

**Европейская экономическая комиссия****Конференция европейских статистиков****Группа экспертов по переписям населения  
и жилищного фонда****Восемнадцатое совещание**

Женева, 28–30 сентября 2016 года

Пункт 5 предварительной повестки дня

**Методы оценки качества и пригодности  
для использования регистров  
и административных источников****Интеграция административных источников данных  
в Италии в целях повышения объема доступных  
данных переписи населения****Записка Национального института статистики Италии<sup>1</sup>***Резюме*

Следующий раунд переписи населения Италии ознаменует собой окончательный переход от традиционной переписи методом поквартирного обхода к регистровой системе, которая обеспечивает комбинирование официальных регистров населения с другими тематическими административными источниками. Использование интегрированной системы регистров позволяет выявлять полезные паттерны в громадных массивах административных данных и создавать профили подгрупп населения. Итальянский национальный институт статистики (ИСТАТ) провел первоначальное тестирование с целью определения обычно проживающего населения на основе использования административных данных, которое принесло полезные результаты. ИСТАТ проанализировал качество регистров и выявил паттерны в данных. Эти паттерны позволяют ИСТАТ классифицировать лиц по конкретным группам, которые также представляют собой критически важное подмножество для определения новой стратегии проведения переписи. ИСТАТ затем определил предварительный порядок расчета численности обычно проживающего населения.

<sup>1</sup> Авторы: Джерардо Галло, Анджела Кьеппа, Валерия Томео и Стефано Фалорс.



## I. Введение

1. Следующий раунд переписи населения в Италии ознаменует собой окончательный переход от традиционной переписи методом поквартирного обхода к регистровой системе. Несмотря на серьезные усилия, предпринятые в ходе переписи населения 2011 года в целях расширения охвата за счет использования различных способов сбора данных и административных данных в качестве базовых совокупностей переписного населения, с переписью некоторых групп населения по-прежнему возникали очень большие трудности при использовании традиционной переписи. Исходя из итогов послепереписного обследования 2011 года, в котором участвовало более 320 000 человек, Итальянский национальный институт статистики (ИСТАТ) сделал вывод о том, что, согласно оценке, примерно 650 000 постоянных жителей Италии не были учтены в ходе переписи и что около 80% из них являлись иностранными гражданами (ИСТАТ, 2014). Эти результаты наряду с огромными затратами на проведение традиционной переписи и значительной статистической нагрузкой на респондентов заставили ИСТАТ принять новую стратегию проведения переписи, опирающуюся на комбинированное использование множественных источников данных и обследований (Crescenzi et al., 2015).

2. В работах Poulain et al. (2013) и Wallgren et al. (2011) отмечалось, что использование регистра населения для расчета численности обычно проживающего населения не позволяет автоматически решить проблемы охвата: точность этого регистра имеет критически важное значение даже в тех странах, которые используют регистровую перепись на протяжении десятилетий. Административные данные о населении страдают проблемами неполного охвата, главным образом из-за трудности регистрации международной миграции. В Италии граждане страны, проживающие за границей, и иностранцы, которые покинули страну на постоянной или долговременной основе, должны исключаться из регистра населения. Однако эмигранты не считают необходимым уведомлять власти о своем отъезде. Кроме того, местные органы власти заинтересованы в сохранении стабильной численности своего населения, считая этих людей «временно» выехавшими, с тем чтобы не исключать их из регистра. Второй критической группой являются иностранцы, обычно проживающие в Италии, но без официального вида на жительство. Хотя охват обычно проживающих иностранцев отслеживается в Италии, некоторые иностранные граждане не желают регистрироваться в регистре населения, особенно граждане – члены Европейского союза. Кроме того, лица, которые утратили основания для пребывания в Италии, и иммигранты, въехавшие без разрешения, не могут быть зарегистрированы в регистре населения<sup>2</sup>. Последние, не охватываемые никакими административными источниками, являются с трудом поддающейся учету подгруппой населения и требуют проведения соответствующего обследования для обеспечения их учета<sup>3</sup>.

<sup>2</sup> С 8 августа 2009 года новый закон (Закон № 94 от 15 июля 2009 году) квалифицирует незаконный въезд или пребывание в Италии в качестве преступления. Таким образом, иностранные граждане, уличенные в незаконном въезде в Италию или пребывании без разрешения на ее территории, считаются виновными в совершении преступления незаконной иммиграции.

<sup>3</sup> К числу с трудом поддающихся учету групп населения относятся бездомные и лица, которые не имеют место обычного жительства в одном и том же месте на протяжении года.

3. Предварительным решением проблемы неполного охвата регистра населения может являться использование интегрированной системы регистров: официальные регистры населения будут увязываться с другими тематическими административными источниками (касающимися трудовой деятельности, учебы, налогообложения, доходов и т.д.). Эта система может использоваться для выявления групп, которые соответствуют национальному или международному определению «обычно проживающего населения». Проблема неполноты охвата регистров населения может быть устранена путем использования сигналов присутствия лиц на территории Италии, поступающих из других регистров; в то время как отсутствие данных о лицах в регистре населения может служить свидетельством избыточного охвата. Расчет переписных данных о численности обычно проживающего населения с использованием системы регистрации является непростой задачей, которая предусматривает разработку нового производственного процесса на основе этой новой базы данных.

4. Исходя из перспективы непрерывной переписи в Италии (*censimento permanente*), которая должна быть в значительной степени ориентирована на оптимизацию административных архивов, ИСТАТ создал рабочую группу с целью тестирования возможностей использования административной информации для целей переписи и для расчета данных о численности населения. Информация о результатах этого тестирования, проведенного исследователями ИСТАТ, приводится в нижеследующих пунктах. В целях эксплуатации потенциала огромного объема информации, содержащегося в административных источниках, ИСТАТ использовал процесс извлечения знаний из баз данных (KDD). Он представляет собой структурированный итеративный процесс, в рамках которого некоторые переменные, подлежащие анализу, создаются в ходе построения. Этот процесс позволяет выявлять полезные паттерны и конкретные подмножества в большом объеме административных данных. Использование и анализ хранилища ИСТАТ и административных регистров проводились в качестве пробного шага в целях определения новой стратегии проведения переписи.

