METHODOLOGICAL NOTE FOR THE APPLICATION OF POPULATION DATA AT MUNICIPALITY SCALE (LAU2) IN THE EUROPEAN UNION (EUROSTAT DATA)

CRISP/UNA (*Leading author: F.A. Mileti*)

INTRODUCTION

The population survey at municipality level (LAU2) is very important for the LANDSUPPORT project. Indeed, demographic data is important (i) to describe the social context of European municipalities and (ii) is required by several of our decision support tools. For example, the Land Take tool requires calculation of the population between given years and comparison of this data with the increase in urbanized surface over the same time period.

To obtain this information at the municipal level (LAU2) on a European scale, it was necessary to associate the data from the LAU2 spatial geometries - which are typically visualized and processed on GIS platforms - with the surveyed population data.

A first screening of EUROSTAT data highlighted the high complexity of the challenge, which is due to a lack of coherence between the information layers obtained for the different member states. These problems were tackled and partially resolved through a gradual, slow and cautious approach that took two months' work.

METHODS

For data processing, the ArcGIS 10.2 software was used.

To allow association of the population data (Excel tables) with the file containing the geometry of the municipalities (shapefile), it was necessary to use a very common procedure in the GIS environment known as "join". By using codes (numeric, textual or alphanumeric) contained in a specific column present in two different tables (for example the data contained in the "code" column of Tables A and B of Figure 1), the join operation allows the merging of data from a hypothetical table A with those from a table B, so generating a table C such as the "Result" in figure 1.

A

City	Code	Population (milion)
Rome	1000	2,873
London	2000	8,136
Paris	3000	2,2

B

City	Code	Area (km2)
Rome	1000	1.285
London	2000	1.572
Paris	3000	105,4

Result

City	Code	Area (km2)	Population (milion)
Rome	1000	1.285	2,873
London	2000	1.572	8,136
Paris	3000	105,4	2,2

Figure 1 – Example of join procedure

An initial data analysis highlighted a series of specific problems in each country (these problems will be dealt with in more detail below). This made the implementation of a single, automatic join procedure for all member states impractical.

To solve these problems and have greater control over the results, the first phase of the work involved segmentation of the shapefile using municipal geometries referring to the year 2013 for the entire European Union (freely downloadable at: https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/communes#communes13), which generated a shapefile for each member state. For the Austrian case, in order to minimize topological coherent problems, **geodata was obtained by data.gv.at, 2017.**

Each shapefile has been associated, using join procedures, with population data for the years 2012, 2015, 2016, 2017, 2018, the years of interest for the Land Take tool (freely downloadable at: https://ec.europa.eu/eurostat/web/nuts/local-administrative-units). Austrian population data **was obtained from Statistik Austria.**

The best practice would have been to use geodata referring to the same years as the population data. For example, the 2012 shapefile for population data collected in that year and so on. This is because the municipal geometries might vary from one year to the next due to administrative reorganization. Associating information on population with municipal geometries for a different year may generate errors of over or underestimation. In our case, the use of geodata referring to the year 2013 is a consequence of the impossibility of finding geodata for the whole European area in the years of interest mentioned above.

For the case study of Austria, it has been possible (thanks to EAA) to find population data, for the years 2012, 2015, 2016, 2017, 2018, referring to geodata downloaded from data.gv.at, 2017. In comparison to the standard procedure, this allowed to overcome the limitation of having the "no data" records. It is desirable that in the near future the UE population dataset will ameliorate as it was in the case of Austria.

Generally, the fields used to associate the municipal geometries with population data have been NSI_CODE and LAU_CODE; however, as already mentioned, there have been several cases for which it was necessary to proceed with join methods adapted to the specific situations in the individual countries (Annex 1).

RESULTS

Five information layers were produced as a shapefile. Each of them described the European population in a specific year (2012, 2015, 2016, 2017 and 2018). These years were chosen in relation to the use of this data in the Land Take tool, which provides soil imperviousness data for the same periods of time.

Within the same shapefile, there are two levels of spatial detail: LAU2 and NUTS2, as shown in table 1.

This is mainly linked to the unavailability (or, due to incomplete available data, the difficulty of application) of population data at the LAU2 level. When the LAU2 data was not available and/or applicable for a particular year in a particular country, the NUTS2 data was applied in its place with the relative geometries (available and downloadable on the EUROSTAT website). The LAU2 resolution data were implemented for 63% of the countries in the Union on average (table 1).

