

Geospatial World Forum

**Workshop “Merging statistics and geospatial information
- Lessons learned towards INSPIRE”
Lisbon, 26/05/2015**

**Standardization and geocoding of place names in the
database of migratory flows**

Alessandro Cimbelli

Istat - Italian Institute of Statistics

Goal

- Identify, map and analyse migratory flows by normalization and geocoding of names of place of birth registered in the residence permits database
- Test and use software for data normalization (openRefine) and spatial open dataset (GeoNames, GADM)
- Understand how to enhance the quality of information of place of birth

What we have done

- Acquisition of two releases of the residence permits database from the Ministry of Interiors (2012, 2013)
- Processing and map production for 5 relevant countries
- Analysis of the multiyear results

<input type="button" value="▼"/> PROVUFF	<input type="button" value="▼"/> SESSO	<input type="button" value="▼"/> NAZIONE_NASCI	<input type="button" value="▼"/> LOUGO_NASCIT.
AQ	M	ALBANIA	VLORE
MI	F	ITALIA	MILANO
MI	F	BRASILE	PIRACICABA
MO	F	ALBANIA	BERAT
PG	M	TUNISIA	CHEBBA
SS	F	COLOMBIA	BOGOTA
MI	M	CONGO	KINSHASA
MI	M	CINA POPOLARE	FUJIAN
PC	M	CROAZIA	ZAGABRIA
RE	M	PAKISTAN	SIALKOT

Tasks

Normalization

- NIPORENI (1 rows)
- NICORENI (1 rows)
- NIHORENI (1 rows)



NIPORENI

Geocoding

luogo_nascita	json	lat	lng
BENI SUEF	{ "totalResultsCount": 1, "geonames": [{ "countryId": "357994", "adminCode1": "18", "countryName": "Egypt", "fclName": "city, village,...", "countryCode": "EG", "lng": "31.09785", "fcodeName": "seat of a first-order administrative division", "toponymName": "Banī Suwayf", "fcl": "P", "name": "Banī Suwayf", "fcode": "PPLA", "geonameId": 359173, "lat": "29.07441", "adminName1": "Muḥāfazat Banī Suwayf", "population": 189624 }] }	29.07441	31.09785

Map production



Dataset

- 5 countries
 - China
 - Egypt
 - India
 - Philippines
 - Ukraine
- 2 release of the residence permits database
 - 2012
 - 2013

Year 2012

Country	Total number of records	Normalized and geocoded records	Rate of normalized and geocoded records
China	21510	21010	97.68 %
Egypt	8195	7933	96.80 %
India	9298	8375	90.07 %
Philippines	7850	7450	94.90 %
Ukraine	7295	3446	47.24 %

Year 2013

Country	Total number of records	Normalized and geocoded records	Rate of normalized and geocoded records
China	19726	19192	97.29 %
Egypt	11431	11336	99.16 %
India	15418	14233	92.31 %
Philippines	6652	5831	87.65 %
Ukraine	13559	5804	42.80 %