5. Первая цель тестирования заключалась в том, чтобы определить предварительный подход к обработке имеющихся административных данных в целях расчета численности обычно проживающего населения. Это предусматривало отбор и оценку качества административных данных. Вторая цель состояла в выявлении профилей и связей в административных данных, а затем в отборе наиболее релевантных характеристик и производной информации из имеющихся в регистрах атрибутов или создании новых элементов в качестве функций исходных. Кроме того, выявление паттернов позволит ИСТАТ классифицировать лиц по конкретным группам, которые представляют собой «критические подмножества» с точки зрения недостаточного и чрезмерного охвата официальных регистров населения и которые должны учитываться при разработке новой стратегии проведения переписи.

## **II. Пробный расчет численности обычно проживающего населения с использованием административных регистров, имеющихся в распоряжении ИСТАТ**

6. Для того чтобы справляться с растущим числом наборов административных данных, приобретаемых для использования в статистических целях, и максимизировать выгоды, вытекающие из огромного объема доступной информации, необходимо создать интегрированную систему административных источников. ИСТАТ движется в этом направлении путем централизации некоторых

функций приобретения, хранения, интеграции и оценки качества административных данных в рамках системы ИСМ (Интегрированная система микроданных). Эта система управляет социальными и экономическими административными данными в отношении: характеристик отдельных лиц и домохозяйств; демографических аспектов, статуса в занятости, уровня образования; мест жительства, работы или учебы; типологии единицы, в которой люди осуществляют свою деятельность или проводят свое время: дома, школы, места работы; типологии взаимосвязей между отдельными лицами, единицами и местами. Этапом интеграции в системе СИУ является процесс увязки и физической интеграции микроданных, поступающих из различных источников: применяемые стратегия интеграции и набор алгоритмов определяются имеющимися связующими переменными. Интеграция означает: i) выявление каждого объекта (лица, хозяйственной единицы) в административных источниках данных с уникальным и стабильным (во времени) идентификационным номером (ИН); ii) определение по каждому объекту логических и физических взаимосвязей между наборами данных, поступающих из различных источников. СИУ была создана с целью поддержки процессов статистического производства: использование ИН СИУ позволяет создавать структуры данных, способные служить отправной точкой для статистических процессов, опирающихся на административные данные. Можно создать тематическую базу данных с использованием синтаксиса для определения правил извлечения на основе использования отчетного периода или статистических доменов или поднаборов переменных или объектов и т.д. Для оценки того, как административные данные могли бы содействовать совершенствованию показателей численности населения, была создана тематическая база данных об административных сигналах в поддержку первоначального эксперимента. Данные, использовавшиеся в ходе тестирования, брались из конкретных административных источников, уже хранящихся и интегрированных в системе СИУ: муниципальное постоянное население, виды на жительство, работники наемного труда и самозанятые, обязательное образование, студенты, пенсионеры, непенсионные пособия, доходы и налоги.

7. Базовый регистр населения по-прежнему находится в ИСТАТ в стадии разработки: был создан специальный тестовый вариант путем объединения всех муниципальных регистров населения на контрольную дату, отобранную для анализа, каковой являлось 31 декабря 2012 года. Перед анализом сигналов из административных архивов необходимо определить отчетный период: присутствие того или иного лица на территории неизбежно связано с определенным временным интервалом. Выбор отчетного периода является ключевым фактором при выявлении и анализе сигналов присутствия: чем продолжительнее этот период, тем легче оценить вес сигнала с точки зрения «непрерывности» и «стабильности» во времени, число и повторение миграционных процессов, трудовую мобильность, курс обучения и т.д. В целях данного тестирования ИСТАТ использовал определение «места обычного жительства», содержащееся в Регламенте 1260/2013 Европейского союза. Исходя из этого ИСТАТ выбрал отчетный период с января года T по декабрь года T+1 с 31 декабря, определив в качестве контрольной даты 31 декабря года T. Период, использовавшийся для оценки сигналов, составил, таким образом, 24 месяца. Каждый сигнал можно было отнести к конкретному лицу и определенному географическому району; сигнал маркировался атрибутом, позволяющим отслеживать лицо в обоих первоначальных источниках, а также атрибутом, касающиеся продолжительности присутствия, выраженного в виде трудовой или учебной деятельности. В тех случаях, когда географическая информация отсутствует, используется условный код для обозначения отсутствия информации.

8. Сигналы могут непосредственно увязываться с индивидуальной единицей или относиться к другому лицу, имеющему связь с данной индивидуальной единицей; в этом последнем случае они являются «косвенными сигналами» и менее сильными, чем прямые сигналы. Кроме того, существует иерархия источников, в которой сигналы, поступающие из «регистров деятельности», касающихся трудовой деятельности и учебы, должны рассматриваться в качестве надежных сигналов, поскольку посещение школы или наличие работы гарантируют присутствие на территории.

9. Кроме того, регистры деятельности содержат более подробную информацию, такую как продолжительность и вид деятельности, которая предоставляет полезные характеристики для анализа. Использование стандартизированного временного периода в отношении продолжительности определенной деятельности позволяет составить схему ежемесячного присутствия. Данная характеристика представляет собой последовательность из 24 цифр, каждая из которых указывает, занималось ли лицо каким-то конкретным видом деятельности и в каком географическом районе страны. Группирование этих последовательностей, очевидных в большинстве сигналов о деятельности, позволяет выявление различных профилей непрерывности (рис. 1).

