04 October 2018

United Nations Economic Commission for Europe

Conference of European Statisticians

Work Session on Migration Statistics
Geneva, Switzerland
24-26 October 2018
Item 4 of the provisional agenda
Big data and migration

Exploring international migration at subnational scale: the Italian case

Note by Istat*

Abstract

Residence permits data can be used as proxies of flow and stock data, with the key added value represented by the possibility to disentangle international migration by work, education, family and other reasons. Yet, several origin-destination dynamics are often lost through country-aggregated data.
This paper highlights the potential of residence permits data in Italy disaggregated at subnational level and geocoded in origin and destination countries. Several examples will be introduced demonstrating how the local dimension can provide data insights on links between communities of origin and receiving areas previously analysed only through scattered case studies.
The study identifies also disaggregated migration networks. For example, for the Chinese case, we can see that – as consequences of migratory chains - the Italian provinces attract in different way the inhabitants of the different Chinese provinces.
The same happens for other countries such as Moldova. Immigrants come from few areas. The most relevant is Chisinau, followed by Calarasi and Hincesti. They are directed towards the North-East of Italy and to Rome. The map shows for some origin territories clear preferences in terms of destination.
Using appropriate visualization tools, this unique framework could be of value to design better informed development and cooperation policies as well as integration and social cohesion measures.

* Prepared by Cinzia Conti and Alessandro Cimbelli

Global and local migration routes

I. Introduction

1. The production of statistics in the social field is undergoing radical changes thanks also to the increased availability and accessibility of administrative records and to the development of ICT (Information and Communication Technology); the new technologies for the management and treatment of individual data, in fact, have enhanced the expansion of administrative databases. In particular, the traditional archives showed significant improvements in the completeness, timeliness and partially in the quality of information collected.

2. Migrations statistics are a field in which the administrative archives could still be better exploited through the use of ICT and in which geographical information seem particularly relevant. Recently Istat, following up on this invitation, has investigated the possibility of using the so-called GI, Geographic Information, contained in the archive of residence permits in order to:

i. Study the origin of migration flows to Italy at sub-national administrative level;

ii. Analyze the placement on the territory of non-EU citizens on the Italian territory to an unbundled local level, including through techniques of record linkage between databases;

iii. Design of tools for a standardized collection of information;

3. About the fully exploitation of the administrative data, ISTAT is currently working to improve the quality of information on the place of birth at a disaggregated territorial level. In this way is it possible to analyse the origin-destination matrix not only at country level, but considering regions and provinces too. Consequently we can more precisely identify migration networks.

II. Data and methods

4. The approval of a European Parliament Regulation on European statistics on migration and international protection - 862/2007/EC - represented a milestone in improving the quantity and quality of information available, as well as a first step towards directing the attention of statistics to the various forms of integration. These improvements have involved above all the statistics on residence permits. The residence permits represent a very rich informative base and their use should not be limited to the purposes of Regulation 862/2007 that should be considered only a starting point for a full exploitation of the information provided by the dataset of residence permits.