The problems for which it was necessary to use the NUTS2 data are attributable to 4 main causes:

(i) the lack of population data;

(ii) the disproportion between the population data collected / available in comparison with the actual number of LAU2 units;

(iii) the difficulty of finding a shapefile with the LAU2 units which was suitable for association with the EUROSTAT population data (this problem was only for Greece);

(iv) other association problems (join) between population data and spatial data.

Table 2 shows these issues for each member state in each year considered in detail.

The green fields in Tables 1 and 2 have population data at LAU2 resolution and might present the following problems:

(i) Presence of municipalities with zero population. This problem refers to the unavailability of data for some specific municipalities or to the lack of join between data and geodata

in that particular row of the table. This is an easily identifiable and quantifiable error (for example by ordering the table with respect to the population data).

(ii) Errors in the attribution of data to the geodata. This may be due to a variety of causes, one of which is that several sectors of the same city may be registered using the same LAU code (see attachment, the case of the city of Bucharest), so generating enormous overestimates. Other causes might be the mergers and divisions of municipalities. These types of errors are less easily identifiable than those in point (i).

Table 1 - Resolution NAU2 / NUTS2 for all countries in the five years considered - in green the countries with LAU2 resolution, in yellow those with NUTS2 resolution, in red the countries with NUTS2 resolution but with no population data

Country	ISO	2012	2015	2016	2017	2018
Austria	AT					
Belgium	BE					
Bulgaria	BG					
Cyprus	CY					
Croatia	HR					
Denmark	DK					
Estonia	EE					
Finland	FI					
France	FR					
Germany	DE					
Greece	EL					
Ireland	IE					
Italy	IT					
Latvia	LV					
Lithuania	LT					
Luxemburg	LU					
Malta	MT					
Netherlands	NL					
Poland	PL					
Portugal	PT					
United Kingdom	UK					
Czech Republic	CZ					
Romania	RO					
Slovakia	SK					
Slovenia	SI					
Spain	ES					
Sweden	SE					
Switzerland	CH					
Hungary	HU					

Table 2 - problems relating to the implementation of the LAU2 resolution data. Abbr. Pop= population

Country	ISO	2012	2015	2016	2017	2018
Austria	AT					
Belgium	BE					
Bulgaria	BG				Few Pop Data	Few Pop Data
Cyprus	CY	No Data		No Data	No Data	No Data
Croatia	HR	No Data				
Denmark	DK	Few Pop Data			Few Pop Data	Few Pop Data
Estonia	EE					Few Pop Data
Finland	FI					
France	FR		No Data	No Data		No Data
Germany	DE		No Data			
Greece	EL	Few Geo Data	Few Geo Data	Few Geo Data	Few Geo Data	Few Geo Data
Ireland	IE	No Data	No Data	No Data	No Data	No Data
Italy	IT					
Latvia	LV	Join Problems	Join Problems	Join Problems	Join Problems	Join Problems
Lithuania	LT	No Data	No Data	No Data	Few Pop Data	Few Pop Data
Luxembourg	LU					
Malta	MT					
Netherlands	NL					
Poland	PL					Join Problems
Portugal	PT					No Data
United Kingdom	UK	Join Problems	Join Problems	Join Problems	Few Pop Data	Few Pop Data
Czech Republic	CZ					
Romania	RO					
Slovakia	SK					
Slovenia	SI	Few Pop Data	Few Pop Data	Few Pop Data	Few Pop Data	Few Pop Data
Spain	ES		No Data	No Data		No Data
Sweden	SE					
Switzerland	CH	No Data	No Data	No Data		
Hungary	HU					

Legend	
	LAU2 population detail data
No Data	Unavailability of the LAU2 population detail data
Few Pop Data	The population data collected at LAU2 detail are too few compared to the actual number of municipalities
Few Geo Data	Availability of population data to detail LAU2, but difficulty in finding a shapefile with the same spatial resolution
Join Problems	Difficult diversification of join between geometry and population data

CONCLUSION

The production of population data for the EU in the years 2012, 2015, 2016, 2017 and 2018 is new and important for LANDSUPPORT and for the Land Take tool.