Рис. 1

**Схема ежемесячного присутствия на основе профилей непрерывности участия в трудовой и учебной деятельности**

Временной период с января 2012 года по декабрь 2013 года																								Профиль присутствия в регистрах занятости и учебной деятельности	
я	ф	м	а	м	июн	июл	а	с	о	н	д	я	ф	м	а	м	июн	июл	а	с	о	н	д		
[Solid black bar]																								1	Постоянное
[Solid black bar]																								2	Исходящие сигналы с присутствием в декабре 2012 года
[Solid black bar]																								3	Входящие сигналы с присутствием в декабре 2012 года
[Solid black bar]																								4	Сигналы присутствия около декабря 2012 года
[Dotted black bar]																								5	Непостоянное, не менее 12 месяцев
[Dotted black bar]																								6	Сезонные сигналы
[Dotted black bar]																								7	Менее 12 месяцев, несезонные
[Single black bar]																								8	Лишь один сигнал присутствия в декабре 2012 года
[Solid black bar]																								9	Присутствие только до декабря 2012 года
[Single black bar]																								10	Случайные сигналы только до декабря 2012 года
[Solid black bar]																								11	Входящие сигналы после декабря 2012 года и присутствия до 2013 года
[Solid black bar]																								12	Входящие сигналы после декабря 2012 года, но не до декабря 2013 года

10. Группа «постоянное» (профиль 1) включает в себя сигналы о непрерывном присутствии и максимальном постоянстве в течение всего отчетного периода. «Входящие» и «исходящие» сигналы (профили 2 и 3) с присутствием в течение более 12 месяцев, в том числе в декабре 2012 года, описывают лиц, которые изменили, начали или прекратили определенный вид деятельности или которые перемещаются по территории. Группа «Около декабря 2012 года» представляет собой сигналы, характеризующиеся непрерывным присутствием, начинающимся в 2012 году и заканчивающимся в 2013 году, причем на протяжении по меньшей мере 12 месяцев (профиль 4). Кроме того, релевантным паттерном является «Непостоянное, не менее 12 месяцев» (паттерн 5). Сигналы «сезонного присутствия» описывают случаи, когда схема паттерна повторяется один раз в год (паттерн 6). Существуют и некоторые менее конкретные паттерны, включая группы сигналов о лишь одномесечном присутствии (паттерн 8), сигналы о присутствии с перерывами продолжительностью менее 12 месяцев (паттерн 7) и, наконец, паттерны сигналов непрерывного присутствия, но толь-

ко в 2013 году. Для целей данного тестирования регистр населения и сигналы деятельности были загружены в базу данных, которая содержала также индивидуальные характеристики, такие как демографические признаки, которые оказались весьма полезными при анализе окончательных результатов.

### III. Группы населения, классифицированные по непротиворечивости между регистрами

11. Оценка непротиворечивости между изучавшимися регистрами обеспечивает классификацию всех случаев, зарегистрированных в экспериментальной базе данных, по группам, которые являются полезными для расчета данных о численности населения. Процесс формирования этих групп был организован в три этапа (см. таблицу 1). Первый этап предусматривал увязку экспериментального регистра населения (РН), сигналов о деятельности из регистров рабочей силы и учащихся (РПСУ) и регистра видов на жительство (РВЖ); на втором этапе, который был призван объяснить отсутствие сигналов о предыдущей деятельности, из числа всех лиц без сигналов в РПСУ определялись пенсионеры и лица, получающие другие пособия; и наконец, регистр налоговых деклараций (РПД) служил источником косвенных сигналов в отношении лиц, которые могли каким-либо образом оправдать свое присутствие в Италии.

Таблица 1

**Схема процесса и показатели численности населения (тыс. человек) в соответствии с основаниями или неопределенностью считаться обычно проживающими в Италии**

Этап I					Этап II			Этап III		
РН	РПСУ	РВЖ	Группа	Численность	РП-ЛПНП	Группа	Численность	РНД	Группа	Численность
61 068	37 704	3 378						20 764		
Сигналы					Сигналы			Сигналы		
Да	Да	–	<b>A</b>	<b>36 618</b>						
					Да	<b>E</b>	<b>14 485</b>			
Да	Нет	–	B	24 450	Нет	F	9 965	Да	<b>G</b>	<b>6 939</b>
									Нет	H
Нет	Да	–	C	1 086						
Нет	Нет	Да	D	351						

Источник: Наша собственная обработка данных ИСТАТ.

Условные обозначения в таблице:

<input type="checkbox"/>	имеющие основания считаться обычными жителями
<input type="checkbox"/>	неопределенность в отношении статуса обычных жителей

Условные обозначения административных источников:

<b>РН</b>	регистр населения
<b>РПСУ</b>	регистры рабочей силы и учащихся
<b>РВЖ</b>	регистр видов на жительство
<b>РП-ЛПНП</b>	регистр пенсионеров и лиц, получающих непенсионные пособия
<b>РНД</b>	регистр налоговых деклараций

12. На первом этапе увязка изучавшихся регистров позволяет определить четыре группы лиц. Группа А, которая насчитывает 36,6 млн. человек, охватывает лиц, зарегистрированных в РН и регистрах РСУ. Группа В, которая насчитывает около 24,5 млн. человек, охватывает лиц, зарегистрированных в РН, но не в источниках РПСУ. Эти лица не зарегистрированы в РН, но с учетом сигналов о деятельности из источников РПСУ включены в группу С, которая насчитывает почти 1,1 млн. человек. И наконец, группа D, которая насчитывает 351 000 человек, включает лиц, зарегистрированных только в РВЖ, но не в источниках РН или РПСУ. На втором этапе группа В была увязана с регистром пенсионеров и регистром лиц, получающих непensionные пособия (пособия по беременности и родам, по безработице и т.д.): в результате этой увязки были созданы два набора в зависимости от того, имели (группа Е) или нет (группа F) эти лица сигналы в этих двух регистрах. Таким образом, группа Е, насчитывающая около 14,5 млн. человек, охватывает лиц в РН с сигналами в регистре пенсионеров и регистре лиц, получающих непensionные пособия (РП-ЛППП); группа F, насчитывающая около 10 млн. человек, состоит из лиц, зарегистрированных в РН, без сигналов о получении пенсии или иных пособий.

