



15 October 2018 г.

Русский

---

## Европейская экономическая комиссия ООН

### Конференция европейских статистиков

#### Рабочая сессия по статистике миграции

Женева, Швейцария

24-26 октября 2018 года

Пункт 4 предварительной повестки дня

Большие данные и миграция

## Исследование международной миграции на местном уровне: итальянский кейс

Доклад ИСтат\*

### *Резюме*

Данные по разрешениям на проживание могут быть использованы в качестве прокси-данных по потоку и фонду данных, причем основная их польза заключается в возможности рассортировать международную миграцию по разным причинам передвижения: работа, образование, семейные и другие причины. Хотя некоторые детали, зависящие от страны происхождения, часто теряются в процессе агрегирования данных в целом по стране. Данный доклад подчеркивает потенциальную пользу данных по разрешению на проживание в Италии, которые дезагрегированы на местном уровне и геокодированы по странам происхождения. В докладе будет представлено несколько примеров, которые демонстрируют, как данные на местном уровне могут разъяснить связь между местом происхождения и областью пребывания, что раньше было возможно проанализировать только с помощью выборочных предметных исследований.

Это обследование выявляет дезагрегированные миграционные сети. Например, в случае с китайским кейсом, мы видим, что как последствия миграционных цепочек, разные итальянские провинции привлекают жителей разных китайских районов.

То же самое происходит с приезжими из других стран, как, например, Молдова. Иммигранты приезжают из нескольких

\* Доклад подготовили Cinzia Conti и Alessandro Cimbelli

районов Молдовы – чаще всего из Кишинева, а также из Каларашского и Хынчештского районов. Люди оттуда в основном едут в северо-восточные районы Италии и в Рим. Карта показывает, что люди из определенных стран предпочитают определенные районы для переселения. Используя соответствующие инструменты визуализации, эта уникальная схема может быть весьма полезна для разработки инструментов и политик сотрудничества на основе более подробной информации, а также мер по интеграции и социальному слочению общества.

## Глобальные и местные пути миграции

### I. Введение

1. Производство статистики в социальной сфере претерпевает радикальные изменения, в том числе, благодаря увеличившейся доступности к административному учету, а также развитию информационных и коммуникационных технологий (ИСТ), новых технологий для управления и обработки индивидуальных данных, которые поспособствовали расширению административных баз данных. Конкретно традиционные архивы были значительно улучшены в смысле их полноты, временного охвата и частично качества собираемой информации.
2. Миграционная статистика это область, в которой административные архивы с конкретной географической информацией могут быть больше исследованы, особенно с помощью использования ИСТ. Недавно, в результате этого приглашения, Истат занялся исследованием возможности использования так называемых географических данных (GI), содержащихся в учете выданных разрешений на проживание, для того, чтобы:
  - i. Изучить происхождение миграционных потоков в сторону Италии на местном административном уровне;
  - ii. С помощью сопоставления разных баз данных проанализировать распределение граждан стран-не членов ЕС на итальянской территории по несгруппированным районам/регионам страны;
  - iii. Разработать инструменты для стандартизованного сбора информации;
3. Что касается исследования полного потенциала административных данных, ИСТАТ в данный момент работает над улучшением качества информации по месту рождения на дезагрегированном территориальном уровне. Таким образом становится возможным проанализировать матрицу происхождения-назначение не только на уровне страны, но также на региональном и районном уровнях. Вследствие этого мы можем более точно определить миграционные сети.

### II. Данные и методы

4. Утверждение Положения Европейского Парламента по Европейской статистике миграции и международной защите – 862/2007/ЕС представляет собой веху в улучшении количеств и качества доступной информации, а также это стало первым шагом к тому, чтобы обратить внимание статистики на разные формы интеграции. Эти улучшения и привели к использованию статистики разрешений на

проживание. Разрешения на проживание представляют собой богатый источник информации, чье использование не может быть ограничено только в целях Положения 862/2007, напротив – оно должно рассматриваться как отправная точка для полного исследования той информации, которая содержится в наборах данных по разрешениям на проживание.

5. База данных разрешений на проживание предлагает много разной информации по гражданам стран-не членов ЕС: возраст, пол, основание для выдачи разрешения, гражданство, семейное положение и т.п. За последние несколько лет интеграция базы данных с другими источниками, благодаря технике сопоставления записей, обеспечила возможность добавить дополнительную информацию в базу данных и исследовать ее с лонгитудинальной перспективы.

