WASABY

WASABY

WAter and Soil contamination and Awareness on Breast cancer risk in Young women

D5.2 Deprivation Indexes Data Collection

WP5 – Ludivine Launay & Elodie Guillaume

V1 – 28th February 2021

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Sommario

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

1. Introduction

The WASABY project focuses on the geographical analysis of population-based cancer incidence data in connection with environmental factors, using breast cancer and water/soil contamination, as an example.
Breast cancer incidence analyses were augmented with socioeconomic data. As individual socioeconomic data were not available in cancer registries, an aggregated deprivation index has been used: the European Deprivation Index (EDI) or a national index, when available, in countries where the EDI could not be constructed.
The European Deprivation Index is constructed according to the Townsend principle of relative deprivation, taking into account multidimensional measures, which depend on many social determinants interacting within the context of deprivation.
The methodology applied to the EDI calculation allowed for the construction of a country-specific ecological deprivation index that best reflects individual experience of deprivation. Intuitively, keeping the home warm is not the same need in the north and in the south of Europe. EDI is constructed using the European Union Statistic Income and Living conditions survey (EU-SILC) along with a national census. (1,2).

In the WASABY project, Work Package 5 (WP5) is dedicated to the deprivation indexes and is lead by the University of Caen, Normandy. The first objective was to constitute a list of experts covering each involved country for the collection of data on deprivation indexes or the deprivation indexes estimate. The second aim was to provide a national version of the European Deprivation Index or another Deprivation Index if its construction is not possible for all countries participating in the study, i.e., France, Germany, Italy, Lithuania, Poland, Portugal, Slovenia, Spain and Northern Ireland. In order to minimize the unavoidable ecological bias, the smallest geographical unit for which census data were available had to be identified to include deprivation (assessed at aggregated level) as a confounder in statistical modeling for breast cancer.

For each country where a national version of the European Deprivation Index (EDI) is not readily available, we had to collect information on census data availability, according to the geographical unit, in order to estimate the possibility construction of a deprivation index. If EDI is not estimable, other deprivation indexes had to be identified and collected. This final report concerns the WP5's activities since January 2018. After a methodological summary about the construction of the index, the data availability for estimating deprivation indexes in the various participating countries are detailed. Results of EDI in each country or national index are then presented and finally, a conclusion on the feasibility of the European Deprivation Index construction in each country and a discussion on analysis with data at the different levels is offered.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

2. Methodology

2.1. Indexes

2.1.1 European Deprivation Index

The method of EDI construction is based on three steps.

- The **first step** aims to construct an individual indicator for deprivation based on the identification of fundamental needs associated with objective and subjective poverty.
- The **second step** aims to identify variables available both at individual (EU-SILC survey) and aggregate levels (census) in each country.
- The **third step** is dedicated to the construction of an ecological deprivation index. In this step, individual indicators of deprivation are explained in a logistic regression by variables identified in the previous step. The regression coefficients became the weightings for the variables measured at the aggregated level. The final index is the sum of these weighted variables.

Step 1: Definition of an individual deprivation indicator

This step is based on the EU-Statistics on Income and Living Conditions (**EU-SILC survey**) in 2011. EU-SILC provides statistics on income, poverty, social exclusion and living conditions in the European Union and covers objective and subjective aspects of these themes in both monetary and non-monetary terms for both households and individuals. To obtain the data, an application has to be submitted to EUROSTAT. As the leader of WP5 had this authorization and the data for all European Countries, construction of EDI was centralized and undertaken by WP5's leader in consultation with experts from each country, to ensure that the results were consistent with the economic situation and the culture of the country.

Step 1.1: Selection of needs

There are 9 items (needs) in the EU-SILC survey common to all European Countries reflecting a financial capacity or incapacity to possess something. These are:

- ability to keep home adequately warm
- capacity to afford paying for one week annual holiday away from home
- capacity to afford a meal with meat, chicken, fish (or vegetarian equivalent) every second day
- capacity to face unexpected financial expenses
- do you have a phone ?
- do you have a colour TV ?
- do you have computer ?
- do you have a washing machine ?
- do you have a car ?

These are considered as potential fundamentals needs where less than fifty percent of the households do not possess because they cannot afford them.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Step 1.2: Selection of fundamental needs associated with objective and subjective poverties

- *Construction of objective poverty*

This variable is calculated from EU-SILC according to many variables relative to income. Poor staus is determined as those whose income is less than 60% of the national median income.

- *Construction of subjective poverty*

Subjective poverty is based on the variable "*Ability to make ends meet*", which is coded as follows:

- with great difficulty
- with difficulty
- with some difficulty
- fairly easily
- easily
- very easily

This variable is dichotomized with logistic regression explaining the objective poverty. The dichotomized variable is defined as poor (=1) / not poor (=0), using all successive combinations of subjective poverty modalities (modality 1 vs 2-6, modality 1-2 vs 3-6…etc.). The best threshold is the model for which the highest Wald test's Chi Square was obtained. A new subjective poverty variable was then created from the modality with the highest value.

- *Selection of fundamental needs associated with both poverties*

In this step, the association of needs selected at step 1.1 (owned by less than 50% of households) with objective and subjective poverty is tested with univariate and multivariate logistic regression. Needs associated with both poverty measures are those considered as fundamental using the 5% significance level. We obtained 'n' missing requirements.

Step 1.3: Construction of the binary individual deprivation indicator

This step aimed to determine the best threshold for the number of missing requirements due to financial inability among those selected in the previous step by first dichotomizing the number of missing requirements. Logistic regression is then used to explain objective poverty and subjective poverty by the number of missing requirements (at least 'n' missing requirements verses no more than n-1). The best threshold is obtained by the model with the highest Wald ChiSquare for both poverties or, if results are discordant, those given by the model with the highest Wald ChiSquare for objective poverty. A binary individual deprivation indicator (IDI) is created by the dichotomization of the number of missing requirements as follows: ≥ best thresholds vs < best thresholds.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Step 2: Variables available both at individual (EU-SILC survey) and aggregate levels (census)

Step 2.1: Selection of variables available in both database

It is necessary to identify variables representing deprivation that are phrased and coded in the same way in EU-SILC and each national census. In this step, partner collaboration is crucial because, depending on the country, survey data is not always freely available, it is not always available in English and it requires particular attention on the restrictions given by the two types of information. For example, sometimes, variable modalities in either the EU-SILC or the census data would have to be aggregated to achieve a perfect correspondence between the two

Step 2.2: Dichotomisation of plurimodal variables according to their relation with IDI

In the next step, all multimodal variables were dichotomised by logistic regression explaining the individual deprivation indicator by the variable recoded. Selection of the modality was completed by taking the best model (wald test) score.

Step 3: Ecological deprivation index

Step 3.1: Harmonization of the census tract according to 2nd step

Census tract level data had to be prepared according to the same dichotomisation of variables as in EU-SILC. This preparation of the census data was sometimes done by the country contacts.

Step 3.2: Selection of variables for ecological step

The Individual deprivation indicator was explained by variables previously identified in a logistic regression model. All variables significant at 5% were selected in the final model.

Step 3.3: Ecological step

The EDI is calculated as the sum of the selected variables in the census (previously normalized) weighted by the coefficients obtained in the previous final model on EU-SILC variables.

$$EDI = \sum_{i=1}^{n} w_i \times var_{i(aggreated\ level)}$$

with:

- w_i: the weight obtained by logistic regression for variable i
- var_i: the percentage calculated on census data at aggregated level for variable i

The EDI was then classified by quintiles of its distribution. The variables included and the classification of geographical units obtained were controlled and validated by the in-country contacts.

All these steps were undertaken in collaboration with the contacts identified in each country. This methodology has already been published elsewhere. (1, 2)

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

2.1.2 Alternative Local Indices

As mentioned earlier, it was not possible to calculate the EDI for all participating countries in WASABY due to the absence of key data. This was the case in two countries, Poland and Germany. In both cases alternative deprivation indices were used and are described here.

Poland

A multidimensional index of deprivation of Polish municipalities was developed using the same methodology as the Poviat Index of Deprivation PID, describe below. It used a conceptualisation of the deprivation phenomenon. Dimensions of deprivation that were considered in creating the index of local deprivation include population income, employment, living conditions, education and access to goods and services[1].

The construction of a synthetic indicator, that is the multi-dimensional index of deprivation, has been conducted in two steps. In the first step, five sub-indices presenting the selected dimensions of deprivation have been created and in the second, the indices have been aggregated to form a multi-dimensional index of deprivation. A detailed procedure of creating the index of deprivation has been as follows[2] :

- Determination whether a given variable is a stimulant or an inhibitor of deprivation
- Standardization of the variables according to the following formula:

  $S_{i=\frac{x_i-\bar{x}}{\sigma x}}$, for stimulant variable

  $S_{i=\frac{\bar{x}-x_i}{\sigma x}}$, for destimulant variable

  With:

  - x_i: the value for variable x
  - $\bar{x}$: the mean for variable x
  - σ_x: the standard deviation for variable x

- Limiting the impact of extreme values on the index value for a given municipalities
- Summing up the standardized values of the variables after the limitation and dividing the result by the number of variables in order to obtain five sub-indices: Wj = $\Sigma si/n$
- Summing up the values of sub-indices and dividing the result by the number of dimensions in order to obtain the synthetic index of deprivation: PID = $\Sigma Wj/n$

With:

n: the number of variableThe index was used to assess the spatial concentration of social and economic phenomena in the poviat and city scale.