13. На третьем этапе этого процесса группа F в конечном итоге была увязана с регистром налоговых деклараций (РНД) для выявления двух других групп населения на основе присутствия (группа G) или отсутствия (группа H) в источниках о лицах, заявленных в качестве иждивенцев для налоговых целей: «супруги», «дети» или «другие члены семьи». Таким образом, группа G, насчитывающая 6,9 млн. человек, охватывает лиц, которые зарегистрированы в РН, но лишь с косвенными сигналами из РНД. Группа H, насчитывающая 3 млн. человек, охватывает лиц, зарегистрированных в РН, но без сигналов из любых других источников.

14. Можно организовать вышеупомянутые группы в соответствии с оценкой по критерию обычного жительства. Таким образом, лиц, относящихся к группам А, Е и G, можно считать имеющими основания рассматриваться в качестве обычно проживающего населения. В свою очередь группы С, D и H могут считаться «неопределенными» жителями, требующими более тщательного анализа. По состоянию на конец 2012 года, около 3 млн. человек были зарегистрированы в группе H, которая представляет собой подгруппу населения, зарегистрированного в регистре населения (РН), без сигналов из других источников. Почти 75% этих лиц имеют итальянское гражданство. Половой состав выглядит хорошо сбалансированным как в случае итальянцев, так и иностранцев, и медианный возраст чуть превышает 40 лет (таблица 2). Распределение по возрастным группам этой подгруппы населения полностью коррелируется с возрастной структурой постоянного населения Италии в 2012 году. Доля итальянских граждан более зрелого возраста (35–64 года) примерно на 14% выше соответствующего показателя иностранцев. Кроме того, существует весьма значительная доля детей в возрасте до шести лет (немногим менее 15% в случае как итальянцев, так и иностранцев). Это объясняется рожденьями после представления налоговых деклараций за 2012 год. Около 400 000 детей не были зарегистрированы в качестве иждивенцев при подаче налоговых деклараций за 2012 год (они будут указаны в налоговых декларациях за 2013 год), и это определяет число детей в регистре населения/рождений, но без сигналов в административных источниках. Таким образом, данную подгруппу населения следует уменьшить на 400 000 человек, поскольку эти дети, несомненно, относятся к семейной ячейке, и в качестве таковых не являются «неопределенными» жителями.

15. Географическое распределение по гражданству (итальянцы в сопоставлении с иностранцами) свидетельствует о наличии значительных различий: в центральных и южных муниципалитетах зарегистрированы самые высокие показатели неопределенных жителей, как итальянцев, так и иностранцев. Итальянцы характеризуются высокой концентрацией во многих муниципалитетах Лацио, Кампании, Сицилии и Калабрии. В свою очередь иностранцы имеют более широкое географическое распределение. В дополнение к южным районам иностранцы проживали в муниципалитетах, расположенных в западной части страны и вдоль Тирренского побережья, особенно в районах Тосканы (рис. 2). На протяжении десятилетий муниципалитеты Тосканы являются излюбленным местом проживания многих пенсионеров из Северной Европы, особенно британских граждан (Warners et al, 1999).

16. Средний или высокий процент людей, имеющих итальянское гражданство, был обнаружен только в северных муниципалитетах вдоль границы со Швейцарией, Францией и Австрией, что объясняется присутствием трансграничных работников, которые обычно проживают в Италии.

17. Если проанализировать возрастную структуру групп гражданства каждой страны, то окажется, что китайская община характеризуется наименьшим медианным возрастом (всего 13 лет) по сравнению с итальянцами и другими иностранными общинами. Среди последних албанская община имеет самый высокий медианный возраст.

18. Кроме того, что касается иностранцев, то если соотношение полов (число мужчин из расчета на 100 женщин) согласуется с половым составом иностранных общин в Италии, распределение по стране гражданства этой подгруппы населения не воспроизводит точно географию иностранного присутствия в Италии. Например, иммигранты из Молдовы, Индии и Перу не входят в число десяти наиболее многочисленных национальных групп, проживающих в Италии (см. таблицу 2).

Таблица 2

**Демографические структуры подгруппы населения Н: страна гражданства, соотношение полов, возрастные группы и медианный возраст**

Страна гражданства	Абсолютные величины (тыс. чел.)	Процентная доля	Соотношение полов	Возрастные группы (в процентах)						Медианный возраст
				моложе 6 лет	6–14 лет	15–34 года	35–64 года	65 лет и старше	Всего	
<b>Всего</b>	<b>3 023</b>	<b>100,0</b>	<b>101,3</b>	14,3	1,5	22,7	55,3	6,3	100,0	<b>40,6</b>
Итальянцы	2 223	73,5	101,4	13,4	0,8	21,5	58,5	5,8	100,0	42,1
Иностранцы (из них)	800	100,0	101,3	17,0	3,5	26,6	44,7	8,1	100,0	35,7
Румыния	134	16,7	90,3	20,0	2,4	34,5	40,8	2,4	100,0	31,9
Марокко	72	9,1	159,6	19,7	2,9	22,5	46,9	8,0	100,0	36,7
Албания	60	7,5	96,1	16,0	1,4	19,0	40,6	23,0	100,0	48,9
Китай	34	4,3	107,7	40,2	10,0	15,2	31,3	3,2	100,0	13,2

