Note on the regionalisation of PATSTAT April 2016

October 2016

Caro Vereyen

The regionalisation procedure is performed on all person addresses in the PATSTAT April 2016 edition, and adopts the new NUTS 2013 codes for 33 countries (EU28 plus Iceland, Liechtenstein, Norway, Switzerland and Turkey). A first step implies the parsing of postcodes and city names from the address field. Next, an automated approach links the retrieved postcodes and/or city names to NUTS codes via a 'NUTS-postcode-city name' concordance list. Further efforts to increase accuracy are pursued by visual inspection and by performing exhaustive quality controls. The procedure results in 98,49% of unique person-ID's (2.332.351 out of 2.368.075) being assigned to a NUTS3 region; and 99,04% assignment when considering the underlying patent amount. The procedure used in this round differs from previous regionalisation exercises (performed for Eurostat), due to issues with the NUTS 2013 source data. More specifically, inconsistencies/errors have been detected when comparing the Eurostat NUTS files 2013 with the NUTS 2010 version. The 2013 files have meanwhile been removed by Eurostat from their website, because of these inconsistencies. In the wake of these issues, and for our regionalisation of PATSTAT addresses, we adopted a two-folded approach in this round. (1) Countries that do not imply NUTS changes, are allocated by the previous NUTS 2010 version. For these countries, the regionalisation procedure allocates addresses, based on postcodes and city-names. (2) Countries that do imply NUTS changes (code changes, merges, boundary shifts and splits) to NUTS 2013 (DE, EL, FR, PL, PT, SI and GB[1]) are allocated twice: once based on the NUTS 2010 version (allocation based on postcodes and city-names) and a second time based on an intermediate, 'remedied' list provided by Eurostat, containing only postcodes linked to NUTS 2013 codes. This allowed us to compare the results obtained from both NUTS versions. All the affected 2010 NUTS codes have been traced and flagged and the new 2013 NUTS codes have been assessed and corrected or discarded (based on information provided by Eurostat). Regions without NUTS changes have been allocated, based on the NUTS 2010 information. A certain amount of addresses remains unassignable as a consequence of coinciding incidents like the poorer source data for NUTS 2013 (postcodes to NUTS) and incomplete address information. For instance in Greece, the NUTS 2010 region for Athens (EL300) is split into 7 new NUTS 2013 codes. Besides, for 401 person-ID's, postcodes are missing in either the address or the postcode-NUTS 2013 list. As a consequence 10,61% of the total amount on person-ID's (n=401) are left unassigned for now on the NUTS3 level. Similar issues affect the coverage rate for Portugal and Slovenia

---

[1] GB's regionalisation was performed with the latest available ONSPD data (May 2016), not with the Eurostat NUTS information. The ONSPD data is an exhaustive up-to-date database of postcodes linked to NUTS codes, and it is freely downloadable through internet. Contrary to the standard regionalisation procedure where postcodes and city names are used for allocating addresses to NUTS codes, the regionalisation of GB addresses is based on postcode information only.

The statistics below give an overview of the NUTS allocation coverage per country (EU28 plus Iceland, Liechtenstein, Norway, Switzerland and Turkey). Unassignable addresses and patents for specific countries are denoted in a separate column as ZZZ). The NUTS allocation for these addresses can be identified as the 'country code + ZZZ'.

Country	Total person ID's	Allocated person ID's	% Allocated person ID's	Total patents	Allocated patents	% Allocated patents	% Person ID's with ZZZ	% Patents with ZZZ
AT	54772	54008	98,61%	107451	106484	99,10%	0,41%	0,23%
BE	63749	63399	99,45%	118396	117943	99,62%	0,25%	0,15%
BG	1696	1657	97,70%	1861	1820	97,80%	1,53%	1,45%
CH	119911	117198	97,74%	296287	291674	98,44%	1,40%	1,12%
CY	571	571	100,00%	937	937	100,00%	0,00%	0,00%
CZ	6139	6030	98,22%	8364	8243	98,55%	0,88%	0,69%
DE	852305	842118	98,80%	2055485	2042007	99,34%	0,35%	0,16%
DK	46708	46233	98,98%	83037	82517	99,37%	0,24%	0,15%
EE	1035	1027	99,23%	1263	1255	99,37%	0,29%	0,24%
ES	63751	63076	98,94%	82284	81542	99,10%	0,17%	0,14%
FI	51340	50863	99,07%	108010	107442	99,47%	0,26%	0,15%
FR	391103	389423	99,57%	743839	741940	99,74%	0,19%	0,11%
GB >= 2000	198215	187195	94,44%	316746	302745	95,58%	0,32%	0,40%
GR	3781	3303	87,36%	4583	4039	88,13%	1,22%	1,07%
HR	1191	1173	98,49%	1778	1757	98,82%	1,18%	0,96%
HU	15132	15049	99,45%	19695	19603	99,53%	0,15%	0,12%
IE	14613	14209	97,24%	22477	22035	98,03%	1,81%	1,28%
IS	1224	993	81,13%	1938	1705	87,98%	17,89%	11,40%
IT	190745	188667	98,91%	315561	312918	99,16%	0,17%	0,13%
LI	1945	1945	100,00%	6901	6901	100,00%	0,00%	0,00%
LT	616	613	99,51%	799	796	99,62%	0,49%	0,38%
LU	4192	4169	99,45%	10953	10927	99,76%	0,55%	0,24%
LV	1091	1083	99,27%	1406	1396	99,29%	0,55%	0,57%
MT	280	262	93,57%	618	586	94,82%	6,43%	5,18%
NL	133593	132375	99,09%	301084	299453	99,46%	0,24%	0,14%
NO	21164	20928	98,88%	31657	31374	99,11%	0,22%	0,18%
PL	10189	9908	97,24%	13337	12994	97,43%	0,56%	0,47%
PT	4258	4004	94,03%	5131	4795	93,45%	1,03%	0,95%
RO	1549	1500	96,84%	1691	1639	96,92%	2,39%	2,37%
SE	99456	98180	98,72%	201370	199851	99,25%	0,30%	0,16%
SI	3382	2966	87,70%	6086	5173	85,00%	0,56%	0,31%
SK	1268	1212	95,58%	1656	1599	96,56%	3,55%	2,78%
TR	7111	7014	98,64%	11802	11685	99,01%	0,32%	0,20%
Sum	2368075	2332351	98,49%	4884483	4837775	99,04%	0,37%	0,24%