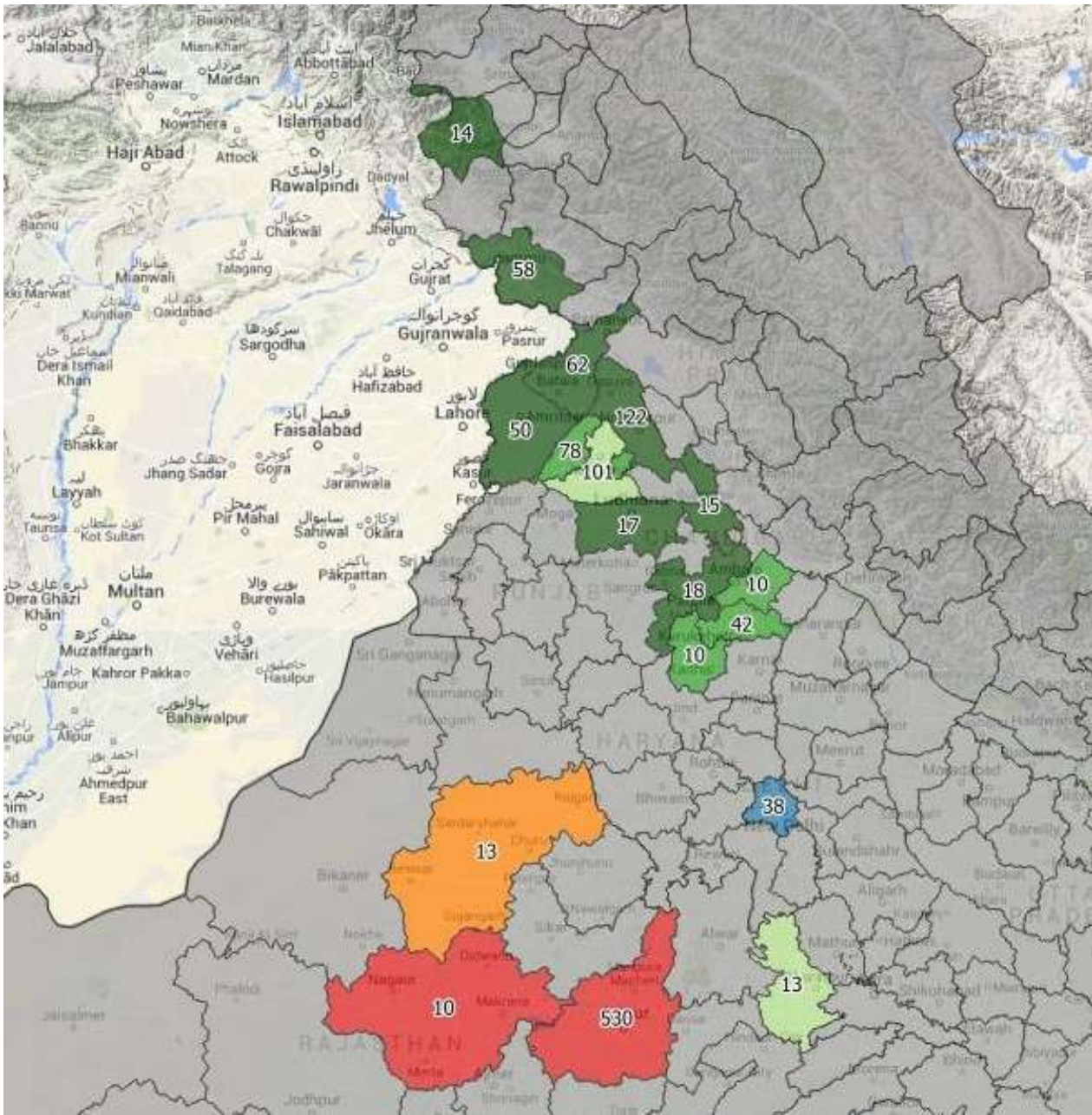
First results

Pro

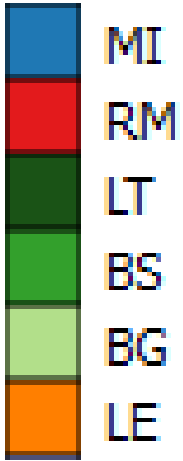
- Very good results for most of the countries (4/5)
- Results are confirmed by demographic analysis