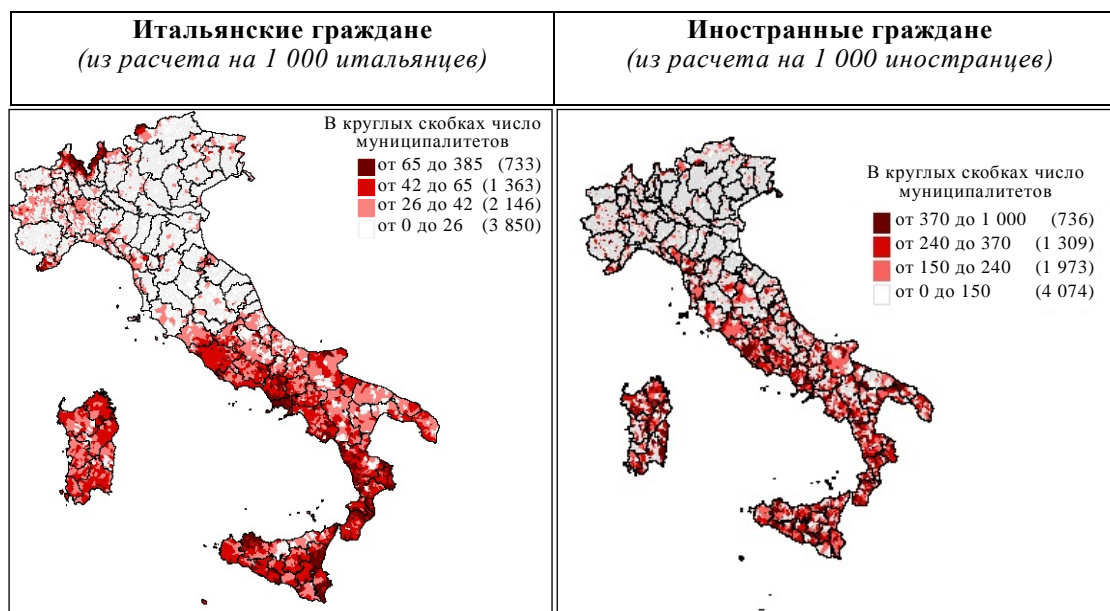


Страна гражданства	Абсолютные величины (тыс. чел.)	Процентная доля	Соотно- шение полов	Возрастные группы (в процентах)					Медианный возраст	
				моложе 6 лет	6–14 лет	15–34 года	35–64 года	65 лет и старше		Всего
Украина	20	2,6	45,3	10,5	2,3	24,8	56,9	5,5	100,0	40,6
Тунис	17	2,2	192,7	18,8	6,4	24,0	47,5	3,3	100,0	34,5
Египет	17	2,2	201,5	22,0	12,4	23,2	40,8	1,6	100,0	29,8
Польша	17	1,9	48,2	7,7	3,4	26,2	59,2	3,5	100,0	38,0
Филиппины	15	2,2	78,7	27,2	5,3	22,3	40,2	5,1	100,0	30,5
Нигерия	15	1,9	99,1	24,3	2,4	38,2	34,0	1,1	100,0	29,9
Другие страны	396	49,5	98,1	12,2	3,2	27,1	47,5	9,9	100,0	37,6

Источник: Наша собственная обработка данных ИСТАТ.

Рис. 2

**Подгруппа населения Н: лица, зарегистрированные в регистре населения (РН), но без сигналов из других источников**



Источник: Наша собственная обработка данных ИСТАТ.

19. В заключение следует отметить, что для целей непрерывной переписи эта подгруппа имеет существенно критическое значение. Ее численность, отсутствие сигналов в регистрах рабочей силы и учащихся и высокая концентрация в южных муниципалитетах служат значительными факторами риска для переписи. Кроме того, из-за продолжающегося экономического кризиса могут иметь место крупные миграционные потоки в другие страны, что в случае неуведомления регистрационных органов может привести к значительному избыточному охвату в регистрах населения.

20. Подгруппа лиц, не зарегистрированных в регистре населения (РН), с сигналами из других административных источников (группа С), насчитывает 1,1 млн. человек. Все лица в этой группе не зарегистрированы в РН, но имеют один или несколько сигналов из одной или более других административных ис-

точников. Группу С можно дополнительно разбить на подгруппы: большинство этих сигналов имеют источником записи о трудовой и учебной деятельности конкретного паттерна непрерывности, что позволяет учитывать вид и продолжительность соответствующей деятельности при анализе этого конкретного населения (см. таблицу 3).

Таблица 3  
**Группа С в зависимости от вида и продолжительности сигналов**

<i>Подгруппы</i>	<i>Источники и виды сигналов</i>	<i>Абсолютные величины</i>
<b>C1</b>	Работники	318 159
<b>C2</b>	Студенты	32 671
<b>C3</b>	Учащиеся начальной/средней школы	58 327
<b>C4</b>	Слабые сигналы присутствия	266 763
<b>C5</b>	Бесполезные сигналы присутствия или отсутствие ежемесячной информации	410 242
<b>Всего</b>		<b>1 086 162</b>

*Источник:* Наша собственная обработка данных ИСТАТ.