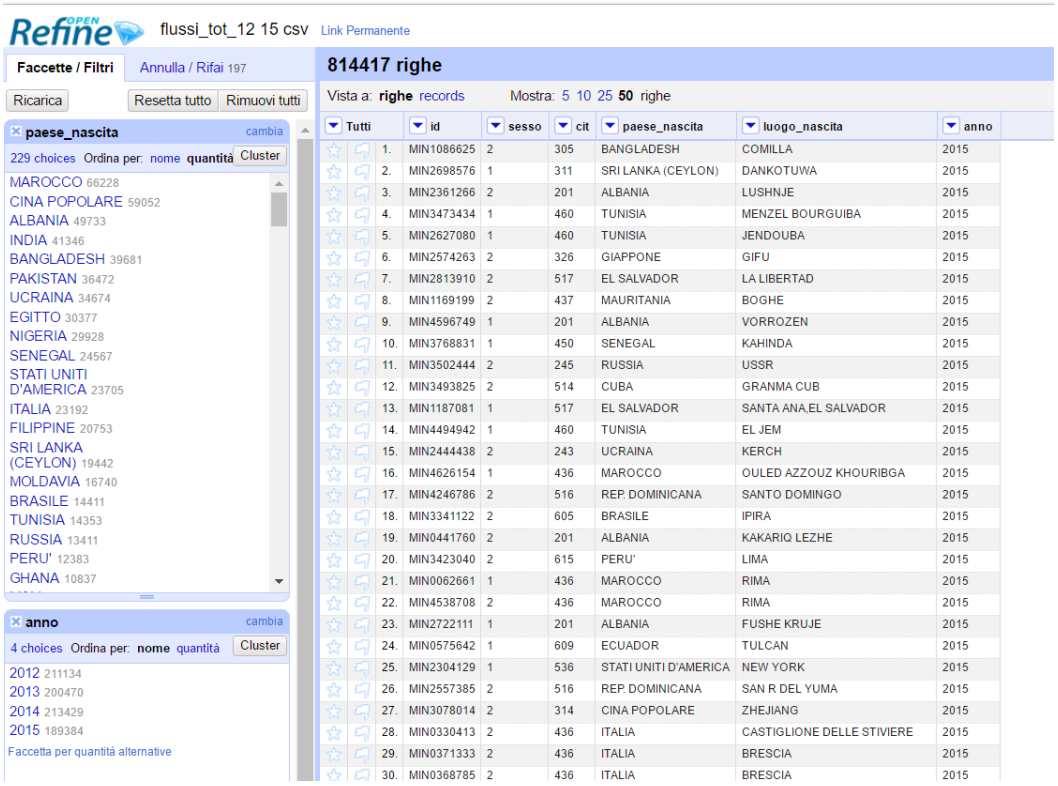
5. The residence permits database offers many information on non-EU citizens: age, gender, reason of the permit, citizenship, marital status, etc. In the recent years the integration of the database with other sources – through record linkage techniques – has allowed to add further information to the dataset exploited also in a longitudinal perspective.

6. The information on the place of birth on residence permits is recorded only in the form of alphanumeric information in a non-obligatory field on residence permits. This makes it difficult to process the variable, which is not standardised and sometimes missing. No assistance is currently provided when entering place names, leading to the possibility of errors and different ways of spelling names according to the language used or the use of different name conventions.

7. The first step in processing this information was to automatically correct the dataset using OpenRefine. Even after this initial processing phase, the dataset continued to contain errors, especially regarding the use of notations in different endonymous languages and names not recognised in the previous phase. We therefore proceeded to standardise the information by linking names and standard dictions contained in the database provided by the GeoNames website, which maintains 8 million place names from all over the world. The "INSPIRE" geographical portal was

also used as a search engine. The correct information was then geo-codified by adding information on longitude and latitude[1].

8. The normalization process has been revised because of the need of having the same place names correctly normalized and recognized for every year. At first it has been created a unique table of data, merging the four yearly releases. The merged table has avoided errors of normalization that could have been different for each year.