It was obtained through a collage of many small elaborations due to the impossibility of finding "key data" to solve problems which were mainly connected to the join between the different data. The works, by virtue of their *digital craftsmanship*, present various critical points to be taken into account, but, for the same reason, they are also in need of further improvement.

ANNEXES (some small examples)

The case of Hungary

One of the join problems was found in Hungary. The NSI_CODE field in figure 1 does not correspond to the LAU CODE field (Fig. 2). It was necessary to truncate the LAU CODE data (figure 2) and verify that the truncating operation did not generate duplications. The truncated data in FIG. 2 (LAU CODE truncated) is now suitable to support the join operation with the NSI_CODE datum of FIG. 1.

This issue regards Hungary for the years 2018, 2017, 2016 and 2015.
The data from Hungary in 2012 present a further, different, problem: they (column LAU2_NAT_CODE in figure 3) are not suitable to support the join operation with the NSI_CODE datum of figure 1; consequently, a zero must be added before the 3-digit number (column LAU2_NAT_CODE modified, Fig. 3).

Table

COMM_RG_01M_2013

FID	Shape *	COMM_ID	Shape_Leng	Shape_Area	OID	COMM_ID_1	CNTR_CODE	NSI_CODE	COMM_NAME
7923	Polygon	HU5121160	0,153858	0,000726	89973	HU5121160150	HU	0150	Ipolyszög
7932	Polygon	HU5121200	0,10926	0,000622	90061	HU5121200152	HU	0152	Somoskoújfalu
7709	Polygon	HU2070810	0,302643	0,00416	87826	HU2070810201	HU	0201	Ráckeresztúr
7877	Polygon	HU5050550	0,214269	0,00212	89509	HU5050550202	HU	0202	Martonyi
7865	Polygon	HU4171700	0,204298	0,002464	89385	HU4171700203	HU	0203	Kakasd
7778	Polygon	HU3181810	0,175627	0,001259	88517	HU3181810204	HU	0204	Narda
7854	Polygon	HU4141440	0,249396	0,003203	89282	HU4141440205	HU	0205	Sávoly
7748	Polygon	HU3080870	0,212763	0,002601	88214	HU3080870206	HU	0206	Gönyu
7753	Polygon	HU3080890	0,118642	0,000548	88268	HU3080890207	HU	0207	Dunaremete
7947	Polygon	HU6151500	0,139717	0,000604	90209	HU6151500208	HU	0208	Mánd
7800	Polygon	HU3201970	0,282117	0,002887	88734	HU3201970209	HU	0209	Gutorfölde
7792	Polygon	HU3201950	0,094485	0,000534	88658	HU3201950210	HU	0210	Maróc
7996	Polygon	HU7060730	0,170114	0,001789	90701	HU7060730212	HU	0212	Csanádalberti
7808	Polygon	HU3201980	0,205118	0,00147	88818	HU3201980213	HU	0213	Mihályfa
7979	Polygon	HU7030370	0,451159	0,003637	90525	HU7030370214	HU	0214	Fokto
7909	Polygon	HU5050680	0,159831	0,001146	89825	HU5050680215	HU	0215	Muhi
7937	Polygon	HU6090940	0,859411	0,017013	90109	HU6090940216	HU	0216	Komádi
7725	Polygon	HU2191850	0,264535	0,002828	87986	HU2191850218	HU	0218	Csopak
7741	Polygon	HU2191910	0,301996	0,004835	88147	HU2191910219	HU	0219	Nemesvámos
7710	Polygon	HU2070820	0,530563	0,006538	87834	HU2070820220	HU	0220	Isztimér
7724	Polygon	HU2191840	0,176519	0,001089	87976	HU2191840221	HU	0221	Balatonfuzfo
7823	Polygon	HU4020290	0,107939	0,000672	88965	HU4020290222	HU	0222	Martonfa
7744	Polygon	HU3080860	0,220319	0,001708	88181	HU3080860223	HU	0223	Vág
7778	Polygon	HU3181810	0,238708	0,002764	88518	HU3181810224	HU	0224	Vát
7812	Polygon	HU4020250	0,166264	0,001386	88854	HU4020250226	HU	0226	Vásárosdombó
7884	Polygon	HU5050570	0,183385	0,000848	89583	HU5050570227	HU	0227	Abaújvár
7909	Polygon	HU5050680	0,255641	0,001866	89826	HU5050680228	HU	0228	Hejokürt
7891	Polygon	HU5050600	0,379335	0,003907	89645	HU5050600229	HU	0229	Tiszabábolna