Cons

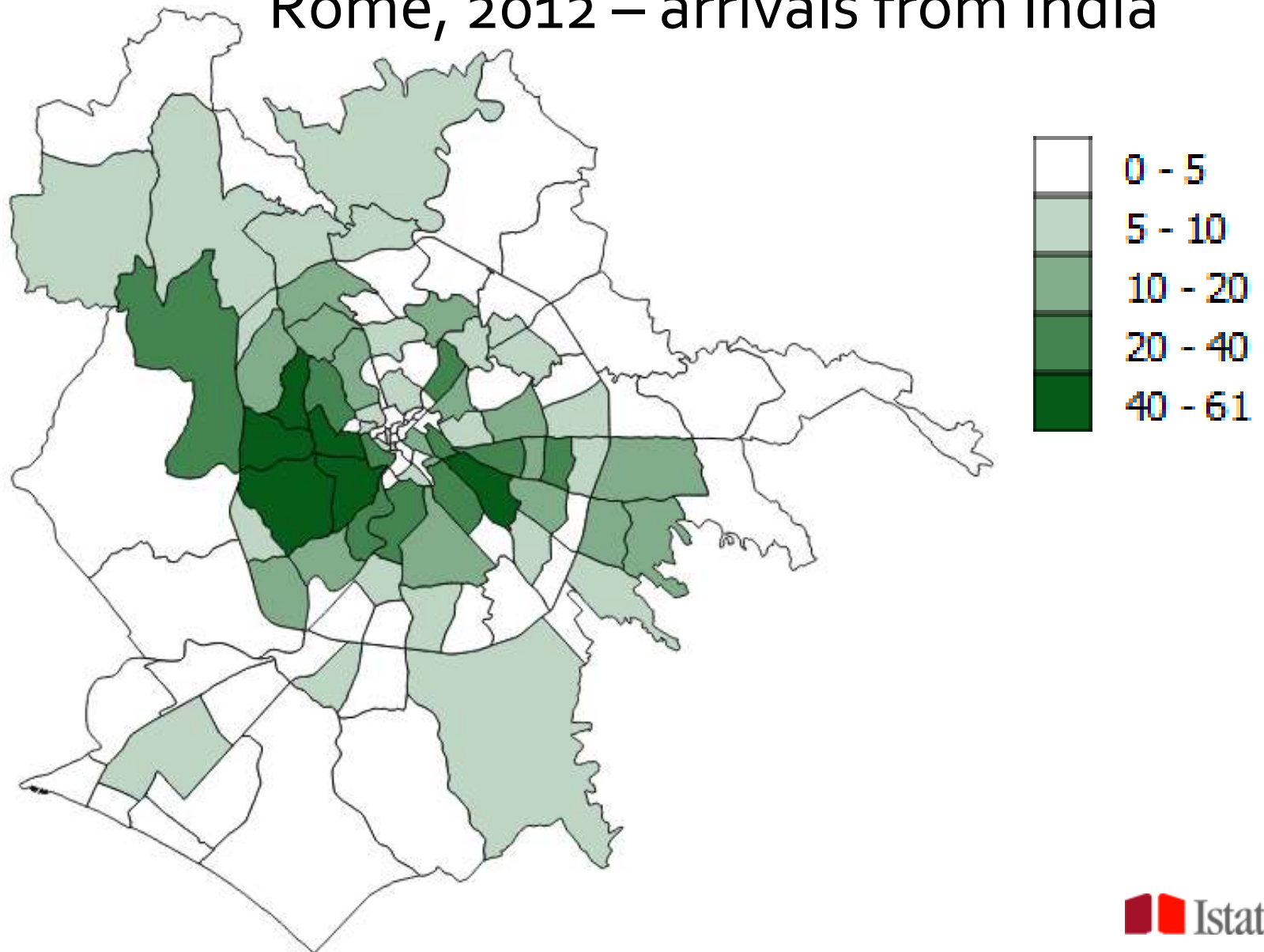
- Very bad results for Ukraine (~ 50% of “place of birth” referring to Ukraine)
- Sometimes the placename is referred to a adm level and in other case to a ppl



India
2012



Rome, 2012 – arrivals from India



What we want still to do

- Publish the results as OGC services
- Analyse the 2014 release (and 2015, if delivered in time)
- Study reconciliation process foreseen in OpenRefine

Thanks for your attention

cimbelli@istat.it