Tutti	id	sesso	cit	paese_nascita	luogo_nascita	anno
1.	MIN1086625	2	305	BANGLADESH	COMILLA	2015
2.	MIN2698576	1	311	SRI LANKA (CEYLON)	DANKOTUWA	2015
3.	MIN2361266	2	201	ALBANIA	LUSHNJE	2015
4.	MIN3473434	1	460	TUNISIA	MENZEL BOURGUIBA	2015
5.	MIN2627080	1	460	TUNISIA	JENDOUBA	2015
6.	MIN2574263	2	326	GIAPPONE	GIFU	2015
7.	MIN2813910	2	517	EL SALVADOR	LA LIBERTAD	2015
8.	MIN1169199	2	437	MAURITANIA	BOGHE	2015
9.	MIN4596749	1	201	ALBANIA	VORROZEN	2015
10.	MIN3768831	1	450	SENEGAL	KAHINDA	2015
11.	MIN3502444	2	245	RUSSIA	USSR	2015
12.	MIN3493825	2	514	CUBA	GRANMA CUB	2015
13.	MIN1187081	1	517	EL SALVADOR	SANTA ANA,EL SALVADOR	2015
14.	MIN4494942	1	460	TUNISIA	EL JEM	2015
15.	MIN2444438	2	243	UCRAINA	KERCH	2015
16.	MIN4626154	1	436	MAROCCO	OULED AZZOUZ KHOURIBGA	2015
17.	MIN4246786	2	516	REP. DOMINICANA	SANTO DOMINGO	2015
18.	MIN3341122	2	605	BRASILE	IPIRA	2015
19.	MIN0441760	2	201	ALBANIA	KAKARIQ LEZHE	2015
20.	MIN3423040	2	615	PERU'	LIMA	2015
21.	MIN0062661	1	436	MAROCCO	RIMA	2015
22.	MIN4538708	2	436	MAROCCO	RIMA	2015
23.	MIN2722111	1	201	ALBANIA	FUSHE KRUJE	2015
24.	MIN0575642	1	609	ECUADOR	TULCAN	2015
25.	MIN2304129	1	536	STATI UNITI D'AMERICA	NEW YORK	2015
26.	MIN2557385	2	516	REP. DOMINICANA	SAN R DEL YUMA	2015
27.	MIN3078014	2	314	CINA POPOLARE	ZHEJIANG	2015
28.	MIN0330413	2	436	ITALIA	CASTIGLIONE DELLE STIVIERE	2015
29.	MIN0371333	2	436	ITALIA	BRESCIA	2015
30.	MIN0368785	2	436	ITALIA	BRESCIA	2015

9. Not-interesting or sensitive information have been cut away from the table in order to reduce the size of the file and to make the processing with OpenRefine faster. The resulting structure of the table, as represented in the previous image, has the Migrant_ID, the gender, the country code and name, the placename and the year.

10. The Geocoding task has been performed directly inside OpenRefine,. The webservices of Geonames can be accessed simply with a limits of 2000 requests per hour and 30000 per day. An exception is thrown when these limits is exceeded. With OpenRefine is possible to set the time interval (in milliseconds) between the queries. A big interval causes long processes but shortening this value could cause the raise of the exception. In this case it is possible to rerun the geocoding only for the ignored records. The Geocoding task has not been changed and it is performed directly inside OpenRefine. Due to the limitation of the 2000 request/hour and the time needed for the processing of the four- year table of a single country, the geocoding is now run on a pivot table of the placenames and no more on the original table. This shortcut has dramatically reduced the size of the table without any impact on the quality of the results. The final results are summarized in the following table:

[1] The work is being performed as part of the 2013 Eurostat Grant: "Merging statistics and geospatial information in Member States" THEME: 08.1.43 – Geographical information system

Table 1 – Rates of normalized and geocoded records by country

Country	Total number of records (in 4 years)	Normalized and geocoded records	Rate of normalized and geocoded records
Morocco	66228	63388	**95,71**
China	59052	58564	**99,17**
Albania	49733	48917	**98,36**
India	41346	40735	**98,52**
Bangladesh	39681	37200	**93,75**
Pakistan	36472	36236	**99,35**
Ukraine	*34674*	*18474*	***53,28***
Egypt	30377	29424	**96,86**
Nigeria	29928	26205	**87,56**
Senegal	24567	22122	**90,05**
USA	23705	23116	**97,52**
Philippines	20753	19994	**96,34**
Sri Lanka	19442	19174	**98,62**
Moldova	*16740*	*9400*	***56,15***
Brazil	14411	14264	**98,98**
Tunisia	14353	13718	**95,58**
Russia	*13411*	*7966*	***59,40***
Peru	12383	12198	**98,51**
Ghana	10837	9736	**89,84**
Mali	10559	8399	**79,54**

11. The table shows very good results for most of the countries, where almost all the records have been normalized and geocoded. In twelve over twenty cases the rate of the corrected records has been of more than 95%. On the other hand, there are three cases of very low rates, below 60%. In these cases the use of Cyrillic charset in the passport, for the name of the birthplace, has often implied the input of the country name instead of the city of birth. Other causes of mismatching rely on the use of short names of the provinces, the format "city-province" use in the field of the birthplace (need to be just one name), the name of very small villages, not present in the Geonames database.