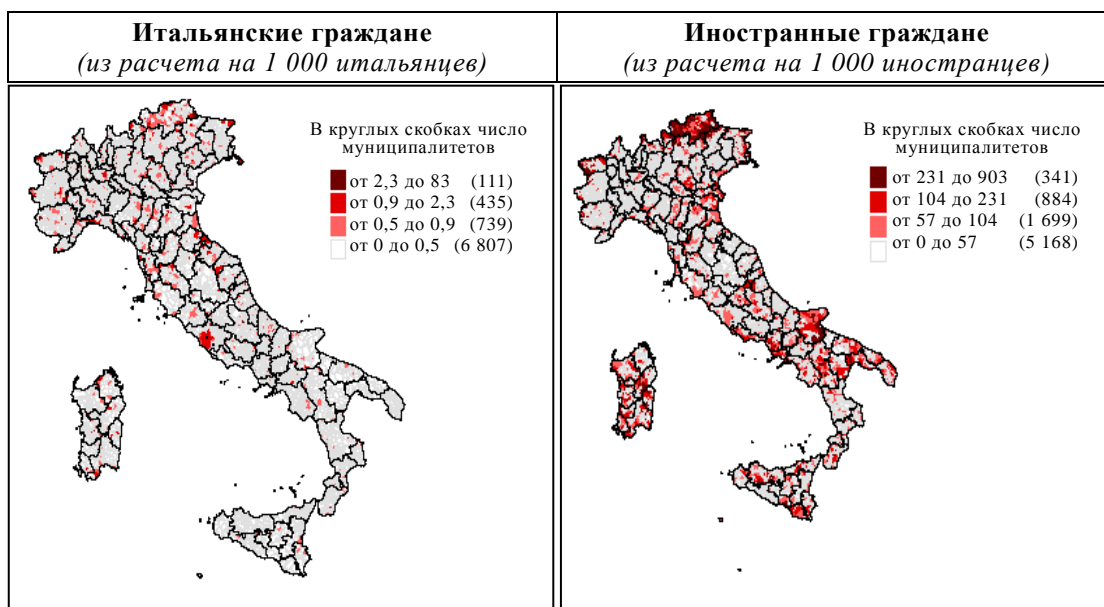
21. Эксплоративный анализ в сочетании с экспертным надзором позволяет произвести надлежащую классификацию, состоящую из трех основных типов паттернов непрерывности. Во-первых, имеются «сильные постоянные сигналы»: профили от 1 до 5 можно считать наиболее релевантными с точки зрения официального определения места обычного жительства с учетом того, что они указывают на по меньшей мере 12 месяцев присутствия, в том числе в декабре 2012 года. Затем идут «слабые непостоянные сигналы»: профили 6 и 7 ежемесячной схемы, которые включают в себя сезонные сигналы и несезонные, периодические сигналы с общей продолжительностью присутствия менее 12 месяцев. Остальные профили, с 8 по 12, определяются как «бесполезные виды сигналов» в данном тестировании, поскольку они не соответствуют определению места обычного жительства ЕС: к их числу относятся сигналы, которые исчезают до декабря 2012 года, или сигналы, которые появляются после декабря 2012 года или полностью случайные сигналы до декабря 2012 года. Последние заслуживают конкретного анализа в будущем, поскольку эти паттерны деятельности, тем не менее, могут быть связаны с некоторыми обычно проживающими лицами.

22. В результате классификации лиц из группы С по виду и продолжительности сигнала было сформировано пять подгрупп (таблица 3). Около 410 000 случаев не являются полезными для анализа (группа С5), поскольку речь идет о лицах с «бесполезным видом сигнала» или лицах, не имеющих ежемесячной информации. Сильно-стабильные сигналы имеют в общей сложности 409 157 лиц: 78% из них (318 000) составляют работники (группа С1), поскольку источником этих сигналов является регистр рабочей силы; остальные 22% приходятся на учащихся начальной или средней школы (58 000) и студентов (около 33 000). На эти группы, по всей видимости, приходится наибольшее число не учтенных в РН лиц: иными словами, речь идет о чрезвычайно важной для расчета официальных данных о численности населения подгруппе. К категории «слабые непостоянные сигналы» (группа С4) относятся 266 763 лиц: хотя эти сигналы имеют общую продолжительность меньше минимального срока, установленного для регистрации в официальном РН, эти случаи требуют дальнейшего анали-

за, поскольку сезонные или прерывистые паттерны могут быть связаны с конкретным видом трудовой деятельности или учебы, а не с характеристиками проживания. Группы С2 и С3 представляют собой два конкретных подмножества лиц, каждое из которых соответствует конкретному регистру, и соответствующим временным паттерном является посещение по крайней мере одного учебного курса. Важно подчеркнуть, что около 90% лиц в этих группах являются иностранцами, а не итальянцами: группа С3 включает в себя детей-иностранцев, обучающихся в итальянских школах без регистрации в официальном регистре населения; группа С2 включает в себя иностранных студентов, которые еще не зарегистрированы в регистре населения. Группа С1 с сильными постоянными сигналами из регистров рабочей силы, может соответствовать более чем одному конкретному подмножеству лиц, поскольку их сигналы имеют своим источником многочисленные различные регистры, касающиеся конкретного типа рабочих мест. Кроме того, она охватывает широкий спектр стран гражданства и возрастных групп (от 18 лет). Географическое распределение лиц в группе С1 в соответствии с гражданством свидетельствует о незначительной численности итальянцев в муниципалитетах вдоль северной границы, вокруг столицы Рима и Республики Сан-Марино (рис. 3).

Рис. 3

**Подгруппа населения С1: лица, не зарегистрированные в регистре населения (РН) с сигналами из других источников**



23. Иностранцы широко представлены в регионе Трентино-Альто-Адидже и в муниципалитетах сельскохозяйственных районов долины По, Апулии, Кампании, Сицилии и Сардинии.

24. Множественный анализ соответствий используется для нахождения связей между категориями имеющихся переменных, которые могут дать полезную информацию о конкретных подмножествах. Переменными, используемыми на этом этапе, являются пол, возраст, гражданство, провинция и тип рабочего места. Проявляются два важных дискриминанта (рис. 4): горизонтальную ось на диаграмме соответствий можно толковать как демографический профиль с увеличением возраста справа налево и с левыми четвертями, характеризующимися присутствием женщин, и с правыми, характеризующимися присутствием муж-

чин; вертикальная ось, как представляется, в большей степени связана со статусом в занятости и сектором трудовой деятельности: секторы домашней работы и сельского хозяйства находятся в нижней части диаграммы, в то время как другие работники государственного или частного сектора сгруппированы в верхней части. Данная диаграмма указывает на существование различных групп: 1) молодые мужчины, работающие в сельском хозяйстве, в основном выходцы из Северной Африки и частично из Албании и бывшей Югославии, сигналы которых расположены в провинциях юга Италии, где сельское хозяйство является главным сектором занятости; 2) другая группа состоит в основном из женщин, работающих в качестве нянь или домработниц, в возрасте старше 35 лет, являющихся выходцами из Филиппин, Восточной Европы (бывшего СССР) или Южной Америки, с сигналами, расположенными в провинциях, в которых имеются крупные города; 3) группа, состоящая из выходцев из стран Западной и Центральной Европы, которые являются работниками наемного труда, самозанятыми, консультантами или нанятыми учреждениями временной занятости, с присутствием лиц в возрасте старше 65 лет и сигналами, расположенными в центральных и северных провинциях Италии; 4) китайцы представлены в том же квадранте, что и выходцы из Западной Европы; 5) румыны не имеют половых или возрастных особенностей, но связаны с секторами сельского хозяйства и домашней прислуги; 6) итальянцы, французы и швейцарцы, находящиеся в этой группе, связаны с провинциями севера, на границе Италии, что предполагает, что они, вероятно, являются трансграничными работниками, живущими за границей, но, как правило, работают в Италию и, следовательно, не зарегистрированы в РН, даже несмотря на то, что они имеют мощные сигналы в регистрах рабочей силы.

