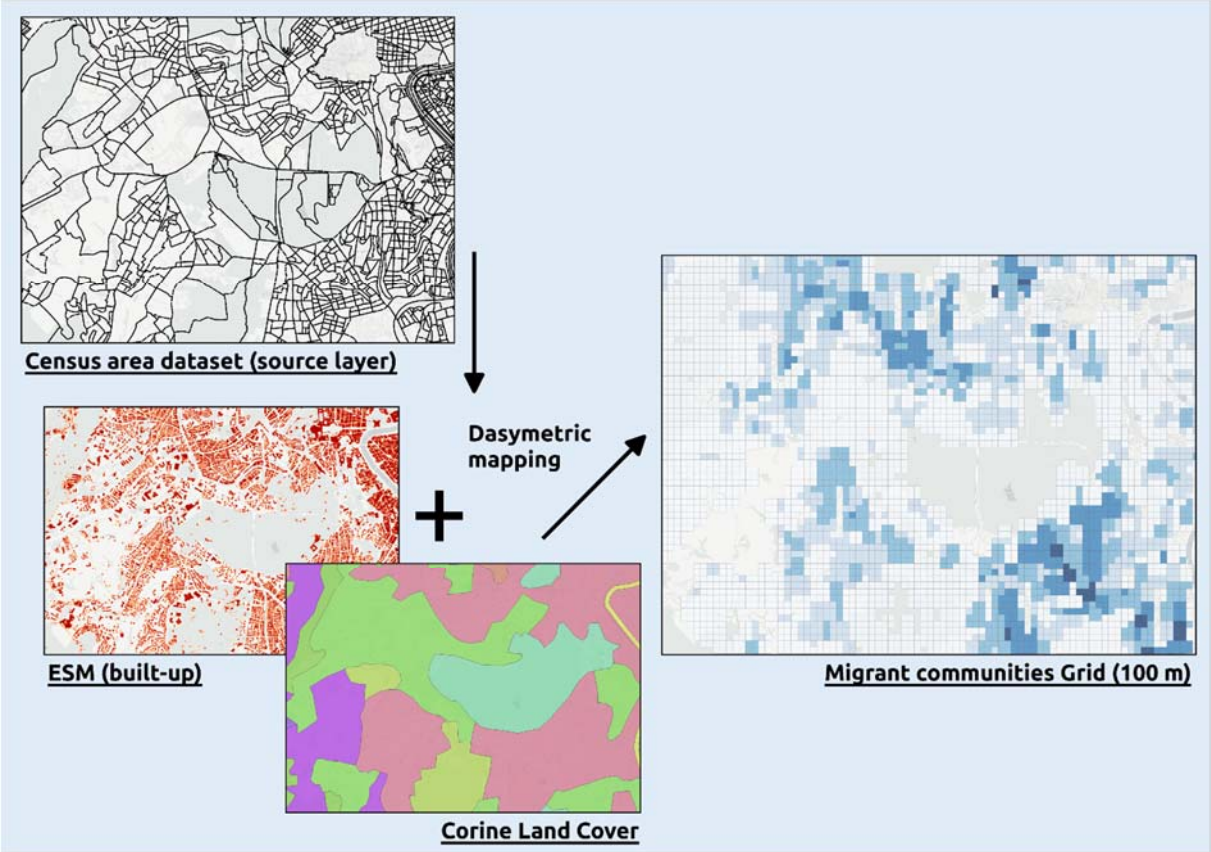


Figure 2 Schematic description of the steps for the processing of the data.

Characteristics of the final data set

The end results of the processing described above is a grid layer, where each cell reports the residential population by origin at three different level of aggregations (country, continent, EU vs. Third country origin).

The following Figure 3 shows an example of three-dimensional visualization that can be derived from such data. The figure shows the absolute number of migrants in London for the 20 main countries of origin.