Figure 1 - Hungary, Shapefile table of attributes

I27 | 7350000

	A	B	C	D	E	F	G	H
1	NUTS 3 CODE	LAU CODE	LAU CODE truncated	validation	LAU NAME NATIONAL	LAU NAME LATIN	CHANGE (Y	POPULATI
2	HU313	01508	0150	-2	Ipolyszög	Ipolyszög	no	615
3	HU313	01526	0152	-49	Somoskőújfalu	Somoskőújfalu	no	2165
4	HU211	02015	0201	-1	Ráckeresztúr	Ráckeresztúr	no	3248
5	HU311	02024	0202	-1	Martonyi	Martonyi	no	409
6	HU233	02033	0203	-1	Kakasd	Kakasd	no	1643
7	HU222	02042	0204	-1	Narda	Narda	no	466
8	HU232	02051	0205	-1	Sávoly	Sávoly	no	521
9	HU221	02060	0206	-1	Gönyű	Gönyű	no	3036
10	HU221	02079	0207	-1	Dunaremete	Dunaremete	no	257
11	HU323	02088	0208	-1	Mánd	Mánd	no	279
12	HU223	02097	0209	-1	Gutorfölde	Gutorfölde	no	946
13	HU223	02103	0210	-2	Maróc	Maróc	no	71
14	HU333	02121	0212	-1	Csanádalberti	Csanádalberti	no	419
15	HU223	02130	0213	-1	Mihályfa	Mihályfa	no	350
16	HU331	02149	0214	-1	Foktő	Foktő	no	1568
17	HU311	02158	0215	-1	Muhi	Muhi	no	443
18	HU321	02167	0216	-2	Komádi	Komádi	no	5341
19	HU213	02185	0218	-1	Csopak	Csopak	no	1742
20	HU213	02194	0219	-1	Nemesvámos	Nemesvámos	no	2664
21	HU211	02200	0220	-1	Isztimér	Isztimér	no	957
22	HU213	02219	0221	-1	Balatonfűzfő	Balatonfűzfő	no	4084
23	HU231	02228	0222	-1	Martonfa	Martonfa	no	201
24	HU221	02237	0223	-1	Vág	Vág	no	460
25	HU222	02246	0224	-2	Vát	Vát	no	672
26	HU231	02264	0226	-1	Vásárosdombó	Vásárosdombó	no	1079
27	HU311	02273	0227	-1	Abaújvár	Abaújvár	no	256

Figure 2 - Hungary, 2017 population data

NUTS_3	LAU1_NAT_CODE	LAU2_NAT_CODE	LAU2_NAT_CODE modified	CHANGE	NAME_1	NAME_2_LAT	POP
HU313	4201	150	0150	no	Ipolyszög	Ipolyszög	639
HU313	4205	152	0152	no	Somoskőújfalu	Somoskőújfalu	2152
HU211	3710	201	0201	no	Ráckeresztúr	Ráckeresztúr	3311
HU311	3502	202	0202	no	Martonyi	Martonyi	374
HU233	4701	203	0203	no	Kakasd	Kakasd	1603
HU222	4808	204	0204	no	Narda	Narda	497
HU232	4406	205	0205	no	Sávoly	Sávoly	505
HU221	3802	206	0206	no	Gönyű	Gönyű	3135
HU221	3804	207	0207		Dunaremete	Dunaremete	241
HU323	4503	208	0208	no	Mánd	Mánd	274
HU223	5002	209	0209	no	Gutorfölde	Gutorfölde	1021
HU223	5003	210	0210	no	Maróc	Maróc	83
HU333	3604	212	0212	no	Csanádalberti	Csanádalberti	434
HU223	5006	213	0213	no	Mihályfa	Mihályfa	386
HU331	3303	214	0214	no	Foktő	Foktő	1606
HU311	3501	215	0215	no	Muhi	Muhi	533
HU321	3902	216	0216	no	Komádi	Komádi	5172
HU213	4903	218	0218	no	Csopak	Csopak	1648
HU213	4908	219	0219	no	Nemesvámos	Nemesvámos	2516
HU211	3705	220	0220	no	Isztimér	Isztimér	979
HU213	4902	221	0221	no	Balatonfűzfő	Balatonfűzfő	4299
HU231	3208	222	0222	no	Martonfa	Martonfa	205
HU221	3801	223	0223	no	Vág	Vág	481
HU222	4808	224	0224	no	Vát	Vát	662
HU231	3203	226	0226	no	Vásárosdombó	Vásárosdombó	1090
HU311	3512	227	0227	no	Abaújvár	Abaújvár	206
HU311	3511	228	0228	no	Hejőkürt	Hejőkürt	315
HU311	3514	229	0229	no	Tiszabábolna	Tiszabábolna	370
HU321	3907	230	0230	no	Kaba	Kaba	5932
HU211	3701	231	0231	no	Etyek	Etyek	4402
HU323	4501	232	0232	no	Baktalórántháza	Baktalórántháza	4302