[1] https://www.researchgate.net/publication/316427509_Local_concentration_of_deprivation_in_Poland

[2] https://www.researchgate.net/publication/316427571_Poviats_threatened_by_deprivation_state_trends_and_prospects

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Germany: The German Index of Multiple Deprivation (GIMD)

Based on an established British method as described by Noble et al. (3), Maier and colleagues developed Indices of Multiple Deprivation (IMD) for Germany: the German Index of Multiple Deprivation (GIMD), as well as its regional versions, e. g. for Bavaria (BIMD) (4). The GIMD consists of seven deprivation domains, which represent single aspects of deprivation (income, employment, education, municipal/district revenue, social capital, environment, and security). Specific indicators were generated from data of official statistics and assigned to the deprivation domains. The weighted single domains were finally combined to an overall index. The GIMD exists at municipal and at district level and is available for two reference years, 2006 and 2010 (4,5). A third generation for 2015 is currently being created.

The German IMD has been proven to be a valid and efficient tool for use in epidemiology and health services research, but also for health policy.

Co-funded by
the Health Programme
of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

2.2 Geographical unit

The countries involved in the WASABY project were: France, Italy, Portugal, Spain, Northern Ireland, Slovenia, Lithuania, Poland and Germany. Belgium left the project last year when step 2 was incomplete. So, it was a requirement that a European deprivation index would have to be available for participating countries, where possible. If not, an alternative local index was used as was the case for Poland and for Germany and explained above. The WP5's first objective was to identify a list of experts covering each involved country for the collection of data on deprivation indexes. In practical terms, the WP5 require each partnership to identify an appropriate and available geographical unit and national index, to obtain census data and to compute the EDI as it has been previously explained– but mainly in step 2 of EDI. As such, the first step was to identify a responsible contact for each country (Table 1).

Table 1: Group of experts for each country

Country	Contact
Italy	Roberto Lillini, Marina Vercelli
Spain	Marc Saez
France	Elodie Guillaume, Ludivine Launay
Portugal	Ana Isabel Ribeiro
Slovenia	Vesna Zadnik
Germany	Ron Pritzkuleit
Lithuania	Ieva Vincerževskienė
Poland	Krzysztof Czaderny - Joanna Didkowska
Nothern Ireland	Adrian Moore – Bruna Pucci

To reduce the risk of potential geographical bias, the deprivation index had to be calculated at the smallest geographical unit for all countries involved in WASABY project. Therefore, registries data would have to be geolocated as precisely as possible. Among registries in the WASABY project, there are two possible situations: data registries are either geolocated or not (Table 2 - details are available in the report entitled "D4.1 Survey results" – page 8-9, 21st June 2019). If not, the quality of the geolocation will depend on the quality of the registered address. If the address is precise, the geolocation via geographical coordinates would be facilitated and therefore using the deprivation index at the smallest geographical unit will be possible. For standardization purposes in this step a tutorial of good practice for geocoding and geolocalization was undertaken for partners (Annex 1)

The construction of an aggregated deprivation index is dependent on census data. Therefore, the geographical unit considered for each country will be the smallest geographical unit for which census data are available (and not necessarily the smallest geographical unit existing in the country). In this aim, a survey was sent to all partners in June 2018 to determine geographical unit existing in the country, the availability of census data and national index for each unit (Table 3). This survey is present in Annex 2, the response of the partners is compiled in table 3.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Table 2: Main geocoded level for cancer registries data involved in WASABY

Country	Geocode level (main)	Available indexes
France	X,Y coordinates, IRIS	EDI 2011
Germany	Municipality	GIMD 2006, GIMD 2010
Italy	X,Y coordinates, census block	EDI 2001 & 2011, national deprivation index 2001 & 2011
Lithuania	Eldership	EDI 2011
Poland	Municipality	National and Local Poland Index 2011
Portugal	Parish	EDI 2001, EDI 2011
Slovenia	X,Y coordinates	EDI 2011
Spain	X,Y coordinates, census block	EDI 2001
UKN	Post code	EDI 2001, EDI 2011, Local deprivation index

Table 3: Geographical unit and available index (listed)

Country	Geographical unit	Available indexes
France	IRIS	EDI 2011
Germany	Municipality	GIMD 2006, GIMD 2010
Italy	Census block	EDI 2001, EDI 2011, national deprivation index 2001 & 2011
Lithuania	Eldership	EDI 2011
Poland	Municipality	National and Local Poland Index 2011
Portugal	Parish, census block	EDI 2001, EDI 2011
Slovenia	Voting unit	EDI 2011
Spain	Census block	EDI 2001
UK(NI)	Small area	EDI 2001, EDI 2011, Local deprivation index

For all countries where the construction of EDI was possible, census data were prepared, firstly, to be in accordance with EUSILC data for step 2 of EDI and secondly to be in accordance with the final model computed at the step 3.3. It consists of calculating for each variable, the percentage according to the threshold obtained in step 2 (for example, percentage of people unemployed, percentage of people with no high diploma…). The preparation of the census data according to the final model was completed by the contact responsible or by WP5'leaders.

Co-funded by
the Health Programme
of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

3. Results

3.1 Geographical units

The EDI would had to be computed at different geographical units according to the availability of census data and the precision of geolocalisation of registries data, which depended on the availability of a precise address (Table 2) and explained in the methodology section above (Table 4).

Table 4: Geographical unit used in each country for the deprivation index calculation in the manner of WASABY project and its characteristics

Country	Geographical unit	Number of unit	Average population (min-max)
France	IRIS	50,867	1,277 (0-11,159)
Germany	Municipality	11,054	1,719 (9-3,469,849)
Italy	Census block	366,863	165 (0-7,647)
Lithuania	Eldership	546	5,434 (260 – 162,360)
Poland	Municipality	2,478	15,510 (1,302-1,764,615)
Portugal	Parish	4,260	2,479 (31-66,250)
Slovenia	Voting unit	3,104	660 (30-4,560)
Spain	Census block	525	3,384 (85-95,675)
UK(NI)	Census Small area	4,537	400(98-3,072)

3.2 European Deprivation Index

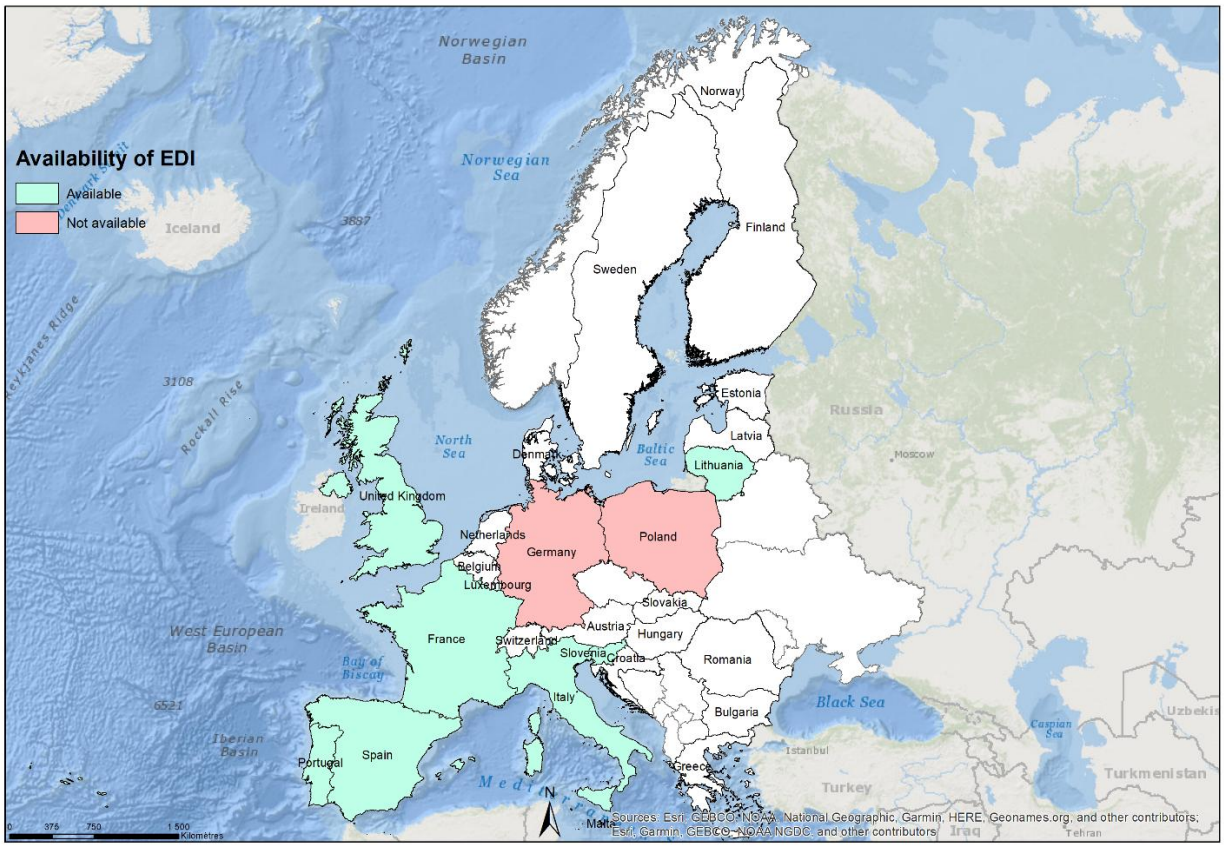
The EDI could not be calculated for 2 countries: Poland and Germany. For Poland, the number of common variables between the EUSILC and census datasets was less than those included in local index. There was therefore the potential that a Polish EDI would be less sensitive to deprivation in the country and for that reason it was decided to use the local index. For Germany, socioeconomic variables in the census are different between the Federal states (Länder) so it is impossible to compute an index for the entire country. Given that, it was decided to use a local index also for this country.