6. Информация о месте рождения в разрешение на проживание вносится только в алфавитно-цифровом виде в поле необязательной информации. Таким образом, могут возникнуть трудности при обработке переменных, если они не представлены в стандартизованном виде или если чего-то не хватает. В данный момент не оказывается никакой технической поддержки при внесении названий территорий, что приводит к возможности погрешности или различного написания названий в зависимости от того, какой язык используется, или использования разных соглашений по использованию названий.

7. Первым шагом в обработке этой информации была автоматическая коррекция набора данных с помощью приложения OpenRefine. Даже после первоначальной фазы обработки в наборе данных все еще содержались ошибки, особенно в случаях использования сокращений на других языках или названий, которые не были распознаны в предыдущей фазе. Поэтому мы продолжили процесс стандартизации информации путем сопоставления имен/названий и их стандартных произношений, содержащихся в базе данных вебсайта GeoNames, где имеется восемь миллионов названий локаций со всего света. Также в качестве поисковика использовался географический портал “INSPIRE”. Затем правильная информация геокодифицировалась путем добавления координат долготы и широты<sup>1</sup>.

8. Процесс нормализации затем еще раз перепроверялся, чтобы правильно скорректированные названия локаций применялись в данных каждого года. Для начала была создана уникальная таблица данных, где объединили четыре ежегодных наименования. В объединенной таблице не присутствуют ошибки нормализации, то есть разночтения, которые могли быть в каждом году.

---

<sup>1</sup> Данная работа проводилась в рамках Гранта Евростата в 2013 году “Совмещение статистической и географической информации в странах-членах ЕС» Тема: 08.1.43 – Географическая информационная система

Refine OPEN flussi\_tot\_12\_15 csv Link Permanente

Faccette / Filtri Annulla / Rifai 197

Ricarica Resetta tutto Rimuovi tutti

814417 righe

Vista a: righe records Mostra: 5 10 25 50 righe

	Tutti	id	Sesso	Cit	paese_nascita	luogo_nascita	anno
1.	MIN1088625	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	MAROCCHO	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	KAKARIO LEZHE	2015	
20.	MIN3423040	2	615	PERU'	LIMA	2015	
21.	MIN0062661	1	436	MAROCCHO	RIMA	2015	
22.	MIN4538708	2	436	MAROCCHO	RIMA	2015	
23.	MIN2722111	1	201	ALBANIA	FUSHE KRUIE	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	

paese\_nascita cambia

229 choices Ordina per: nome quantità Cluster

MAROCCHO 66228

CINA POPOLARE 59052

ALBANIA 49733

INDIA 41346

BANGLADESH 39681

PAKISTAN 36472

UCRAINA 34674

EGITTO 30377

NIGERIA 29928

SENEGAL 24587

STATI UNITI D'AMERICA 23705

ITALIA 23192

FILIPPINE 20753

SRI LANKA (CEYLON) 19442

MOLDAVIA 16740

BRASILE 14411

TUNISIA 14353

RUSSIA 13411

PERU' 12383

GHANA 10837

anno cambia

4 choices Ordina per: nome quantità Cluster

2012 211134

2013 200470

2014 213429

2015 189384

Faccetta per quantità alternative

9. Несодержательная или деликатная информация была исключена из таблицы с целью сокращения размеров файла и ускорения процесса обработки с помощью OpenRefine. Полученная структура таблицы, как показано на примере выше, содержит идентификационный номер мигранта, пол, код страны и название, название локации и год.