Figure 3 - Hungary, population data 2012.

The case of the city of Bucharest

Another problematic case, generated by the join operation between spatial and population data, is that of the city of Bucharest. The join operation generates a significant overestimation of the population. The LAU CODE for the city of Bucharest (RO321 in figure 4b) is used for all 6 sectors of the city (Fig. 4 °). In this case, recognition and correction of this problem was only possible through a check between the total rows and those associated by join.

Table

RO_pop5

FID	Shape *	CNTR_CODE	NSI_CODE	NUTS_CODE	NAME	P2012	P2015	P2016	P2017	P2018
694	Polygon	RO	179132	RO321	Bucuresti Sector 1	192429	190866	210614	210291	211248
695	Polygon	RO	179132	RO321	Bucuresti Sector 2	192429	190866	210614	210291	211248
696	Polygon	RO	179132	RO321	Bucuresti Sector 3	192429	190866	210614	210291	211248
697	Polygon	RO	179132	RO321	Bucuresti Sector 4	192429	190866	210614	210291	211248
698	Polygon	RO	179132	RO321	Bucuresti Sector 5	192429	190866	210614	210291	211248
699	Polygon	RO	179132	RO321	Bucuresti Sector 6	192429	190866	210614	210291	211248
1706	Polygon	RO	95060	RO213	Iasi	317020	332271	362142	368818	373507
2788	Polygon	RO	155243	RO424	Timisoara	306466	303725	332983	331862	330014
919	Polygon	RO	54975	RO113	Cluj-Napoca	304802	304520	321687	322572	323484
1000	Polygon	RO	60419	RO223	Constanta	299049	296823	317832	316263	314816
1315	Polygon	RO	75098	RO224	Galati	287182	284986	304340	303111	303069

NUTS 3 CODE	LAU CODE	LAU NAME NATIONAL	LAU NAME LATIN	POPULATION
RO321	179132	Municipiul Bucureşti	Municipiul Bucuresti	2102912
RO213	95060	Municipiul Iaşi	Municipiul Iasi	368818
RO424	155243	Municipiul Timişoara	Municipiul Timisoara	331862
RO113	54975	Municipiul Cluj-Napoca	Municipiul Cluj-Napoca	322572
RO223	60419	Municipiul Constanţa	Municipiul Constanta	316263
RO411	69900	Municipiul Craiova	Municipiul Craiova	304089
RO224	75098	Municipiul Galaţi	Municipiul Galati	303111
RO122	40198	Municipiul Braşov	Municipiul Brasov	290348

Figure 4 - City of Bucharest - overestimation of the population

Topological problems

During the construction of the 5 shapefiles (2012, 2015, 2016, 2017, 2018) topological problems were presented. An example is the topology of Austria, it has been extracted from data.gv.at (2017), has not been used as such but has required a re-adaptation of the boundary line to make it fit with the LAU2 geometries of the neighboring countries (Eurostat).