Finally, the EDI is available for 7 countries: France, Italy, Spain, Northern Ireland, Portugal, Slovenia and Lithuania (carte).

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Carte: availability of EDI in country involved in WASABY Project



For Northern Ireland, the combined EUSILC data for the United Kingdom was used. While observation data for Northern Ireland alone could be extracted there were too few observations to conduct the analysis and run the multivariate regression. So all steps computed using EUSILC data were completed using the UK data. For the final step, the Northern Ireland specific census dataset was used.

Step 1: Definition of an individual deprivation indicator

Step 1.1: Fundamental needs

Fundamental needs where those among the nine common european needs, where less than fifty percent of households do not possess it because they can not afford it, and were associated with both objective and subjective poverties.

Most of the fundamental needs selected in the first step were common at all countries (Table 5).

Table 5 reports for each need and each country, the percentage of households that could not afford the need. A need that is considered as fundamental is highlighted in bold, according to the previous definition.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Table 5: Selection of fundamental needs

Fundamental needs for people (%)	France	Italy	Portugal	Spain	UK	Slovenia	Lithuania
Meat or fish or vegetarian equivalent	**7.9**	**13.2**	**3.5**	**3.3**	**5.4**	**12.4**	**26.8**
One week annual holidays	**28.5**	**46.8**	50.3	**40**	**27.9**	**35.4**	52.9
Unexpected financial expenses*	**32.5**	**39.5**	**29.7**	**37.8**	**35.1**		64.1
Keep home adequately warm	**6.6**	**18.5**	**26.9**	**6.4**	**6.6**	**6.4**	**37.9**
Phone (including mobile phone)	**0.2**	**0.2**	**2.8**	**0.7**	**0.1**	**0.3**	**3.2**
TV	**0.3**	**0.4**	**0.7**	**0.2**	**0.2**	**0.7**	**0.8**
Computer	**3.9**	**3.4**	**8.8**	**5.2**	**3.4**	**5.6**	**11.9**
Washing machine	**1.1**	**0.5**	**2.5**	**0.3**	**0.5**	**0.5**	**5.2**
Car	**4.0**	**3.3**	**9.7**	**5.7**	**6.5**	**5.5**	**16.7**

For example, in Lithuania, “One week annual holidays” and “Unexpected financial expenses” cannot be afforded by more than fifty percent of households and were therefore not considered as fundamental. In Slovenia, more than fifty percent of households cannot afford to pay for “Unexpected financial expenses” so it too was not considered as a potential fundamental need and was not included.

Between all the countries countries we observe reasonable uniformity in the percentage of people who cannot afford a need, with the exception of “Meat or fish or vegetarian equivalent” which ranged from 3.3% for Spain to 26.8% for Lithuania. Lithuania seems to stand out from theother countries with a much higher percentage.

Step 1.2: Subjective poverty

The dichotomisation of subjective poverty was achieved according to modality 1 "with great difficulty" for Portugal and Lithuania and modality 2 which grouped together "with great difficulty" and "with difficulty" for France, Italy, Slovenia, Spain and UK. The percentage of households perceiving themselves poor and who are objectively poor was around 5% for UK, Lithuania, France and Portugal and closer to 11% for Spain, Slovenia and Italy (Table 6).

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Table 6: Percentage of household perceiving themselves poor and that are objectively poor

Country	Threshold for subjective poverty	Percentage of household perceiving themselves poor	Percentage of household perceiving themselves poor and that are objectively poor
France	2	25.5	6.0
Italy	2	37.2	12.5
Portugal	1	19.0	7.7
Slovenia	2	32.0	11.3
Lithuania	1	12.9	5.4
Spain	2	27.6	10.5
UK	2	15.1	5.4

Step 1.3: Construction of the binary individual deprivation indicator

The individual deprivation indicator in each country was constructed according to a different number of lacking fundamental needs due to financial constraints. There were 3 fundamentals needs for Lithuania and UK, 4 for Slovenia, 5 for Italy and Spain and 6 for France (Table 7). Most fundamental needs are common to all countries. We note that a 'TV' is never considered as a fundamental need, 'Phone' is considered as fundamental only in Portugal, and a 'washing machine' is fundamental in Portugal and Lithuania.

Households were defined as deprived if they were lacking at least 1 fundamental need among those selected for UK, Slovenia and Lithuania and 2 for France, Italy, Portugal and Spain. Between 19% and 38.4% of households were considered as deprived by the individual deprivation indicator.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Table 7: Fundamentals needs included in the definition of the individual deprivation indicator

Fundamental needs for people	France	Italy	Portugal	Spain	UK	Slovenia	Lithuania
Meat or fish or vegetarian equivalent	x	x	x		x		x
One week annual holidays	x	x		x	x	x	
Unexpected financial expenses	x	x	x	x			
Keep home adequately warm	x	x	x	x		x	
Phone (including mobile phone)			x				
TV							
Computer	x			x		x	x
Washing machine			x				x
Car	x	x	x	x	x	x	
Minimal number of fundamental needs lacking	**2**	**2**	**2**	**2**	**1**	**1**	**1**
% of deprived households	**25.5**	**37.0**	**19.1**	**32.4**	**30.6**	**38.4**	**32.8**

Step2: Variables available both at individual (EU-SILC survey) and aggregate levels (census)

Step 2.1: Selection of variables available in both databases

This step requires the identification of variables available in both the EU-SILC and in the census datasets. These variables have to be comparable and p identically defined. It is the most difficult and time consuming process. Sometimes, it is necessary to group together modalities to accomodate for less detailed data. It is also neccessary to understand the census data and methodological aspects of its construction. That is the reason why it is important to have a contact in the host country to ensure that there is no misunderstanding or confusion due to language barriers with the description of census variables. Nevertheless, it did not guarantee any misunderstanding on the part of the person responsable for its construction.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Step 3: Ecological deprivation index

The final regression model provides the weighting that will be applied to selected variables to compute EDI in the ecological step (Table 8).

The final variables included in the calculation of EDI for each country differ, and when they are shared they differ by the weightings they bring to the measure of deprivation. As described in the introduction, the methodology is the same but it allows for adaptation of the final index based on the economic and cultural situation of each country by measuring the same subjective and objective concepts of deprivation.

Table 8: Variables included in the final model and their weightings by Country.

Domains	Variables	France	Portugal	Italy	Slovenia	UK	Lithuania	Spain
Social exclusion	Foreign nationality	0,51	1,04		0,37			
	No country of birth				0,32	0,46		
	Crime/Vandalism							0.49
Household data	Overcrowding*	0,24	0,96			0,32		0.99
	Household composition*	1,13				0,70		
	Household size*	0,87			0,32	0,50	0,50	
Basic amenities of housing	No bath or shower	0,91		1,09	2,42			1.33
	No indoor flushing		1,73					
	No detached house					0,63	0,33	
Home ownership	Non-owner	1,1	1,19	1,01	0,22	0,82		0.73
Car	No car	0,95				0,79		1.74
Marital status	Not married				0,36	0,27	0,48	0.37
Employment status	Disabled at work							
	Unemployed	0,93	0,27	0,31				
Education level	No higher education*	0,94	0,51	0,92	0,87	0,64	1,15	1.30
Occupation	Status in employment*		0,62	0,63	0,55	1,32		0.95
	Occupation: low income occupations	**0,65**	**0,37**		**0,70**	**0,39**	**0,67**	0.62

The European Deprivation Index produces a numerical value for each geographical unit in a country (Table 9).

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Table 9: Distribution of European Deprivation Index in each country

Distribution of EDI (geographical unit)	Mean	Std	Min	P25	Median	P75	Max
France (IRIS)	0	4.56	-16.40	-2.76	-0.93	1.40	55.65
Italy (census tract)	0	2.09	-27.2	-1.27	-0.317	1.05	9.17
Portugal (parish)	0	2.98	-7.85	-1.87	-0.42	1.37	30.36
Slovenia (voting unit)	0.01	3.839	-7.43	-2.529	-0.847	1.555	40.32
Nothern Ireland (Small Area)	0	4.69	-8.66	-3.77	-1.002	3.54	13.92
Lithuania (eldership)	0	1.90	-7.16	-0.94	0.31	1.24	5.59
Spain (census block)*	0	4.19	-16.5	-2.6	-0.28	2.29	55.03