10. Геокодирование проводилось напрямую в приложении OpenRefine. Доступ к веб-ресурсам географических названий может быть предоставлен просто путем ограничения количества запросов до 2000 в час и 30000 – в день. При превышении этого лимита выдается сигнал исключения. В OpenRefine возможно установить интервал времени (в миллисекундах) между запросами. Большой интервал удлиняет процесс, но его сокращение может привести к выдаче сигнала исключения. В таком случае возможно будет перезапустить геокодирование только для пропущенных записей. Сама задача геокодирования не отменяется и выполняется прямо в рамках OpenRefine. По причине лимита 2000 запросов в час и времени, необходимого для обработки таблицы, содержащей четырехлетнюю информацию по одной стране, геокодирование теперь проводится в сводной таблице названий локаций, а не в первоначальной таблице. Такое упрощение значительно сокращает размер таблицы и при этом не приводит к ухудшению качества результатов. Окончательные результаты суммированы в следующей таблице:

Таблица 1 – Процент нормализованных и геокодифицированных записей по странам

Страна	Общее кол-во записей (за 4 года)	Нормализованные и геокодированные записи	Процент нормализованных и геокодированных записей
Морокко	66228	63388	95,71
Китай	59052	58564	99,17
Албания	49733	48917	98,36
Индия	41346	40735	98,52
Бангладеш	39681	37200	93,75
Пакистан	36472	36236	99,35
Украина	34674	18474	53,28
Египет	30377	29424	96,86
Нигерия	29928	26205	87,56
Сенегал	24567	22122	90,05
США	23705	23116	97,52
Филиппины	20753	19994	96,34
Шри-Ланка	19442	19174	98,62
Молдова	16740	9400	56,15
Бразилия	14411	14264	98,98
Тунис	14353	13718	95,58
Россия	13411	7966	59,40
Перу	12383	12198	98,51
Гана	10837	9736	89,84
Мали	10559	8399	79,54

11. В таблице видны очень хорошие результаты по большинству стран, где почти все записи были нормализованы и геокодифицированы. В двенадцати из двадцати случаев процент нормализованных записей превышает 95%. С другой стороны есть три случая очень низкого процента – ниже 60%. В этих случаях использование кириллического алфавита в паспорте в названии места рождения часто приводило к тому, что вместо названия места рождения вводилось название страны. В других случаях ошибки вызваны использованием сокращений в указании районов, использованием формата «город-район» в поле о месте рождения (должно быть только одно название), а также использования названий очень маленьких деревень, которых нет в базе данных GeoNames.

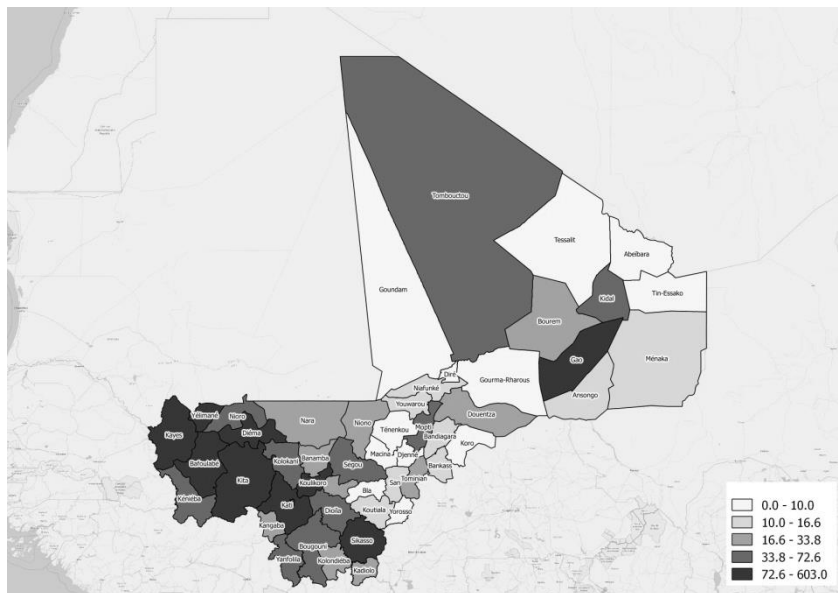
12. После нахождения широты и долготы с помощью приложения GeoNames и экспорта окончательных таблиц в формат csv или tsv, становится очень легко импортировать их в QGIS и обозначить места рождения на карте. Таким образом, мы получаем более подробную картину изучаемых стран, и регионов, откуда начинается миграция в Италию, которые обозначены на специальных цветных картах. Административно-хозяйственные границы в виде изломанных замкнутых линий используются для карт, которые можно загрузить с вебсайта GADM<sup>2</sup> в векторном формате, с Google Earth kmz или из геобазы данных ESRI.