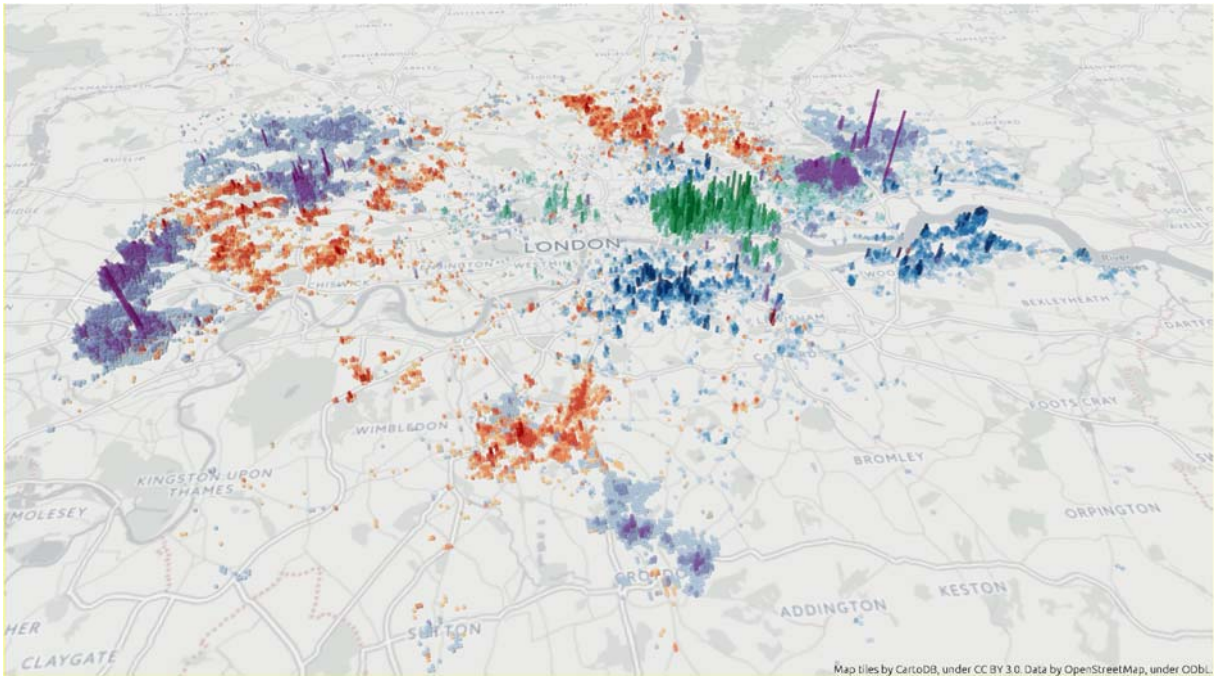


Figure 3 Absolute number of migrants in London for 5 main countries of origin (Nigeria blue, Bangladesh green, India purple, Poland red). The highest bars correspond to a population of 260 persons for a given country of origin living in the cell of 100 by 100 m).

The grid layer, has resolution of 100 m, and is projected using the equal area Europe an Terrestrial Reference System (ETRS-LAEA; EPSG 3035).

The data sets include two dimensions: the cell identifier (GRID_ID), the origin (c_birth) and one measures which represents the population count (pop). The string of cell identifier besides providing a unique identification includes the X, Y ETRS-LAEA coordinates of the cell centroid (e.g. 2499650N4149350E, is the cell centroid having coordinate 2499650 north and 4149350 east).

The dasymetric mapping, generated decimal values for the population count. We avoided any rounding of these decimal values to reduce differences in the totals in respect of the aggregated statistics.

The origin of migrants is reported using a standardized ISO country code (ISO 3166-1 alpha-3 code). The code OTH is used to indicate the residential population aggregated for confidentiality reasons or for cases where it was not possible to convert the citizenship or country of birth in an ISO country code. The national population is included in the dataset to allow the calculation of total population and shares of migrants.

Besides the main data set of gridded data by specific country codes, two additional datasets were generated, aggregating respectively the country of origin by continents and by EU vs. not-EU origin.

The following Table 2 shows the main statistics for the final three final data sets, highlighting also the extent to which the data captures the national aggregation information (e.g. for Germany the data set cover areas for a total population of around 68 million people in the case of the EU/non-EU counts, whereas the total population reported by Eurostat is approximately 80 million in 2011).

Origin by country

	DEU	ESP	FRA	GBR	IRL	ITA	NLD	Grand Total
Population	63.38M	40.64M	24.88M	59.49M	4.04M	34.70M	14.62M	241.74M
Migrants	5.05M	1.63M	1.30M	4.82M	0.51M	1.75M	0.41M	15.47M
Nr of cells	2.42M	1.73M	0.70M	2.56M	0.49M	1.45M	0.56M	9.91M
Nr of LAU	10,490	7,724	2,121	9,498	3,439	7,142	408	40,822
Nr of origins	140	21	125	20	5	148	120	178

Origin by continent

	DEU	ESP	FRA	GBR	ITA	NLD	PRT	Grand Total
Population	69.39M	44.75M	26.15M	62.74M	54.83M	16.01M	7.33M	281.19M
Migrants	10.98M	4.99M	2.51M	7.62M	4.02M	1.73M	0.49M	32.35M
Nr of cells	2.48M	3.57M	0.92M	4.08M	4.33M	0.64M	1.57M	17.59M
Nr of LAU	10,496	7,766	2,670	9,498	8,045	408	3,728	42,611

Origin by EU vs Third country

	DEU	ESP	FRA	GBR	IRL	ITA	NLD	PRT	Grand Total
Population	69.39M	44.75M	26.15M	62.74M	4.52M	54.83M	16.01M	7.33M	285.71M
Migrants	10.98M	4.99M	2.51M	7.63M	0.77M	4.02M	1.73M	0.49M	33.12M
Nr of cells	2.48M	3.57M	0.92M	4.08M	1.33M	4.33M	0.64M	1.57M	18.92M
Nr of LAU	10,496	7,766	2,670	9,498	3,441	8,045	408	3,728	46,052

Table 2 Summary statistics for the three data set including respectively population counts by specific country, continent and EU vs. Third country of origin.