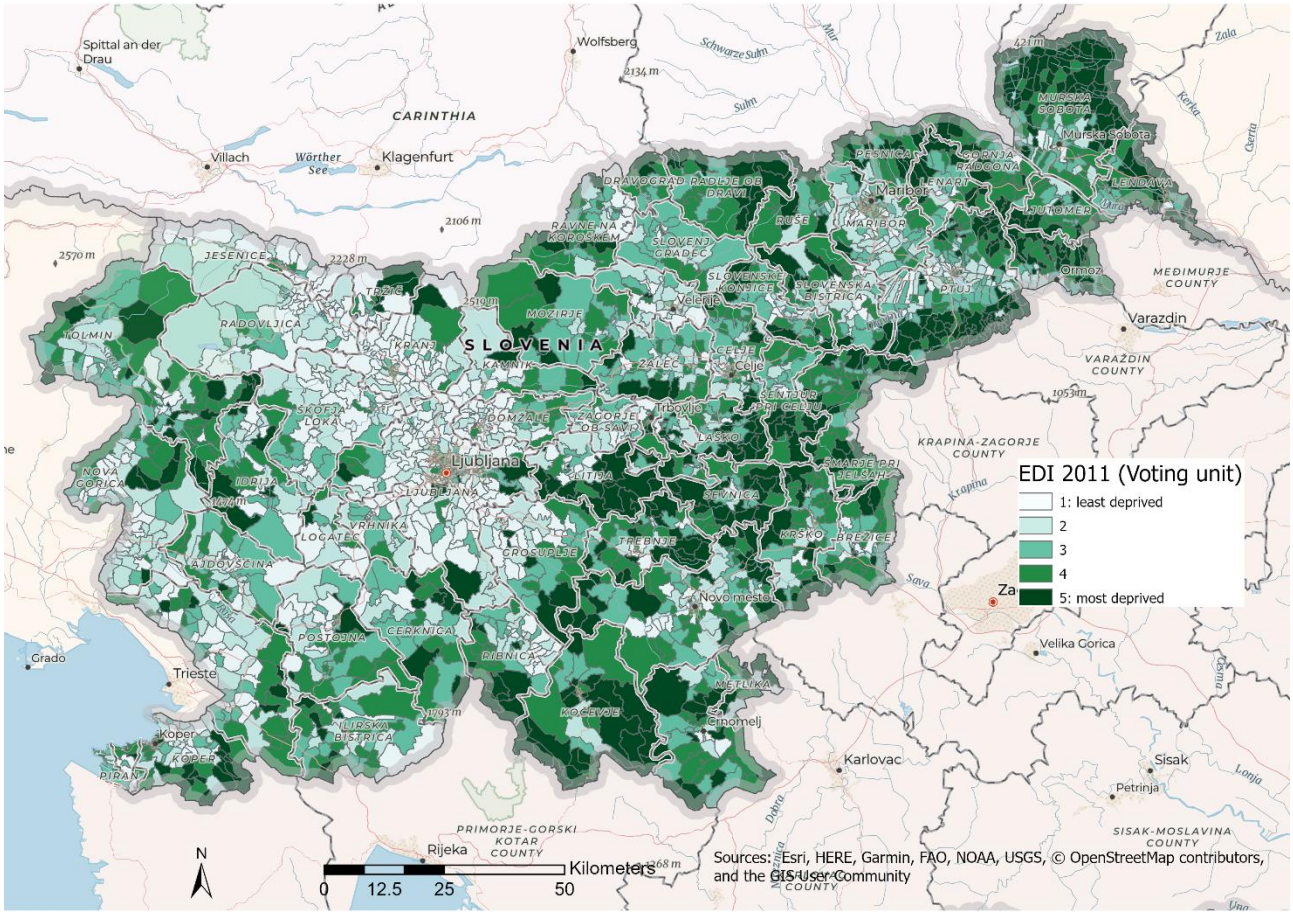
*based on 2001 census data

The European Deprivation Index value was classified according to the quintile of its distribution and mapped to highlights deprived area. Maps were obtained from the contacts in each country (with input from national statistical institutes) and incorporated into WP5. For Continental France, Continental Portugal, Slovenia, Northern Ireland and Poland, it has been completed by partners responsible for WP5. For Italy, it has been completed by Roberto Lillini and for Germany, by Werner Maier who computed the German IMD. Given that the indices have come from different sources and are themselves different, the colour schemes for mapping were not the same than those for EDI.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Slovenia

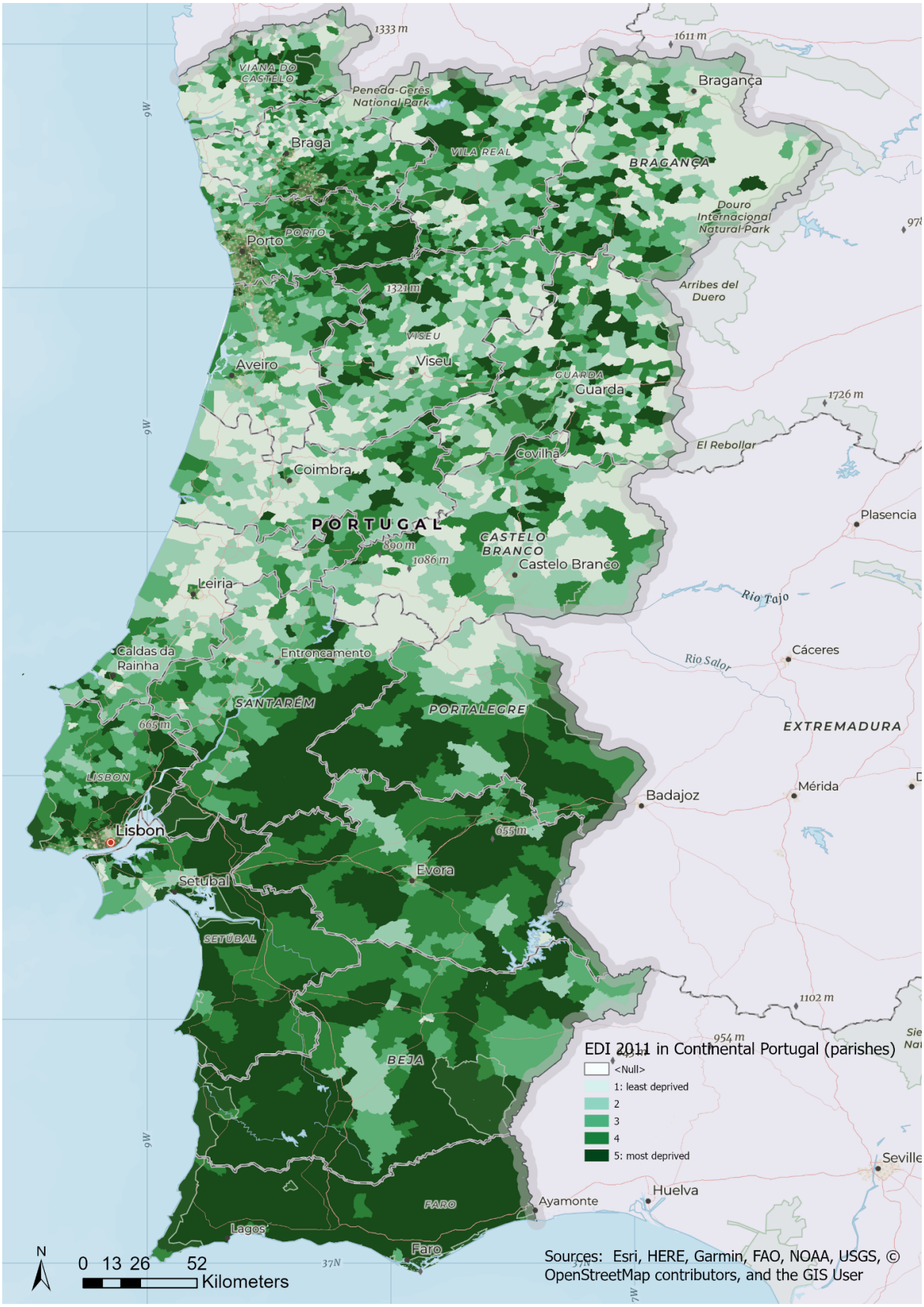


Deprived areas ar more concentrated in eastern Slovenia (along the border with Hungary) and in the South East of Slovenia (along the border with Croatia). Ljubljjana, the capital, is a heterogenous area incorporating some of theleast deprived areas and deprived areas (6) .

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Continental Portugal



Derpived areas are more concentrated in the south of Portugal, and in the North West near Porto (7).