### III. Первые результаты

13. Данный анализ мест рождения иммигрантов, конкретно мигрантов, просящих убежище<sup>3</sup>, может быть особенно полезен для более ясного понимания миграционных потоков из определенных регионов мира. Использование дезагрегированной информации о территориях происхождения этих людей может открыть новые перспективы сотрудничества между странами происхождения и назначения.

14. Если мы рассмотрим область Западной Африки, имеющую конкретное отношение к Италии, поскольку оттуда недавно прибыла волна иммигрантов, то сможем определить их основные миграционные пути и даже выделить последствия конфликтов или терроризма. Данные рассматривают прибывающие потоки в период с 2012 до 2015 гг из Мали (рис. 1) и Нигерии (рис. 2) – две наиболее часто встречающиеся страны происхождения последней волны иммигрантов в Италию.

**Figure 1 – Иммигранты из Мали, зарегистрированные в Италии в период с 2012 по 2015 гг по регионам/районам рождения**



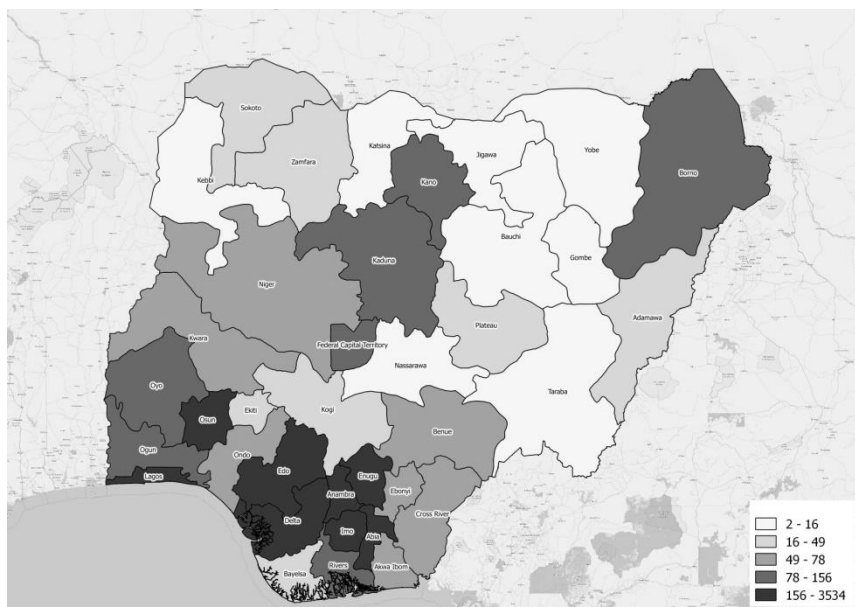
<sup>2</sup> <http://www.gadm.org>

<sup>3</sup> Конечно же у данного анализа есть свои ограничения, поскольку место рождения и место последнего проживания не всегда одно и то же.

Источник: ИСТАТ, данные Министерства внутренних дел

15. В анализе видно, что основные миграционные потоки идут из городских районов, а также еще с некоторых территорий, подвергшихся вооруженным столкновениям и террористическим нападениям. В Мали это города Бамако и Каес, а также северные города – Гао, Кидаль и Томбукту, которые в 2012 году были оккупированы исламскими фундаменталистами (рис. 3). Что касается Нигерии, миграция в Италию идет в основном из штатов Эдо, Дельта, Бенин-Сити и Лагос – городских районов недалеко от побережья. Кроме этого мигранты в Италию едут из областей Борно, считающейся штаб-квартирой Боко Харам, а также с территорий Кано и Кадуна, где весьма активны фундаменталистские организации (рис. 4).