Differences between the total population included in the data sets in respect of the official population statistics for each country stem from the suppression of values below confidentiality thresholds and in the case of France also from the limited coverage of the original census data which was only including municipalities with more than 5000 inhabitants.

The absence of countries in the table indicate that data from the NSI was not available for the given aggregation level.

A lookup table was created, to link the population grid dataset to different classifications of administrative units. This table covers the entire territory of the eight Member Sates and includes cell identifiers, additional geographical coordinates according to WGS84 reference system, and codes and names for different levels of classification of administrative units: Territorial Units for Statistics (NUTS1,2,3), Functional urban areas (FUA) according to the OECD classification and Local Administrative Units (LAU) level 2 (Communes local area).

By joining the main tables and the lookup table it is possible to aggregate and filter the population by administrative units.

Maps of the data can be produced using the grid coordinates included in the main tables or linking the codes of administrative units to shape files which can be found at http://ec.europa.eu/eurostat/web/nuts/overview.

One example of calculation of indicators is provided in the Figure 4 below.

The figure shows diversity indexes (Shannon entropy) calculated for local administrative units in Germany, Italy, France and Netherlands.

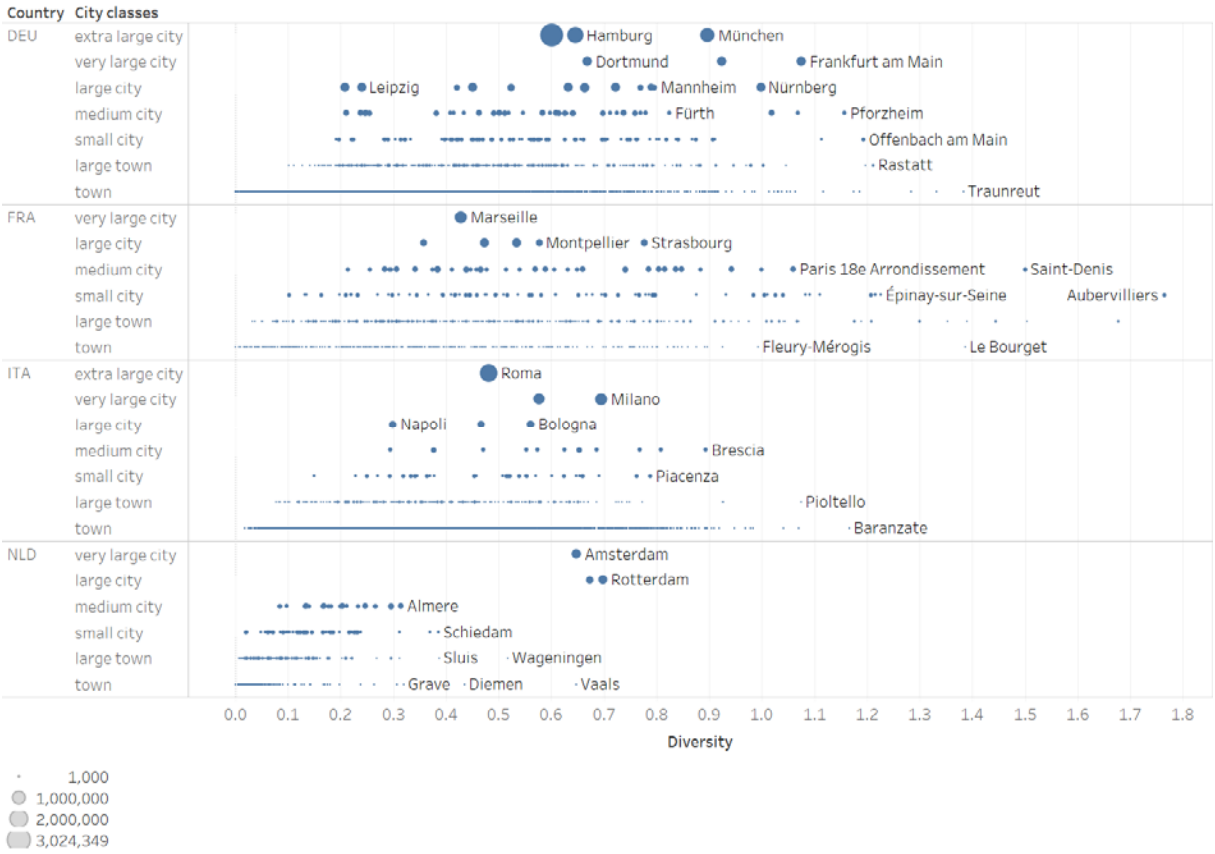


Figure 4 Diversity index (Shannon entropy) calculated for all local administrative units in 4 MS.