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

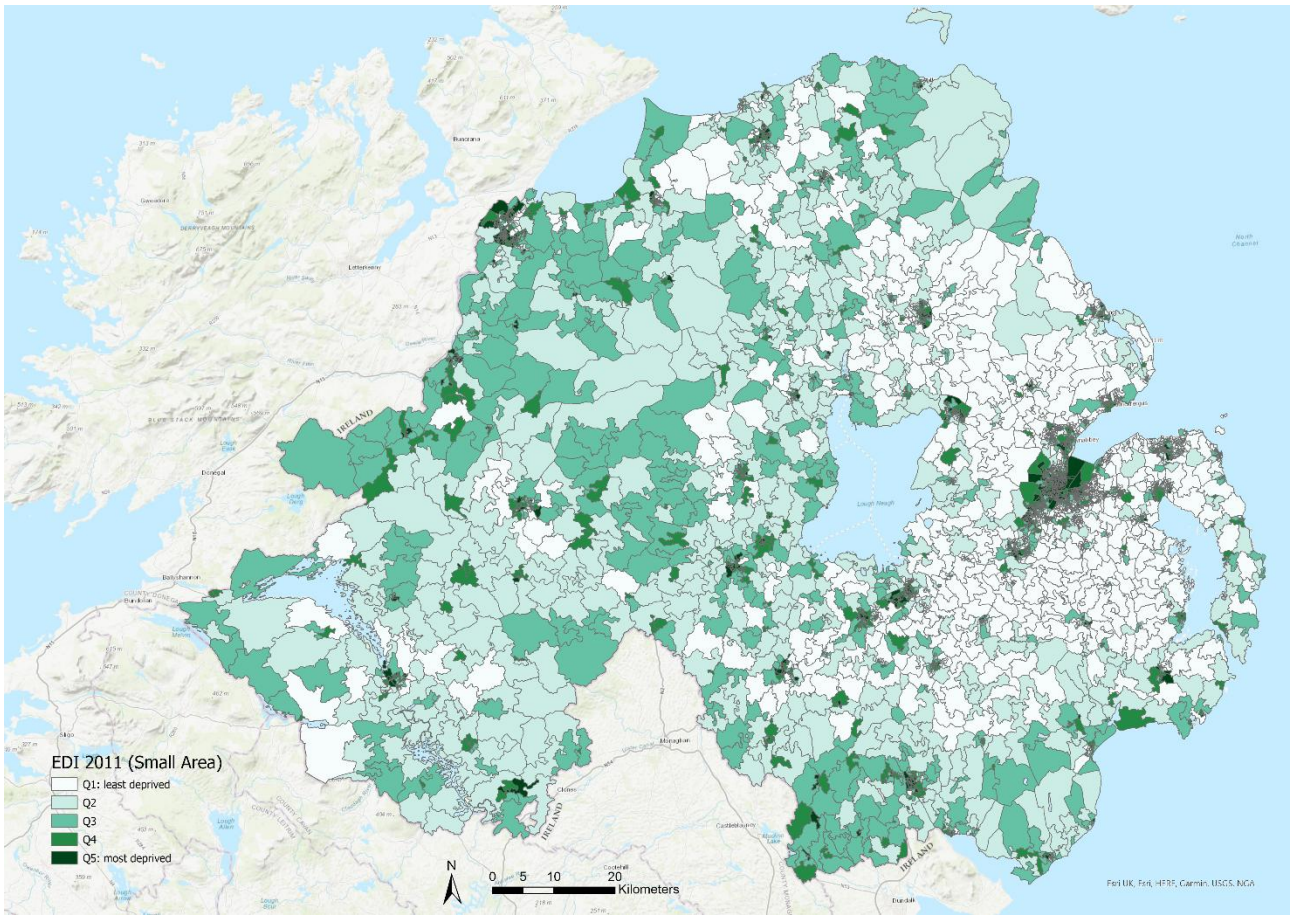
Continental France

Deprived areas are spread across continental France. In general, they are more concentrated in big cities (e.g. in Paris, the capital) and very rural IRIS. Corsica also has many deprived IRIS.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Northern Ireland



In Northern Ireland the most deprived areas are in the inner city areas of the two main cities, Belfast and Derry/Londonderry, as well as the more rural areas in the west and south of the province. The least deprived areas tend to be in small clusters within the cities and larger urban areas and in the commuter belt around Belfast.

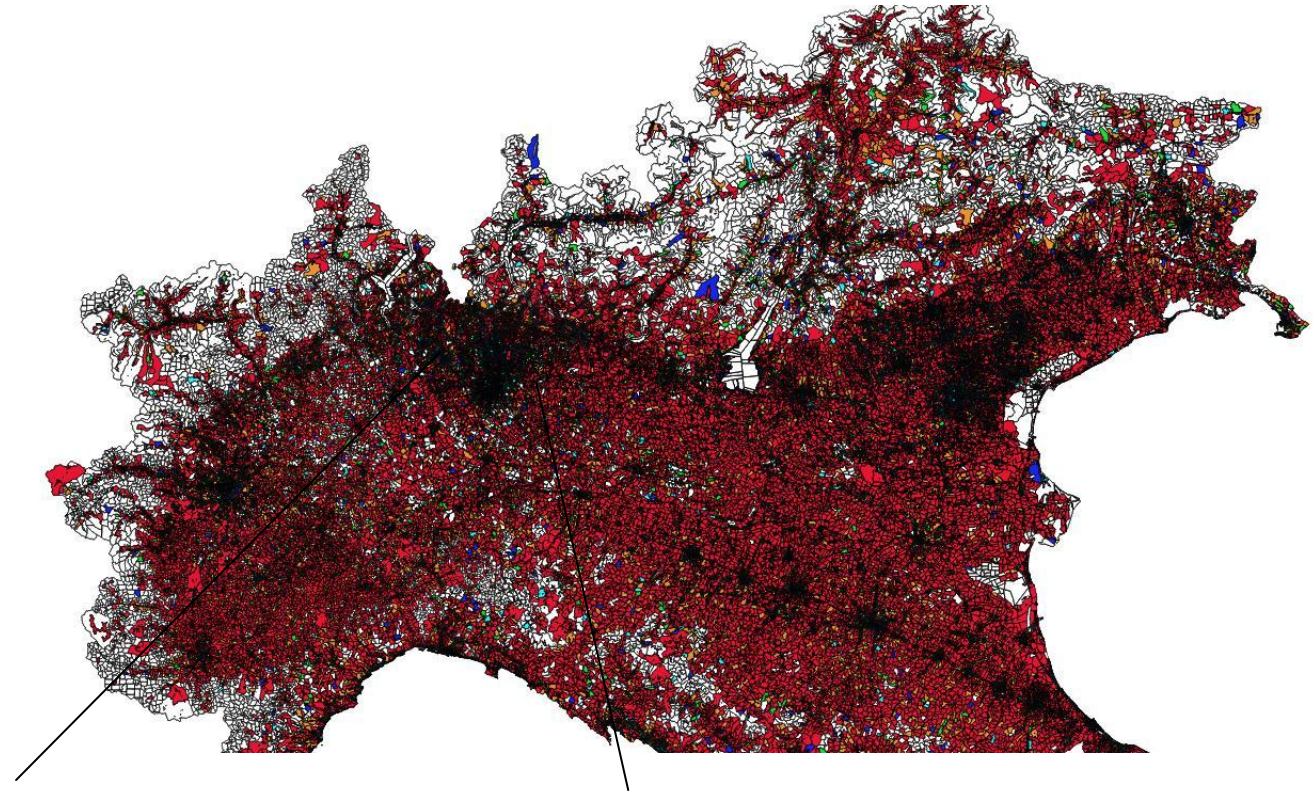
Italy

There is a problem in the Italian EDI geographic representation: the census units are too many and too small to be accurately presented on a standard page if the whole territory is reported. The resulting image is a not well defined (i.e. mostly gray territory with only some occasional spots surfacing). Therefore, a representation in three parts, North, Central and Southern Italy are presented in order to give a better description of the EDI distribution.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Northern Italy and Milan metropolitan area (zoomed)



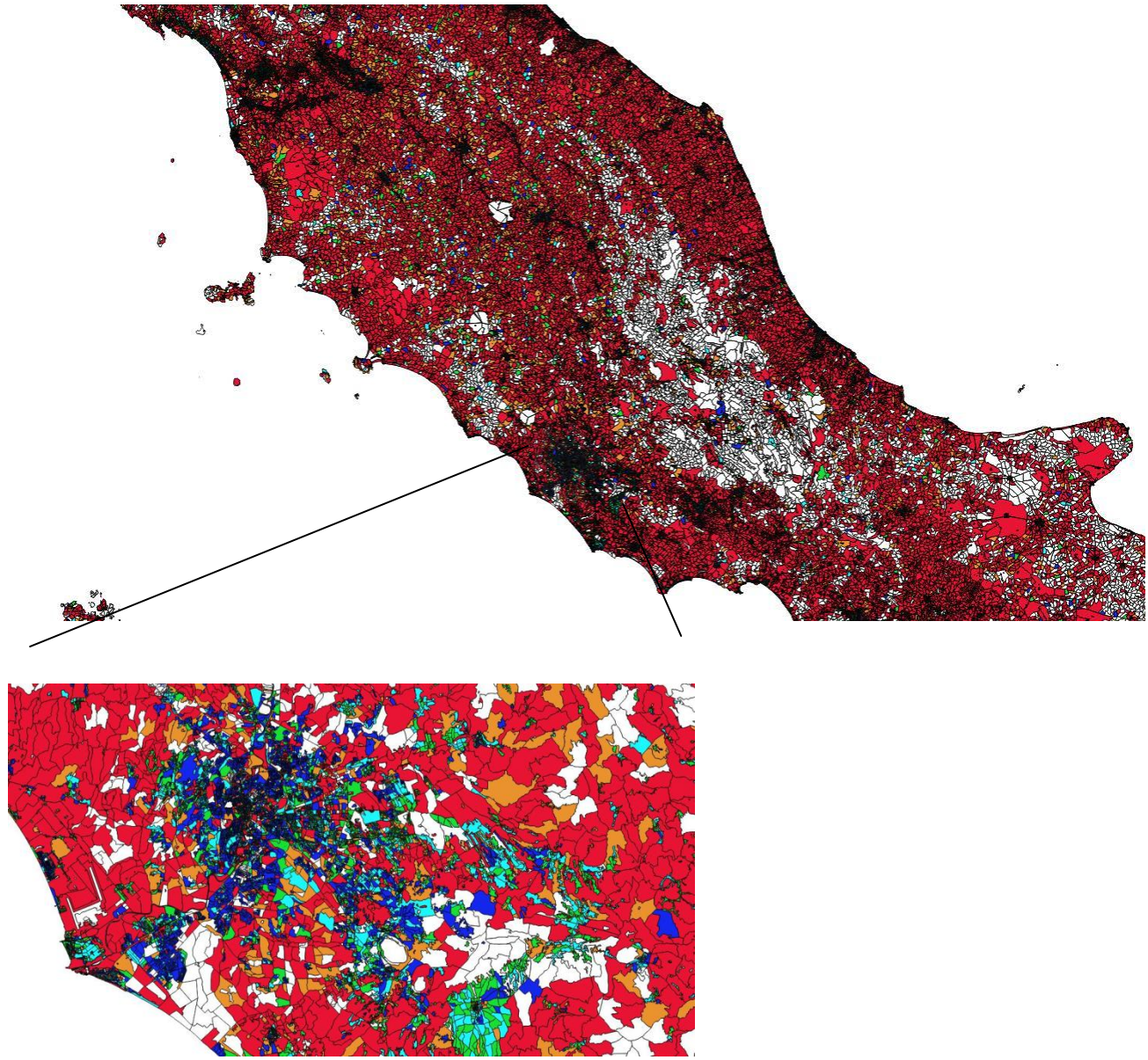
1: least deprived
2
3
4
5: most deprived

The less deprived (groups 1, 2 and 3) are mainly concentrated in the medium-to-large urban areas: in the map they show as “black spots” (due to the great number of CTs to be represented). A zoom on the Milan metropolitan area shows this aspect of the EDI distribution. Red and orange areas are distributed mainly in the rural areas; white areas are CTs where2011 Census data were not collected (for various different reasons).