12. Once obtained the latitude and longitude from Geonames and exported the final tables in csv or tsv format, it is quite easy to import them in QGIS and represent the birthplaces as points on a map. A better picture of the places of the studied countries, where possibly start the flows to Italy, are given by intensity maps. The polygon administrative boundaries used for the maps are freely available from the GADM[2] website as shapefile, Google Earth kmz or ESRI geotadabase.

III. First results

13. This analysis of the places of births could be particularly useful to better understand the migration flows from some areas of the world, in particular migrations of asylum seekers[3]. The use of disaggregated information about the territories of origin could open new perspectives also in terms of cooperation between origin and destination countries

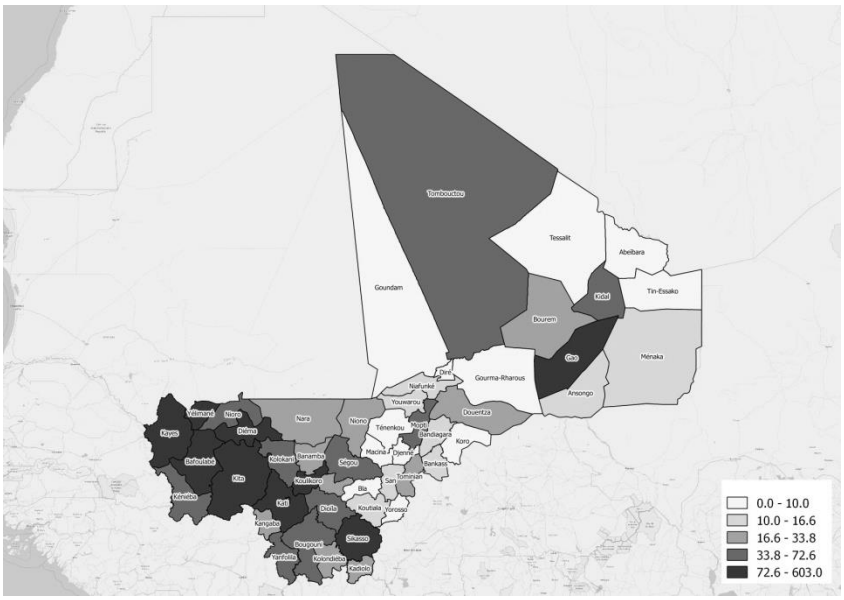
14. If we consider an area particularly relevant for Italy in terms of recent in-flows of immigrant as West Africa we can clearly identify the principal migration routs and even distinguish the effects of the conflicts and of terrorism. The data take into account the incoming-flows reported between 2012 and 2015 from Mali (fig. 1) and Nigeria (fig. 2), two of the most relevant countries for recent migrations towards Italy.

[2] http://www.gadm.org

[3] Of course this analysis has the limit of consider the place of birth and not the place of previous residence.