The diversity index reflects both the distribution of population across different origins and the variety of migrants groups. A high value of the index may derive from a high number of origins and shares of population more evenly distributed across several origins.

The figure exemplifies the potential of comparative analyses across different sizes of cities and countries. What is interesting to note is that also small cities and towns may show a higher diversity and concentration of migrants with respect to larger cities.

Limitations

Despite all the efforts put in harmonizing the different original data sets, several limitations and assumptions should be considered when using the final data to calculate measures of concentration, diversity and segregation.

The dasymetric method is in an estimation and therefore the figures about population in each cell should not be interpreted as the exact number of people residing in the area. We have applied several validation procedures to verify that the aggregation of the data reproduces the totals for the original census areas provided by the NSI. However, differences in respect of totals reported at country level in official statistics are still present given the suppression of confidential values made by the NSI and in our additional filtering.

One assumption used in the dasymmetric mapping is that the distribution of the nationalities in the original census area is uniform, therefore the aggregation in the original data at census area level does not take into account the possible clustering of residential areas for specific migrants' communities.

The ESM raster dataset used as weighting factor for the dasymmetric mapping represents only building footprints and does not take in account the vertical dimension of human settlements. The redistribution may result in underestimation of population counts for cells occupied by tall buildings hosting many people in a small surface.

The data sets only provide a static representation of the population at the reference year of the 2011 census. Since 2011, the concentration of the migrant population may have dramatically changed. Urban dynamics and changes in the ethnic composition of neighbourhoods are very interesting phenomena which cannot be explored with the proposed data set.

Comparison between cities in the same countries is legitimate since the same rules for confidentially and aggregation of data have been applied in the original data sets by the NSI. However, comparison across countries has to take into account of the differences in the original data set which are only in part compensated by the processing methods applied. Table 1 provides some guidance on which country combinations have similar aggregations levels along the two dimension of geography and origins.

AS previously mentioned, the data set which is here described allows calculating indicators about residential concentrations diversity and segregation. These indicators assume that places where people live influences their interaction and integration outcomes. However, migrants interact and integrate in the receiving society in many other places linked to education, work and leisure activities and increasingly also in social media. Measures of

integration should expand the number of dimensions to be observed and consider interaction possibilities in a more encompassing social and economic space not exclusively based on residential patterns.

Acknowledgements

We thank the National Statistical Institutes (ISTAT (Italy), INE (Spain), DESTATIS (Germany), INSEE (France), CBS (Netherlands)) for providing the original data from the 2011 Census through ad hoc extractions and the statistical services in Ireland and UK for exposing detailed census data through their online data dissemination platforms.

References

1. Massey DS, Denton NA. The Dimensions of Residential Segregation. Soc Forces. 1988;67: 281. doi:10.2307/2579183

2. Reardon SF, O'Sullivan D. Measures of spatial segregation. Sociol Methodol. 2004;34: 121–162.

3. Peach C. The mosaic versus the melting pot: Canada and the USA. Scott Geogr J. 2005;121: 3–27. doi:10.1080/00369220518737218

4. Johnston R, Poulsen M, Forrest J. The Geography of Ethnic Residential Segregation: A Comparative Study of Five Countries. Ann Assoc Am Geogr. 2007;97: 713–738. doi:10.1111/j.1467-8306.2007.00579.x

5. Ferri S, Syrris V, Florczyk A, Scavazzon M, Halkia M, Pesaresi M. A new map of the European settlements by automatic classification of 2.5 m resolution SPOT data. Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International. IEEE; 2014. pp. 1160–1163. Available: http://ieeexplore.ieee.org/abstract/document/6946636/

6. Florczyk AJ, Ferri S, Syrris V, Kemper T, Halkia M, Soille P, et al. A New European Settlement Map From Optical Remotely Sensed Data. IEEE J Sel Top Appl Earth Obs Remote Sens. 2016;9: 1978–1992. doi:10.1109/JSTARS.2015.2485662

7. Pesaresi M, Huadong G, Blaes X, Ehrlich D, Ferri S, Gueguen L, et al. A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results. IEEE J Sel Top Appl Earth Obs Remote Sens. 2013;6: 2102–2131. doi:10.1109/JSTARS.2013.2271445

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*):

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the internet (http://europa.eu).

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

28770 EN

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi: 10.2760/0199

ISBN 978-92-79-73420-5