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Central Italy and Rome metropolitan area (zoomed)

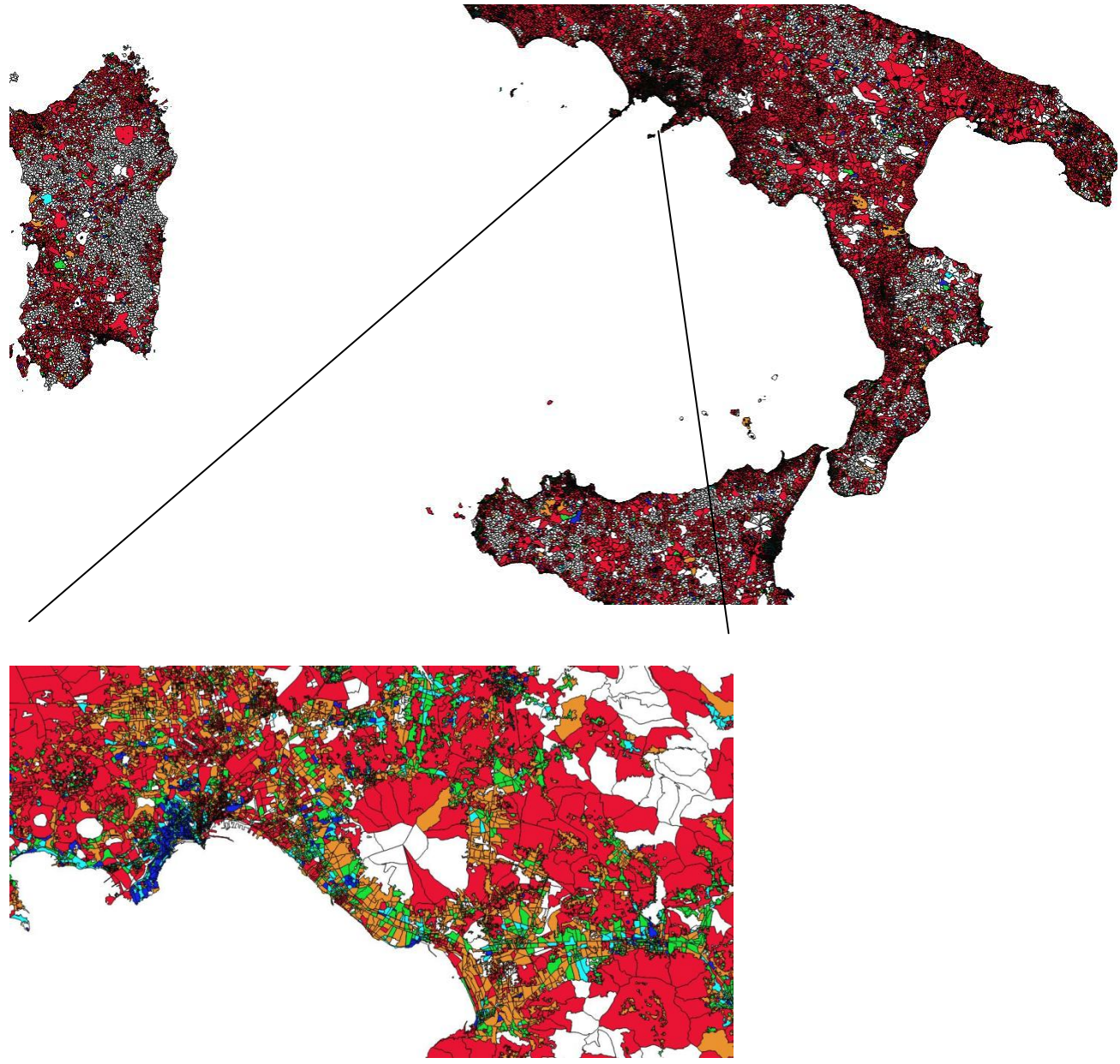


Similar results can be seen for the central part of Italy. Also here, the red and orange CTs are mainly present in the rural areas, where connection with the largest municipalities (i.e., Florence, Ancona, Perugia or Rome) are more difficult due to the presence of mountain ranges all along the peninsula.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Southern Italy and Naples metropolitan area (zoomed)



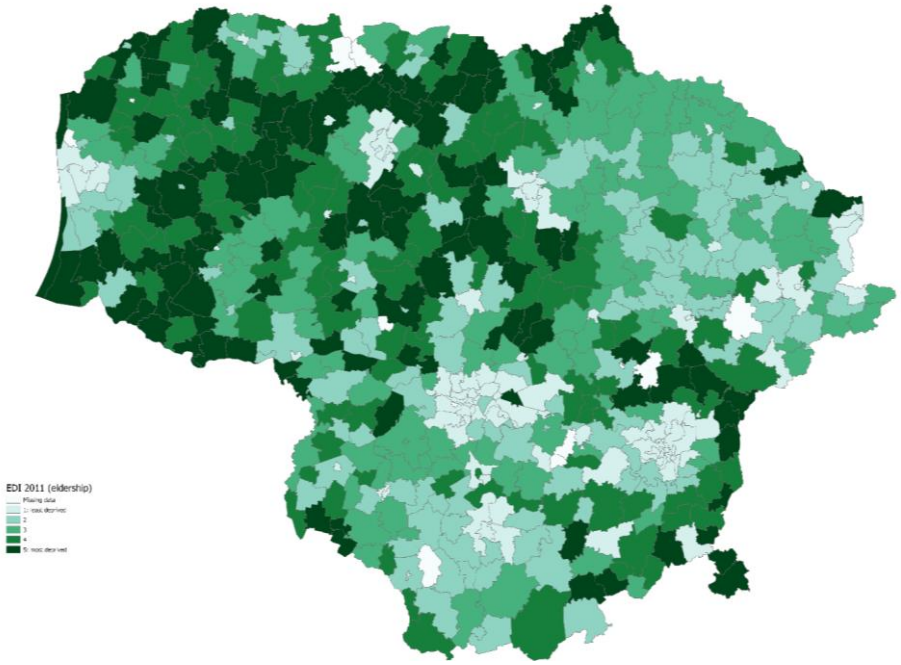
1: least deprived
2
3
4
5: most deprived

The Southern area of Italy and the two largest islands (Sicilia and Sardinia) present with a similar situation as the other parts of Italy. It should be noted that, for this Census, an issue arose whereby this region had the highest number of CTs where data were not collected.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Lithuania



Derpived areas are more concentrated in the west of Lithuania.

Spain

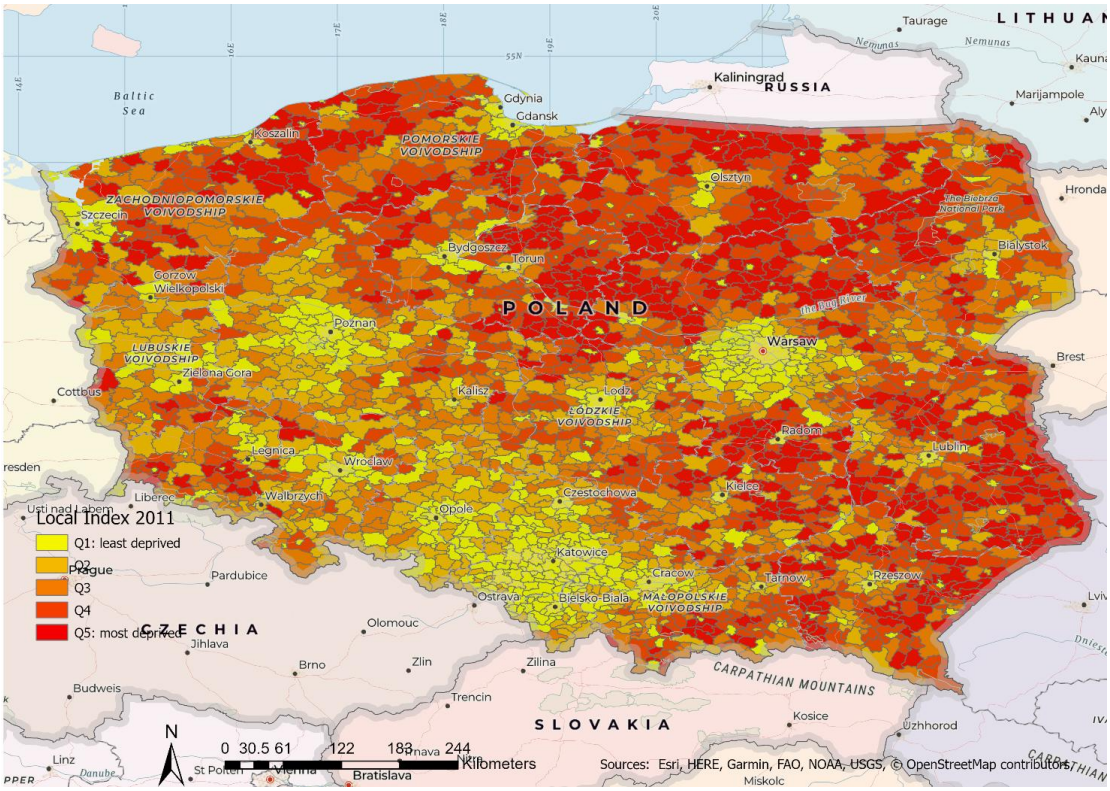
As shapefile at census tract was not obtained, EDI would not be mapped for Spain.