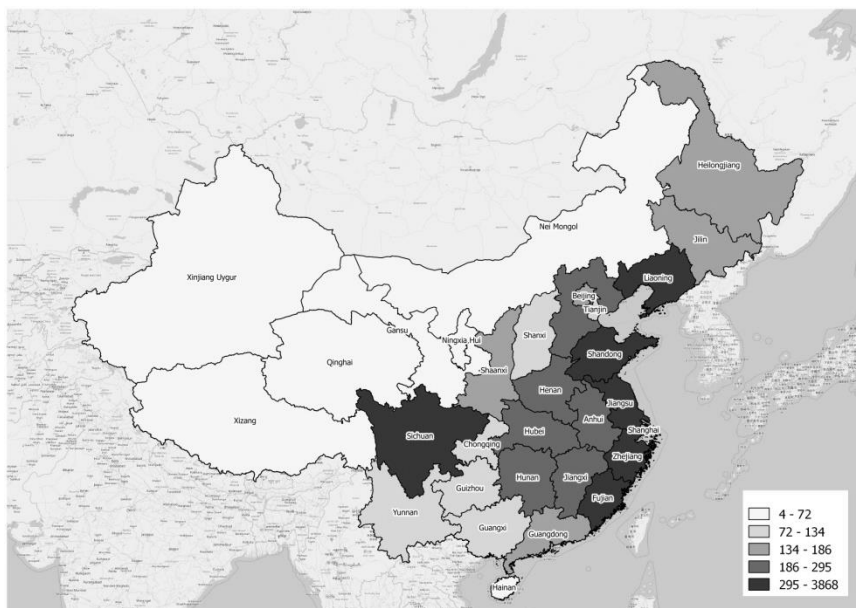
**Рисунок 2. Иммигранты из Нигерии, зарегистрированные в Италии в период с 2012 по 2015 гг по регионам/районам их рождения**



Источник: ИСТАТ, данные Министерства внутренних дел

16. В то же самое время эти данные помогают нам получить более ясную информацию по трудовой миграции и миграции с целью воссоединения с семьей. Анализ этой информации объясняет существование транснациональных сетей, а также она может быть полезной для изучения потоков денежных переводов на деагрегированном территориальном уровне. Китайская община одна из наиболее значительных в Италии, где на 2016 год было выдано почти 334 000 разрешений на проживание. Если мы изучим места рождения китайских иммигрантов, прибывших в Италию в 2015 году, мы обнаружим, что более 79% из них родились в провинции Чжецзян (рис. 3).

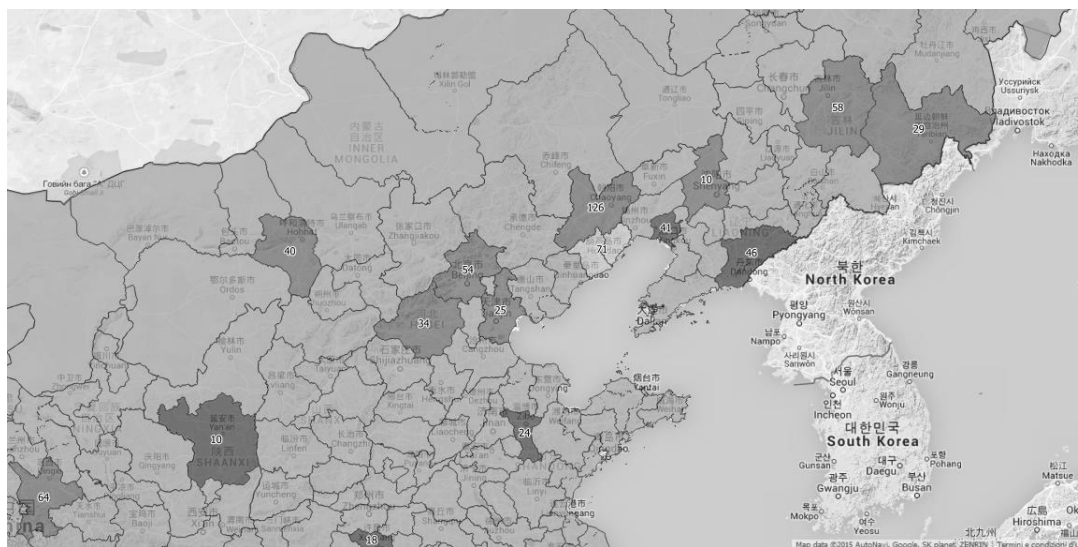
**Рисунок 3. Иммигранты из Китая, зарегистрированные в Италии в 2015 году по регионам рождения**



Источник: ИСТАТ, данные Министерства внутренних дел

17. В то же самое время, мы можем определить дезагрегированную сеть китайской миграции. Например, мы видим, что в результате миграционных цепочек, разные итальянские провинции привлекают жителей разных китайских территорий (рис. 4).