*Prepared by Alessandro Cimbelli and Cinzia Conti

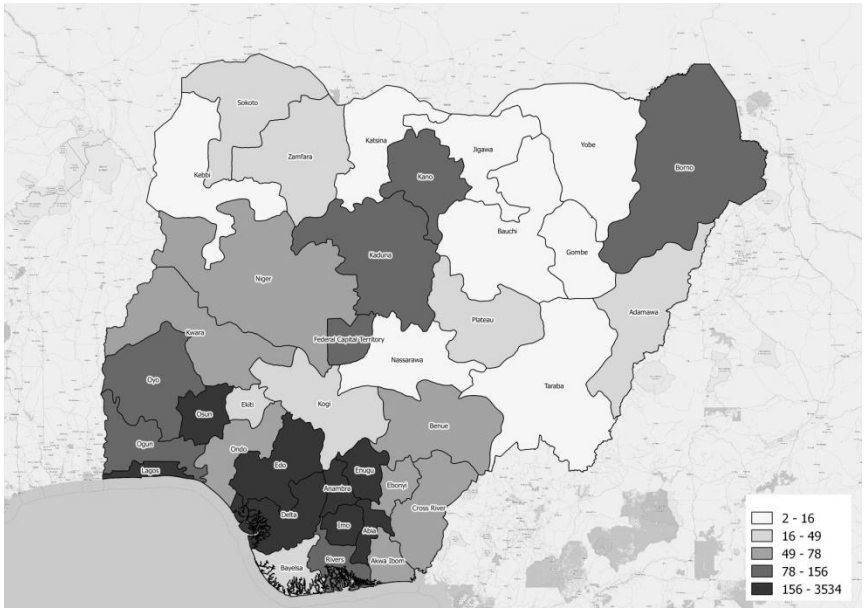
Figure 1 - Immigrants from Mali registered in Italy between 2012 and 2015 by region/province of birth



Source: Istat on data of Ministry of Interior

15. The analysis highlights migration flows coming from urban areas, but also from some territories affected by clashes and terroristic attacks. For Mali are strongly involved the cities of Bamako and Kayes, but also cities in the North of the country, as Gao, Kidal, and Timbuktu, that in 2012 had been occupied by Islamic fundamentalists (Figure 3). As regard to Nigeria, migrations to Italy affect principally the Edo State, the Delta State, Benin City, and Lago, urban territories near the coast. Nevertheless migrants towards Italy come also from the area of Borneo, considered as the headquarter of Boko Haram, and also the territories of Kano and Kaduna where the fundamentalist organization is very active (Fig.4).

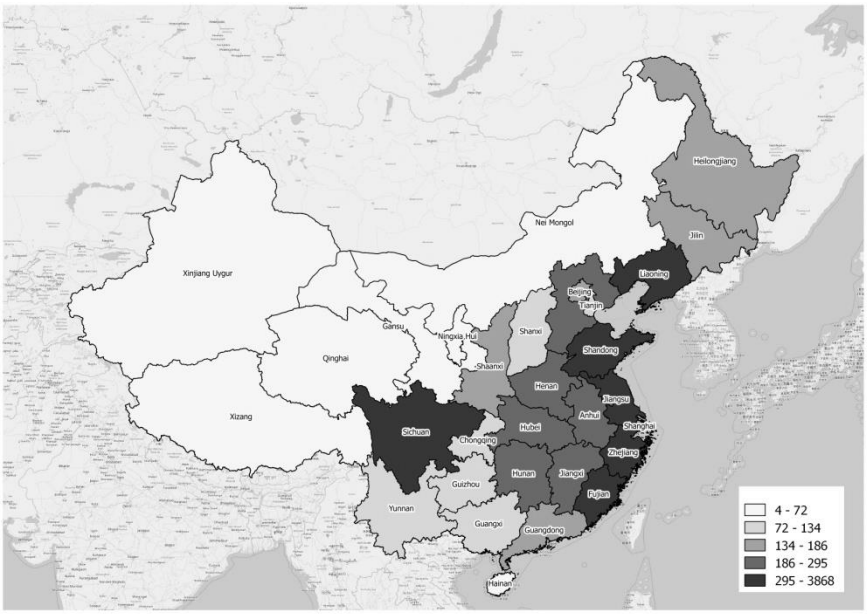
Figure 2 - Immigrants from Mali registered in Italy between 2012 and 2015 by region/province of birth



Source: Istat on data of Ministry of Interior

16. At the same time, this kind of data let us have better information also about labour migrations and family reunifications. The analysis of these information enlightens the existence of transnational networks, and they could be also useful for studying the remittances flows at a disaggregated territorial level. The Chinese community is one of the most important in Italy with almost 334,000 residence permits at the beginning of 2016. If we study the place of birth of Chinese arrived in Italy in 2015 we discover that more than 79% were born in the Zhejiang province (Fig.3).