Co-funded by
the Health Programme
of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

3.3 Local Index

Poland



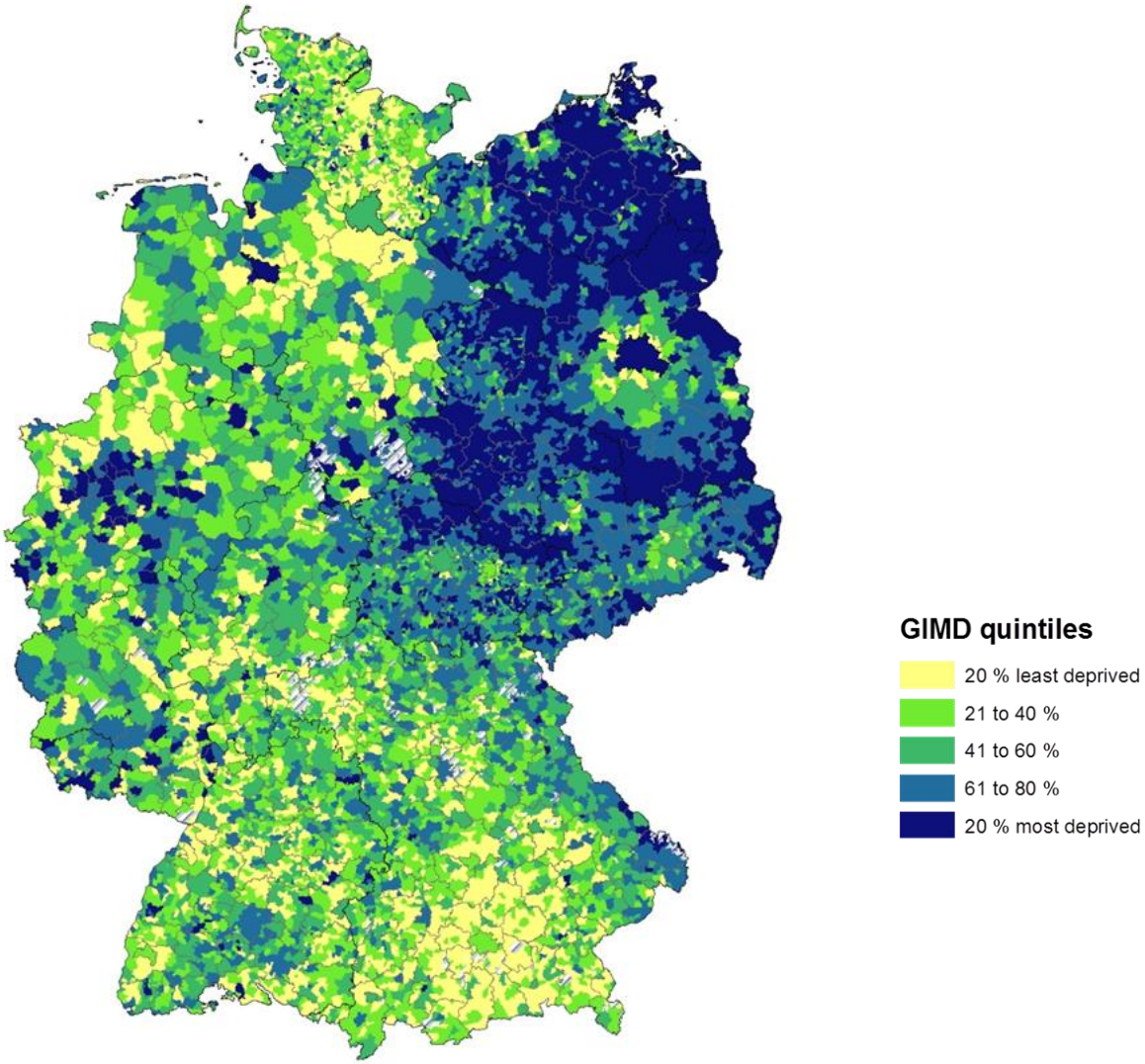
More deprived areas are located in the north of Poland and in the west, around Lublin, Rzeszow, Radom. At municipality level, Warsaw, the capital, and Poznan, Szczecin, Opole and Ostrava are among the least deprived cities.

Co-funded by the Health Programme of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Germany

The German Index of Multiple Deprivation (GIMD) at municipality level



Most deprived municipalities are mainly located in the North west of Germany. Berlin, the capital, is among the most deprived areas as is Düsseldorf (in teh mid-eastern region) . Both are surrounded by some of the least deprived municipalities. Among the least deprived areaa, are Hamburg in the north and Munich in the south.

Co-funded by
the Health Programme
of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

4. Conclusions

WASABY is a feasibility project based on a geographical analysis of population-based cancer incidence data in connection with environmental factors. Deprivation indexes have been used as cofounders. WP5 has identified a group of experts and contacts in each country involved in the project. Available data were compiled to evaluate the feasibility of the construction of a European Deprivation Index in each country. If this construction was not possible a national index has been used instead. Even if the methodology of the European Deprivation Index, based on a European-wide survey and national census data, could have been developed to be transferable/replicable in all European countries, some difficulties still arise. These difficulties are various and include issues around the available geographical units, availability of data but also with communication between partners.

With regard to the geographical unit, as it was explained in this report, it has to be the smallest unit possible to reduce the ecological bias (i.e. the error that arises when a deprivation index from aggregated data is used as a proxy of individual socioeconomic position). This geographical unit is dependent on the country concerned and there is a considerable heterogeneity between them. Even though the geographical units are heterogeneous, the EDI is constructed in the same way in all countries. As there are considerable differences in the populations in geographical units between countries, there is the potential for heterogeneity in the ecological bias and this raises the question of the comparability of results between countries. This comparability is more difficult between countries like Germany or Poland were the EDI could not be calculated. In fact, the concept of relative deprivation on which EDI is based, is not the concept used by the other indexes and as such, indexes are not directly comparable. Census data are not always free to access and can be available only in the native language. Another problem concern the effects of the EU-SILC survey for Northern Ireland, within the total UK survey. This highlights the importance of having within country partnerships with someone legitimate to ask this data. Moreover, for countries like Slovenia for example, where data can be available at the individual level, the requirements for data protection and security do not allow extraction of data to another computer other than those present in the office of national statistics. It implies that someone in the country, authorized by the national institute, goes and complete the final step in their offices. Moreover, as the deprivation index is compiled at an aggregated level using census data, this unit is constraint by it. The smallest geographical unit for which the deprivation index can be computed is the smallest geographical unit for which census data are available. Moreover, using indices at an aggregated area level requires data geocoding at this scale and so, for some units (smaller than municipalities), this means having availability of very precise geographical coordinates. Some registries already have this but for some of them, additional geocoding had to be undertaken. For those having geographical coordinates from available geocoding, it was possible to link EDI at the smallest unit for which census data was available. For those not having it (due to a lack of precise addresses), the geographical unit considered for EDI was constrained and not necessarily the smallest unit aligned with census data.

Finally, the methodology of EDI is complex and is further challenged when language barriers are present. It is not always clear that the methodology is as fully understood as it should be and it highlights the importance of good communication at each step of the process to be sure that what is undertaken is done so in accordance with the appropriate methodology and protocols for the index in each country.

Co-funded by
the Health Programme
of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Bibliographie

1. Pornet C, Delpierre C, Dejardin O, Grosclaude P, Launay L, Guittet L, et al. Construction of an adaptable European transnational ecological deprivation index: the French version. J Epidemiol Community Health. nov 2012;66(11):982‑9.
2. Guillaume E, Pornet C, Dejardin O, Launay L, Lillini R, Vercelli M, et al. Development of a cross-cultural deprivation index in five European countries. J Epidemiol Community Health. mai 2016;70(5):493‑9.
3. Noble M, Wright G, Smith G, Dibben C. Measuring Multiple Deprivation at the Small-Area Level. Environ Plan Econ Space. 1 janv 2006;38(1):169‑85.
4. Maier W. [Indices of Multiple Deprivation for the analysis of regional health disparities in Germany : Experiences from epidemiology and healthcare research]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. déc 2017;60(12):1403‑12.
5. Schederecker F, Kurz C, Fairburn J, Maier W. Do alternative weighting approaches for an Index of Multiple Deprivation change the association with mortality? A sensitivity analysis from Germany. BMJ Open. 2019;
6. Zadnik V, Guillaume E, Lokar K, Žagar T, Žakelj M, Launoy G, et al. Slovenian version of the european deprivation index at municipal level. Slov J Public Health. 6 avr 2018;57(2):47‑54.
7. Ribeiro AI, Launay L, Guillaume E, Launoy G, Barros H. The Portuguese version of the European Deprivation Index: Development and association with all-cause mortality. PLoS ONE. 5 déc 2018;13.