**Рисунок 4. Иммигранты из Китая, зарегистрированные в Италии в 2012 году по регионам рождения, и итальянские провинции, где они впервые были зарегистрированы и получили разрешение на проживание**

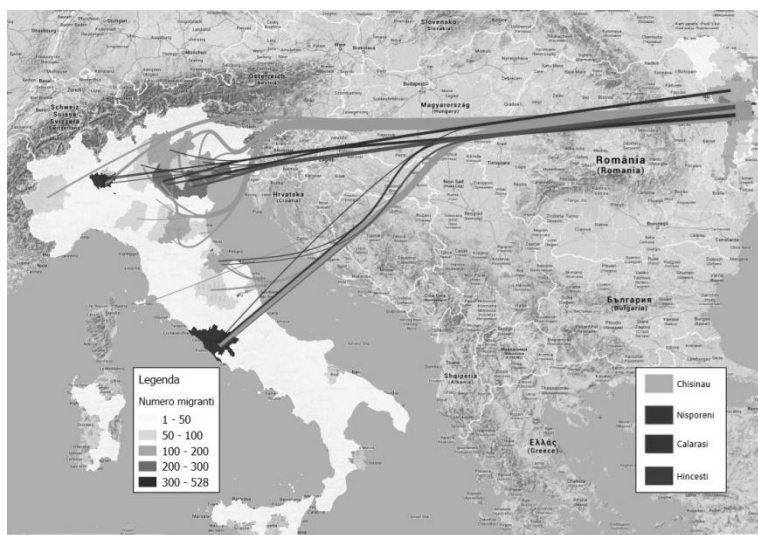


Источник: ИСТАТ, данные Министерства внутренних дел



18. То же самое происходит и с другими странами, например с Молдовой. Иммигранты прибывают из нескольких молдавских регионов – прежде всего из Кишинева, а также из Каларашского и Хынчештского районов. Они в основном направляются в северо-восточные районы Италии и в Рим. На карте ясно видно, выходцы из каких территорий, какие районы Италии предпочитают (Рис. 5).

**Рисунок 5. Молдавские иммигранты, прибывшие в Италию, по районам их первой регистрации и местам рождения.**



Источник: ИСТАТ, данные Министерства внутренних дел

#### IV. Дальнейшее развитие

19. Опыт данного проекта «Стандартизация и гео-кодификация названий локаций в базе данных миграционных потоков» предоставил нам возможность сделать широкий обзор, а также провести углубленное обсуждение использования разнообразных источников данных по мигрантам. Если быть конкретными, исследования, изложенные на этих страницах, сделали особый вклад в анализ миграционных потоков в качестве процессов, которые могут быть представлены в виде маршрутов, обозначенных на географических картах. Данное исследование также показывает, что административные источники можно использовать намного интенсивнее.

20. Важно развивать межведомственное сотрудничество для получения более качественных административных данных. В конце исследования ИСТАТ предлагает Министерству внутренних дел разработать инструмент для правильного сбора и предоставления информации по местам рождения, которую не нужно будет потом исправлять.

21. В использовании этих данных кроется много неожиданного потенциала, который может помочь отследить миграционные маршруты. Больше усилий необходимо посвятить дальнейшему исследованию новых стратегий интеграции крупных архивов, регистров и спутниковых снимков с целью улучшения качества информации, получения более точных цифр, локаций, времени и характеристик

миграционного феномена. Спутниковые снимки могут дать более ясную картину стадий миграционного процесса на континентах, где гораздо сложнее получить статистические данные по внутренней миграции.

22. Новый аналитический подход в сегодняшнее время возможен благодаря инновационным инструментам и нестандартным источникам данных. Комбинированное использование спутниковых снимков, зарегистрированных данных, информации о визах, разрешениях на проживание, списках населения, доступных на местном уровне, дает возможность проанализировать миграцию по стране происхождения мигрантов, по местам их прибытия, по внутреннему передвижению иностранного населения и их конечному месту проживания. Это дает очень полезную информацию политическим деятелям, что помогает им разрабатывать программы сотрудничества.

23. В 2018 году ИСТАТ начал сотрудничество с Европейской Комиссией – Центр совместных исследований – с целью разработки новых анализов и инструментов для визуализации данных по маршрутам миграции, а также для того, чтобы поделиться опытом Италии с другими странами.

## V. Справочная литература

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](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.