Рис. 4  
**Подгруппа С1: иностранные работники с постоянными сильными сигналами и их связь между категориями**



25. Такого же рода анализ проводился по группе С4, которая охватывает лиц со «слабыми непостоянными сигналами»: результаты указывают на те же факторы и те же связи между категориями, что и в случае группы С1 (рис. 5). Представляется целесообразным проводить анализ обеих групп – С1 и С4, с сильными и слабыми сигналами одновременно. К числу релевантных факторов, помимо демографического характера и статуса в занятости, добавляется постоянство сигнала, которое выступает как новый существенный фактор, характеризующийся контрастом между сильными постоянными сигналами в одном направлении и слабыми, непостоянными сигналами в противоположном направлении. Подмножества лиц, как упоминалось выше, во многом являются схожими, однако можно определить сильную связь между непостоянными сигналами и конкретным видом трудовой деятельности и гражданством: слабые, непостоянные или периодические паттерны тесно связаны с сельским хозяйством и молодыми людьми – выходцами из Румынии, Албании и других стран бывшей Югославии.

Рис. 5

**Подмножества сильных постоянных сигналов и непостоянных сигналов: связь между категориями**



26. Группа D насчитывает около 350 000 человек и определяется только сигналами из регистра видов на жительство (см. таблицу 1). Таким образом, «неопределенные жители» в группе D все являются иностранцами выходцами из стран, не входящих в ЕС, и не зарегистрированными в РН. Что касается информации, содержащейся в этом источнике, то срок действия вида на жительство может использоваться в ежемесячной схеме, которая подвергалась обработке в отношении сигналов, содержащихся в других административных источниках. Данное преобразование позволило выявить иностранцев, которые удовлетворяли требованиям в отношении пребывания в Италии по состоянию на 31 декабря

2012 года в течение по меньшей мере 12 месяцев. И наконец, около 61,5% лиц в группе D могут рассматриваться в качестве обычных жителей, даже если они не имеют сигналов о трудовой деятельности или учебе или другой информации, подтверждающей их присутствие на территории Италии.

#### IV. Выводы

27. Методика извлечения знаний из баз данных оказалась весьма эффективной в целях эксплуатации потенциала огромного объема информации, содержащегося в административных источниках. Однако дополнительным подспорьем для процесса принятия решений ИСТАТ может стать интеграция процесса извлечения знаний из баз данных с моделями прогнозирования. Выбор конкретных административных источников из всех имеющихся в наличии и определение их иерархии являются важнейшими элементами: регистры рабочей силы и учащихся занимают более высокое место в иерархии по сравнению с другими изученными источниками, особенно с точки зрения объема более подробной информации о территориальном уровне и продолжительности деятельности.

28. Признаки присутствия на территории извлекаются из источников и могут эффективно использоваться для повышения качества регистров населения. Вместе с тем, поскольку эти сигналы могут также описывать временное или случайное присутствие, необходимо провести процесс характеристики путем построения производных переменных, которые затем позволят выявлять случаи постоянного присутствия, которые соответствуют определению места обычного жительства международных правил.

29. Результаты данного тестирования свидетельствуют о том, что непостоянные сигналы присутствия присущи определенным секторам трудовой деятельности (в частности, сезонным, сельскохозяйственным работам). Таким образом, классификация сигналов должна осуществляться на основе более точных и многомерных критериев, при использовании которых вес сектора трудовой деятельности не будет вносить никаких искажений в оценку численности обычно проживающего населения. На основе административных источников могут быть рассчитаны «косвенные» сигналы, опирающиеся на связи между лицами. Они являются полезными для оценки присутствия на территории лиц, которые не демонстрируют прямых сигналов, поскольку они не работают или не учатся и не получают другого дохода (например, находящиеся на иждивении члены семьи).

30. В ходе данного тестирования ИСТАТ определил предварительный рабочий подход к интеграции использования административных источников и официальных регистров населения для расчета численности обычно проживающего населения. Этот подход позволяет определить группу лиц, подлежащих включению в состав обычно проживающего населения Италии на заданную контрольную дату. Эта группа возможных обычных жителей насчитывала 62,6 млн. человек в 2012 году (по состоянию на 31 декабря) и 62,4 млн. человек в 2013 году. Из этой группы лиц можно выделить три основные подгруппы:

- а) подгруппа, присутствующая в регистре населения, без сигналов из других источников (3,0 млн. человек в 2012 году);
- б) подгруппа, присутствующая в регистре населения, с сигналами из других источников (58,1 млн. человек в 2012 году);

с) подгруппа, не присутствующая в регистре населения, но с сигналами из других источников (1,5 млн. человек в 2012 году).

31. Демографические переменные, особенно пол, возраст и страна гражданства, а также местонахождение сигнала на территории, продемонстрировали свою высокую значимость для определения профилей конкретных подгрупп населения. Более подробный анализ характеристик подгрупп населения, подверженных риску чрезмерного и недостаточного охвата, высветил некоторые важные элементы и конкретные подмножества лиц. Разделение на подгруппы и подмножества представляет большой интерес: с одной стороны, оно позволяет идентифицировать подгруппы, требующие дальнейшего тематического анализа (например, иностранные общины, которые обычно избегают регистрации, но занимаются конкретными видами трудовой деятельности, или лица, посещающие определенные территории); с другой стороны, эти же группы могут послужить основой для определения стратегии проведения переписи, опирающейся на смешанную методику, сочетающую в себе конкретные обследования и соответствующие статистические модели.