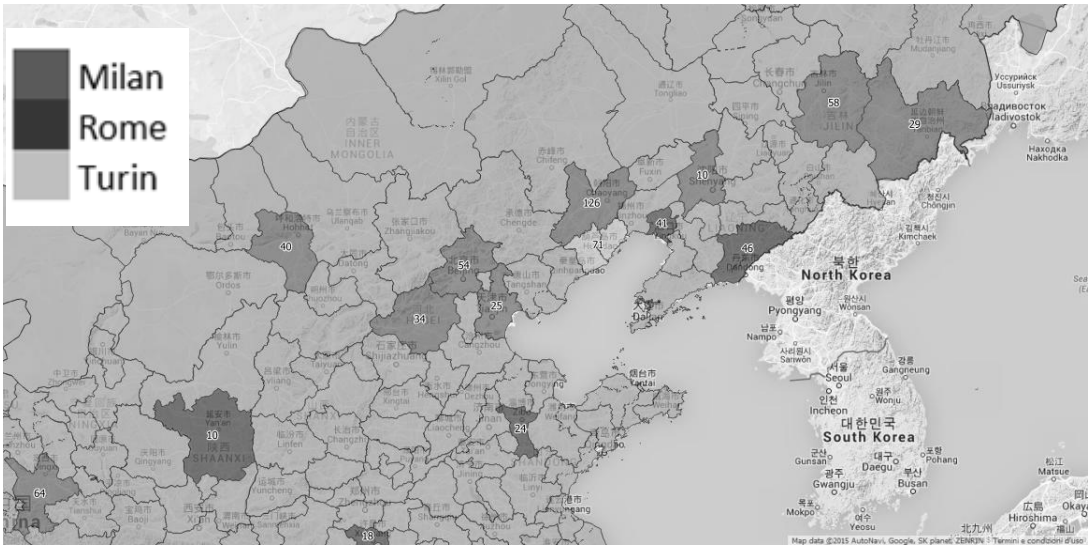
Figure 3 - Immigrants from China registered in Italy in 2015 by region of birth



Source: Istat on data of Ministry of Interior

17. At the same time, we can identify the disaggregated Chinese migration network. For example, we can see that – as consequences of migratory chains - the Italian provinces attracts in different way the inhabitants of the different Chinese territories (Fig.4).

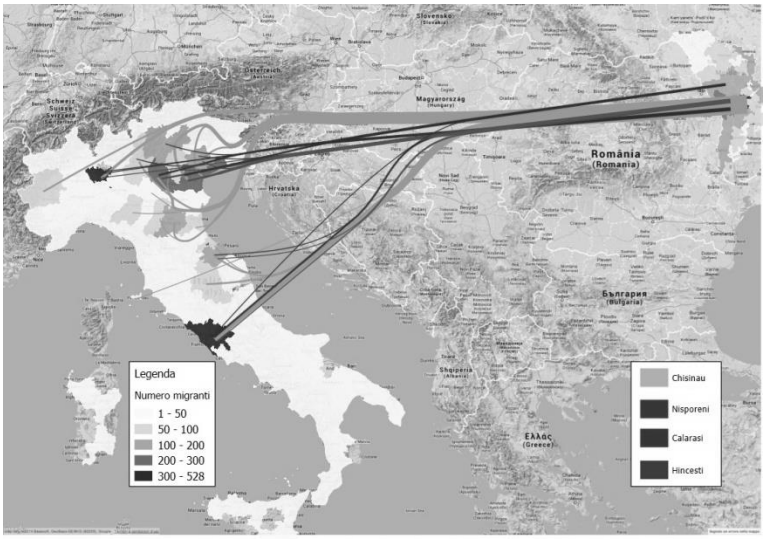
Figure 4 - Immigrants from China registered in Italy in 2012 by region of birth and Italian province of first registration of the residence permit