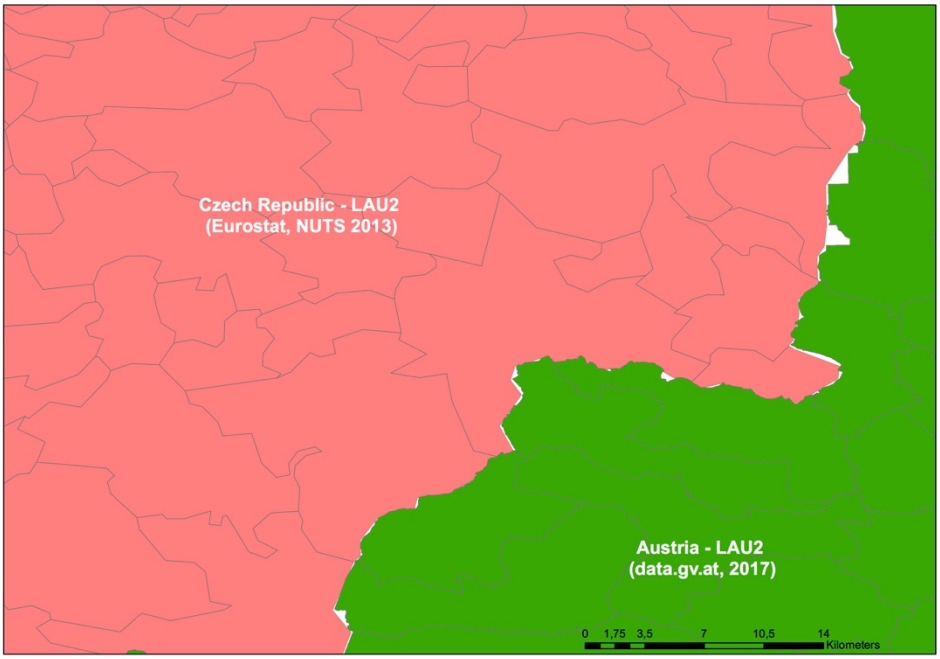

Figure 5 - Topological problems at the Austrian border

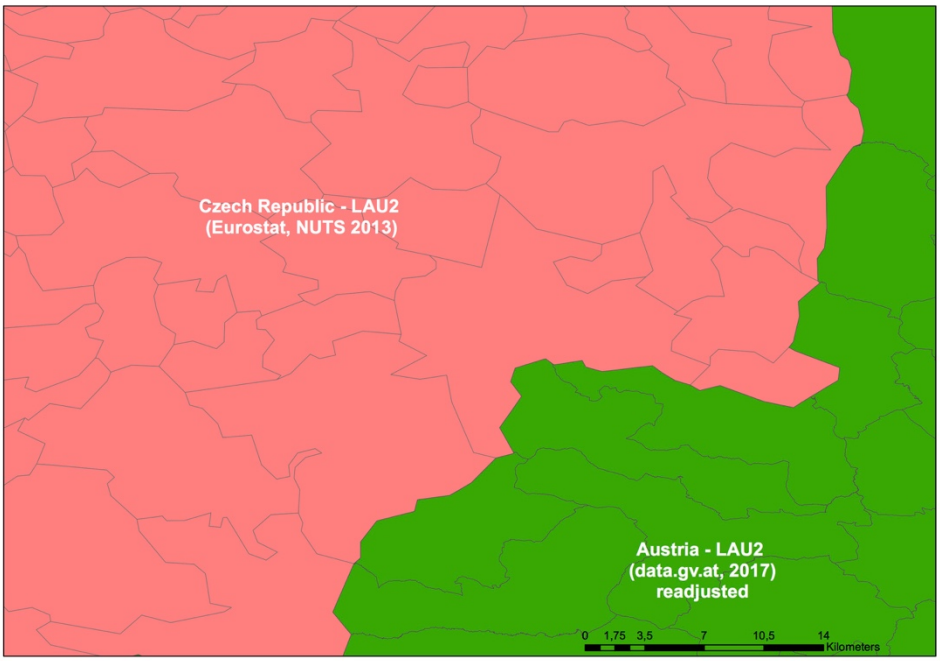

Figure 6 - Resolution of topological problems at the Austrian border

Another example of topological problems is the one that emerged for the simultaneous presence (in the same shapefile) of nations at LAU2 resolution (Eurostat 2013) with nations at NUTS2 resolution (Eurostat 2016). These topological problems, visi-

ble at the interface of the two spatial resolutions (figure 7) have been almost completely solved by replacing the NUTS2 2016 geometries with NUTS2 2013 geometries (figure 8).