32. Что касается 3 млн. человек, которые являются результатом потенциально избыточного охвата, то три четверти из них имеют итальянское гражданство, а иностранцы в среднем более чем на шесть лет моложе итальянцев. В данной подгруппе представлены в основном лица трудоспособного возраста (от 15 до 64 лет). Помимо домохозяйек, в силу своего географического распределения эта подгруппа населения связана с географическими районами, где более высок уровень безработицы, такими как муниципалитеты юга и некоторые районы центральной части страны.

33. Что касается недостаточного охвата, то анализ позволил выявить наличие нескольких конкретных подмножеств. Непрерывные сигналы, указывающие на стабильное присутствие на территории, касаются чуть более 400 000 человек, которые, по всей видимости, представляют собой неучтенное в регистре населения подмножество: речь идет главным образом об иностранных гражданах. Географическое местоположение и конкретное гражданство имеют важнейшее значение для выявления трансграничных работников, для которых допустимо отсутствие в регистре населения. Анализ слабых (непостоянных) сигналов подчеркнул тот факт, что некоторые из них могут быть связаны с лицами, стабильно присутствующими на территории, и по этой причине представляется целесообразным улучшить характеризацию прямых сигналов.

34. Анализ силы сигнала на основе его непрерывности во времени является лишь отправной точкой в использовании продольных данных, которые могут обрабатываться вместе с административными источниками. Основная цель будущих тестов должна, таким образом, заключаться в изучении возможностей использования продольных моделей в отношении ряда лет для расчета оценок численности подгрупп населения, которые являются более стабильными по отношению к колебаниям на рынке труда, с которых поступают сигналы.

## Справочные материалы

Argüeso A., Vega, J.L. (2014). A population census based on registers and a «10% survey» methodological challenges and conclusions. *Statistical Journal of the IAOS*. 30(1): 35-39.

Bonifazi C., Martini C. (2014). The Impact of the Economic Crisis on Foreigners in the Italian Labour Market. *Journal of Ethnic and Migration Studies*. 40(3).

Cibella N., Gallo G., Pezone A., Tuoto T. (2015). The integration between the 2011 census Post Enumeration Survey Data and Administrative Data. The Analysis on Hard-To-Count Population. Paper presented to the Population Days Conference. Palermo: 4-6 February.

Citro, Constance F. (2014). From multiple modes for surveys to multiple data sources for estimates. *Survey Methodology*. 40(2): 137-161.

Crescenzi, F., Sindoni, G. (2015). The Combined Use of Multiple Data Sources in the Population Census. Paper presented to the Unece Group of Experts on population and Housing Censuses. Geneva: 30 September – 2 October.

Di Bella G., Ambroselli, S. (2014). Towards a more efficient system of administrative data management and quality evaluation to support statistics production in Istat. Paper presented to the European Conference on Quality in Official Statistics Q2014. Vienna: 2-5 June.

EMN European Migration Network (2012). Practical responses to irregular migration: the Italian case. Edited by EMN National Contact Point, Idos, Rome, 2012. See: [http://ec.europa.eu/dgs/home-affairs/index\\_en.htm](http://ec.europa.eu/dgs/home-affairs/index_en.htm).

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*, U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Eds. AAAI/MIT Press: Cambridge.

Gallo G., Paluzzi E., Benassi F. (2014). The 2011 Italian experience towards supported-census for measuring migration. Paper presented to the Unece Wok Session on Migration Statistics. Chişinău Republic of Moldova: 10-12 September.

ISTAT, 2014. La misurazione della qualità del 15° Censimento generale della popolazione e delle abitazioni: i risultati dell'indagine di copertura (PES). Seminario del 27 giugno, Roma, <http://www.istat.it/it/archivio/126014>.

Jensen, P. (1983). Towards a register based statistical system- some Danish experience. *Statistical Journal*. 1(3): 341-365.

Lanzieri, G. (2013). On a New Population Definition for Statistical Purposes. Paper presented to the Fifteenth Meeting of Group of Experts on Population and Housing Censuses. Geneva: 30 September – 3 October 2013.

ONS Office for National Statistics (2003). *Census Strategic Development Review – Alternatives to a Census. Review of International Approaches*. Information Paper. London: United Kingdom. Office for National Statistics

Poulain M., Herm A. (2013). Le register de population centralisé, source de statistiques démographiques en Europe. *Population*. 68(2): 215-247.

Statistics Canada (2011). Preliminary Report on Methodology Options for the 2016 Census. <http://www12.statcan.gc.ca/strat/Preliminary%20Report%20on%20Methodology%20Options%20for%20the%202016%20Census.pdf> (accessed April 17, 2016).

Statistics Denmark (1995). *Statistics on persons in Denmark – a register-based statistical system*. Eurostat: Luxembourg.

Statistics Finland (2004). *Use of registers and administrative data sources for statistical purposes– best practices in Statistics Finland*. Handbook 45. Statistics Finland: Helsinki.



UNECE (2007). Register-based statistics in the Nordic countries: review of best practices with focus on population and social statistics. United Nations Publication, ISBN 978-92-1-116963-8.

UNECE (2011). Using Administrative and Secondary Sources for Official Statistics - A Handbook of Principles and Practices. United Nations Publication.

UNECE (2013). Population Definitions at the 2010 Censuses Round in the Countries of the UNECE Region. Paper presented to the Fifteenth Meeting of Group of Experts on Population and Housing Censuses. Geneva: 30 September – 3 October 2013.

Wallgren A., Wallgren B. (2011). To understand the Possibilities of Administrative Data you must change your Statistical Paradigm! Proceedings of the Survey Research Methods Section. American Statistical Association, Invited Papers.

Warners A.M., King R., William A.M. and Patterson G. (1999). The well-being of British expatriates retirees in southern Europe. *Ageing and Society*. 19(6): 717-740.

Zhang, Li-Chun (2012). Topics of statistical theory for register-based statistics and data Integration. *Statistica Neerlandica*. 66, (1), 41-63.

---