Co-funded by
the Health Programme
of the European Union

EU funding disclaimer: This project has received funding from the 3rd European Union Health Programme under Grant Agreement PP-2-5-2016 (# 769767)

Annex 1: Geocoding guide

Background

Many studies have shown a link between health indicators (incidence, survival, care…) and social inequalities but the underlying mechanisms remain unknown. To better quantify these inequalities and compile results in a comparable manner between countries, it is important to measure the social environment in the same way. Measuring social environment requires using the smallest geographical unit for which census data are available to limit the potential for ecological fallacy. This smallest unit is often determined by the accuracy of geographical coordinates which is why geocoding, which enables correspondence between an address and x,y coordinates, is often necessary.

This is a guide to help you to geocode your addresses for WASABY Project.

What is geocoding?

Geolocalisation consists of making a link or cross-reference between an address and its geographical coordinates. For example, the Cancer Centre for Normandy – Centre François Baclesse – located 3 av général Harris, 14076, Caen, France has for geographical coordinates (49.203529, -0.354513) in the WGS84 coordinate system.

Why geocoding?

Having precise coordinates for patients included in a study allows the inclusion of information at all geographical units: from the smallest to the largest unit (e.g. IRIS for France). It is necessary to evaluate accessibility to health care centers or professionals (for example using distance) and to undertake environmental analyses which aim to determine the effect of some pollutants on health and to calculate an ecological deprivation index. In our previous example, geographical coordinates correspond to IRIS number 141181404.

What is needed?

Precise address

Information required is a precise address – registered by registries - with house number, street type, street name, postal code, municipality or city (other code specific to country, for example in France, insee code which is specific for each municipality).

Be aware that this information is considered as directly identifying individuals and is covered by national data protection authorities so specific authorization could be required.

All this information has to be either in a common field or in separate field according to software you will use.

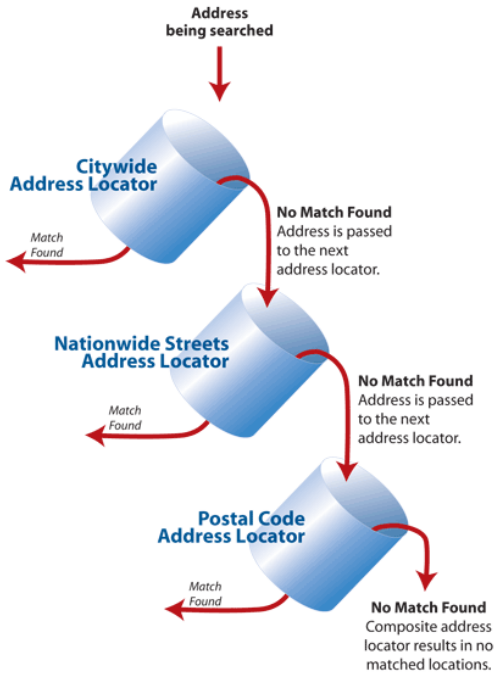
- Geographic Information System (GIS)
- The most famous commercial GIS are Mapinfo® (Pitney bowes) and ArcGIS® (ESRI). QGis is a GIS freely available and is more and more used by researcher or collectivity (https://qgis.org/en/site/forusers/download.html). There is also qVSIG or GRASS GIS (the list is not exhaustive). As QGis is widely used there are a lot of tutorials on the internet and a lot of forums about a variety of topics. It should be the most appropriate for discover GIS.Maps

To make the link between an address and geographical coordinates, maps and mapping software are needed. Some mapping solutions are commercialized by the most popular proprietory gIS companies (for example ESRI) but the price can be expensive according to the product you need. Free data are also available, for example, from openstreetmap.

How does it work?

Process is conducted a step by step cascade (see figure below). Addresses are used with all information (number, type, name, postal code, locality). If an exact match is found, the software goes to the next address. If not, the next location level (using only type, street name, postal code and locality) is considered and so on…

Figure: illustration on geocoding process



Source: http://help.arcgis.com/fr/arcgisdesktop/10.0/help/index.html#//00250000003r000000

So, geolocation can be conducted at different levels according to the quality of information used or available:

- Level 1: Number, type, street name, postal code, locality
- Level 2: Type and street name, postal code, locality
- Level 3: Postal code, locality so municipality level which often is the city hall

Caution

Be careful of:

- The coordinate system (the repository in which elements are represented in space) you will use. It has to be the same between all maps used (if not, maps will not be stackable). The coordinate system depends on your country but you can use ETRS89 Lambert Conformal Conic Coordinate Reference System which is used by Eurostat.
- The geocoding level you will use: some geographical units could not be used in accordance with the aim of the study. For example, geocoding to municipality will not be relevant for a municipality that is divided into smaller geographical units (for example, in France with IRIS).

The example of the procedure performed by Caen

GIS used is ArcGIS 10.7 provided by ESRI France. Maps for geocoding are BD Adresse for ArcGIS Advanced® (provided by ESRI France and IGN). Geocoding process has been adapted for social inequality studies. To limit ecological biases in inequality studies, it is important to use the smallest geographical unit available. In France, the smallest geographical unit for which census data are available is IRIS so the European Deprivation Index will be used at this unit. Municipalities with less than 5,000 inhabitants are an IRIS by themselves. Municipalities with more than 5,000 inhabitants are divided into IRIS. Therefore, addresses coming from municipalities with more than 5,000 inhabitants could not be located at street name level (if the street crosses more than one IRIS) or municipality level.

There are different steps in geocoding process:

- preparation of address with formatting street name, locality (to optimize automatic geocoding) programming with SAS® 9.4 Software
- quality control of original data (for example, locality in accordance with insee code) programming with SAS® 9.4 Software,
- preparation of data to allow differentiation according to locality type (more or less than 5,000 inhabitants) in GIS programming in python,
- geocoding (mainly correction of addresses with no correspondence)
- allocating the corresponding IRIS for all coordinates
- geocoding quality control : addresses automatically matched are in accordance with original locality as with IRIS
- adding European Deprivation Index according to geographical unit (mainly IRIS)

Annex 2: Survey sent to the group of expert to determine the best geographical unit to construct EDI, the availability of census data and deprivation indexes

Country: _____

Contact name: _____

As you know, your country is involved in WASABY Project which aim to identify areas with higher cancer rates, so to study whether pollutant contamination may be a cause for increased cancer risk. As part of Work Package 5 (construction of the European Deprivation Index, EDI), we need some information to determine the best geographical unit to compute EDI for your country. Indeed, to limit ecological bias, EDI have to be computed at the smallest geographical area for which census data are available. This area will determine the geocoding level for registries data for each country. That is why we sent you this survey in addition to the one sent by WP4. Do not hesitate to contact people who can answer to these questions in your country or to indicate the more pertinent contact to obtain these informations.

About geographical area

For some countries, socioeconomic data could be not available in national census data but available in regional statistics as in Germany. So we need to investigate the availability in census of such information for the different geographical unit.

Can you cite the different geographical area (administrative and non-administrative) in your country (for example, district, municipality, census block…). Cite them by the smallest to the largest scale. For all of them, precise if census data are available. Provide a little description if some area are not disposable for all units (for example in Germany, administrative region does not divide all Bundesländer). If some geographical area are dependent of the number of population (or another variable), please precise it in description (for example in France, IRIS concerned only all the municipalities with more than 10,000 people, a part of the municipalities with more than 5,000 inhabitants and less than 10,000; municipalities with less than 5,000 inhabitants are not divided and are an IRIS by themselves).

For example, in France:

Geographical area	Description	Number of units	Mean	min	max	Corresponding NUTS	Availability of the census
Region						2	Yes
Department						3	Yes
Municipality		36,664	1,443	0	1, 926,595	LAU	Yes
IRIS	All municipalities with more than 10,000 inhabitants, a part of municipalities with more than 5,000 inhabitant and less than10,000; municipalities with less than 5,000 inhabitants are not divided and are an IRIS by themselves	50,867	1,277	0	11,159		Yes
Grid							Little information

			Population					
Geographical area	**Description (if necessary)**	**Number of units (if available)**	**Mean (if available)**	**Min (if available)**	**Max (if available)**	**Corresponding NUTS* (if available)**	**Availability of the census**	**Socioeconomic indicator available**

***The NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the economic territory of the EU: NUTS 1=major socio-economic regions, NUTS 2 = basic regions for the application of regional policies, NUTS 3: small regions for specific diagnoses. To meet the demand for statistics at a local level, Eurostat maintains a system of Local Administrative Units (LAUs) compatible with NUTS. These LAUs are the building blocks of the NUTS, and comprise the municipalities and communes of the European Union. (http://ec.europa.eu/eurostat/web/nuts/background; http://ec.europa.eu/eurostat/web/nuts/local-administrative-units).

About census

What is the name of the institute in charge of it: _____

What is its website: _____

Is the census exhaustive?

□ Yes

□ No, Precise the size of the survey sample: _____

What is the type of census?

□ Traditional

□ Rolling

□ Sample based administrative data based

□ Sample-based

□ Other (Precise:_____)

Is it conducted annually?

□ Yes

□ No. Please provide the specific timing:_____

Does it include socioeconomic data?

□ Yes

□ No. Precise, if these data are available by another mean and how: _____

Reference

If there is some reference that can complete these information, please, write it:

1. _____
2. _____
3. _____
4. _____