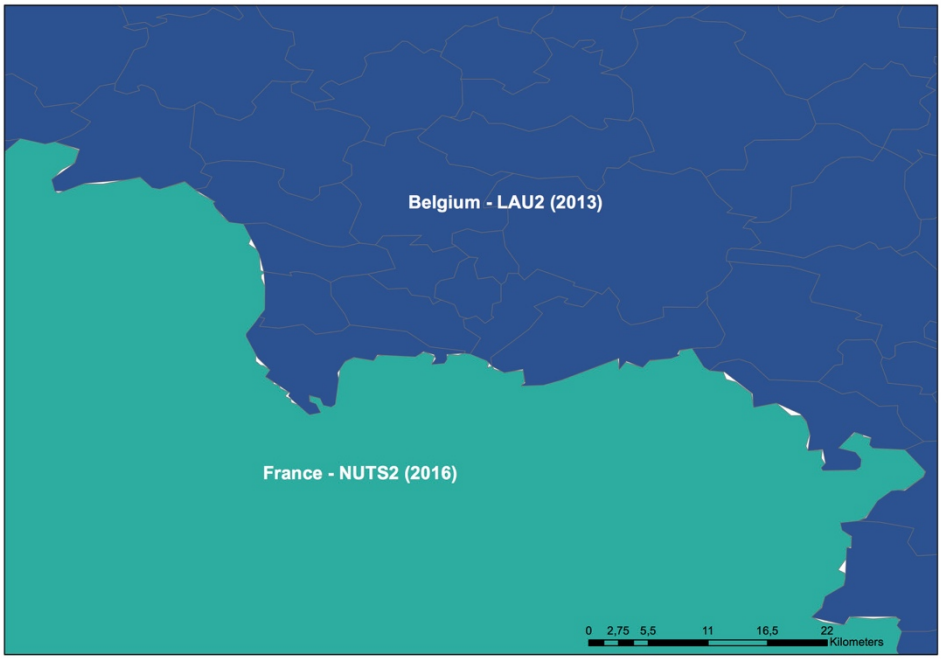

Figure 7 - Topological problems related to the simultaneous presence of nations at LAU2 resolution (Eurostat 2013) with nations at NUTS2 resolution (Eurostat 2016)

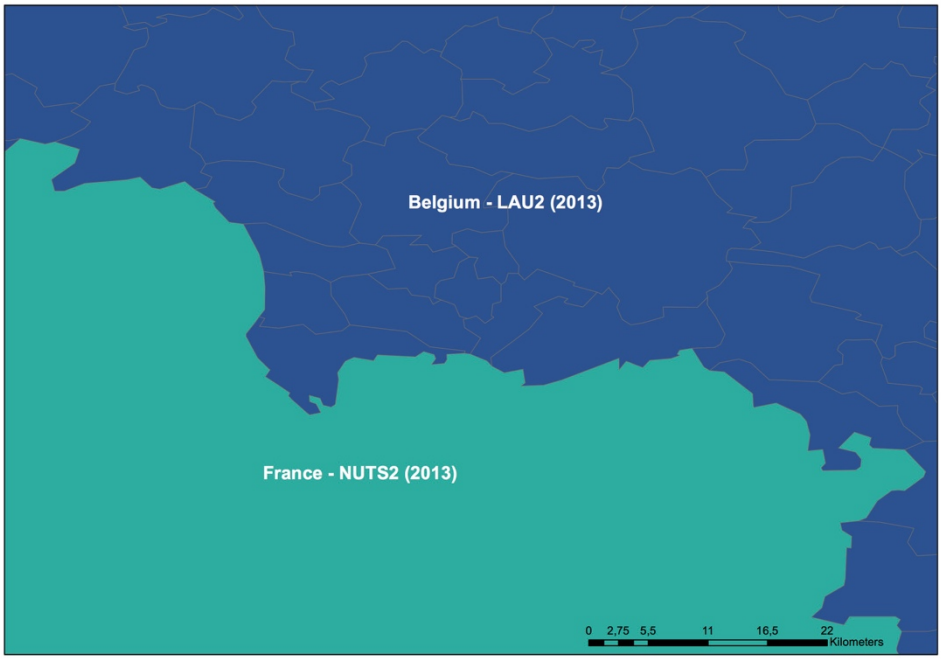

Figure 8 - Solving topological problems by replacing the NUTS 2 2016 geometries with NUTS 2 2013 geometries