Source: Istat on data of Ministry of Interior

18. The same happens for other countries such as Moldova. Immigrants come from few areas. The most relevant is Chisinau, followed by Calarasi and Hincesti. They are directed towards the North-est of Italy and to Rome. The map shows for some origin territories clear preferences in terms of destination (Fig.5).

Figure 5 – Moldovans arrived in Italy in 2012 by province of first registration and place of birth.



Source: Istat on data of Ministry of Interior

IV. Further developments

19. The experience of the project on "Standardization and geocoding of place names in the database of migratory flows" has offered the opportunity to make a wide review of, and an in depth discussion on the use of different kind of data sources on migrants. In particular, the studies presented in these pages have given a specific contribute to the analysis of migration flows as processes that can be represented by routes tracked on geocoded maps. The study shows that it is possible to exploit more intensively administrative sources.

20. It is important to develop inter-institutional cooperation in order to obtain a better quality of administrative data. At the end of the study Istat suggested to the Ministry of Interior to develop a tool to allow information on place of birth to be collected correctly and shared without the need for ex post corrections.

21. There is much unexpressed potential in the use of such data sources in order to track migration routes. More efforts should be devoted to a further investigation of new strategies of integration between large archives, registers, and satellite images in order to enhance the quality of information, to be more precise on the numbers, places, time and the characteristics of the migration phenomenon. The satellite images could contribute in better understanding the steps of migration process in the continents for which is more difficult to have statistical data on internal migration.

22. A new analytical approach is nowadays possible thanks to innovative tools and non standard data sources. The combined use of satellite images, reported data, visas, residence permits, population lists available at local level, enables to analysing migrations, looking at the country of origin, the point of arrival, the internal movements of foreign population, and their final settlement. This results in a very useful information for policy makers also in order to develop cooperation programmes.

23. In 2018 Istat has started a cooperation with European Commission – Joint Research Centre – in order to develop new analysis and instruments for data visualization of migration routes and also with the purpose of sharing the Italian experience with other countries.

V. References

24. Cimbelli A., Conti C., Deriu, A. “The use of Big Data in studying migration routes” in: Cathleen M. Stützer, Martin Welker, Marc Egger (Edited by) Computational Social Science in the Age of Big Data. Concepts, Methodologies, Tools, and Applications", Herbert von Halem Verlag (Cologne, Germany), 2017 ISBN: 978-3-86962-267-5

25. Conti, C., Gabrielli D., Strozza, S. (2012). Dati amministrativi per le statistiche ufficiali sulle migrazioni” in: Rivista di Economia, Demografia e Statistica, Volume LXVI N. 1.

26. De Backer, O. (2014), Big Data and International Migration.

27. European Commission (2012). Report from the Commission to the European Parliament and the Council on the implementation of Regulation (EC) No 862/2007 on Community statistics on migration and international protection. Brussels, 20 September

28. European Big Data Value cPPP (2014), Strategic Research and Innovation Agenda.

29. Eurostat, “Big data – an opportunity or a threat to official statistics?”, Economic Commission for Europe Conference of European Statisticians Sixty-second plenary session Paris, 9-11 April 2014, ECE/CES/2014/32 http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2014/32-Eurostat-Big_Data.pdf.

30. Freeman, R. B. (2006). People Flows in Globalization. In The Journal of Economic Perspectives, Volume 20, Number 2, Spring 2006, pp.145-170(26), American Economic Association Publisher

31. Istat, (2016). Final Report on "Standardization and geocoding of place names in the database of migratory flows", Grant Agreement No. 08143.2013.004-2013.443, edited by Cimbelli, A.