

Распространение: Общее
27 марта 2012 г.

Оригинал: английский

Европейская экономическая комиссия

Конференция статистиков Европы

Заседание экспертной группы ЕЭК ООН – Евростат по проведению переписи с использованием регистров

Женева, 22-23 мая 2012 г.

Сессия 3: Наличие, полнота и качество данных из регистров и других источников

Использование регистров населения в 15-й переписи в Италии: вызовы и предварительные результаты

Записка ISTAT, Италия¹

Резюме

Одним из главных нововведений Итальянского национального института статистики (ISTAT) во время прошлой переписи населения, проведенной в октябре 2011 г., было использование административных регистров для содействия проведению переписи. Данный документ имеет две цели: а) задокументировать усилия, предпринятые ISTAT в целях обеспечения пригодности регистров для использования в целях проведения переписи, и показать основные вызовы, с которыми столкнулся ISTAT и связанные с очисткой, интеграцией и стандартизацией данных, и б) представить некоторые предварительные, основанные на имеющихся на сегодня метаданных переписи, результаты в отношении охвата муниципальных регистров населения, а также степени успешности вспомогательных регистров в устранении недостатков муниципальных регистров населения.

¹ Раздел II подготовлен Anna Pezone и Federica Pellizzaro (SCD/F), раздел III.1 подготовлен Daniela Casale (SCD/C), раздел III.2 подготовлен Luca Mancini (MTC/A), раздел IV подготовлен Luigi Marcone (MTC/A).

I. Введение

1. Одним из главных нововведений Итальянского национального института статистики (ISTAT) во время прошлой переписи населения, проведенной в октябре 2011 г., было использование административных регистров для содействия проведению переписи. Во время проведения переписи счетчики не посещали миллионы домохозяйств, этим домохозяйствам переписные листы были направлены по почте на адрес, указанный в муниципальных регистрах населения (*Liste anagrafiche comunali*, далее - LAC) того муниципалитета, где проживал глава домохозяйства.
2. Ошибки в охвате вследствие рождений и смерти, а также изменений адреса, произошедших между датой рассылки и датой переписи, устранялись путем инструктирования счетчиков о действиях, которые необходимо предпринять в случае отклонений во время работы на местах.
3. В соответствии с законодательством необходимо постоянно обеспечивать актуальность LAC для того, чтобы в любое время можно было получить точную картину о населении, проживающем в пределах муниципалитета. Однако случаи проживания домохозяйств или отдельных лиц, не зарегистрированных в муниципалитете, и/или их проживания не по адресу, по которому они обычно проживают, не столь редки.
4. В целях определения их места проживания и учета использовались вспомогательные административные регистры, в том числе Национальный налоговый регистр, архив разрешений на проживание и адресный регистр.
5. Данный документ имеет две цели: а) задокументировать усилия, предпринятые ISTAT в целях обеспечения пригодности регистров для использования в целях проведения переписи, и показать основные вызовы, с которыми столкнулся ISTAT и связанные с очисткой, интеграцией и стандартизацией данных, и б) представить некоторые предварительные, основанные на имеющихся на сегодня метаданных переписи, результаты в отношении охвата муниципальных регистров населения, а также степени успешности вспомогательных регистров в устранении недостатков муниципальных регистров населения.

II. *Liste Anagrafiche Comunali*

6. LAC – муниципальный регистр, содержащий информацию на уровне лиц и домохозяйств по всем лицам, зарегистрированным в качестве резидентов в определенном муниципалитете в определенный момент времени. В месяцы, предшествовавшие проведению переписи, LAC служили ряду целей, в том числе: а) определение соответствующих единиц переписи (домохозяйство или сожительство); б) распечатка, рассылка и создание пароля для доступа к переписным листам через Интернет; в) предварительная загрузка данных в систему управления обследованиями (*Sistema di gestione della rilevazione*, далее – SGR).
7. ISTAT определил единый протокол передачи данных для всех муниципалитетов со следующей технической спецификацией: а) название, структура и формат файлов, включая инструкции для декодирования переменных, таких как «связь», «семейное положение» и «пол»; б) характеристики, формат и длина переменных.

8. Электронные файлы отправлялись с использованием безопасного протокола (https) в целях обеспечения неприкосновенности частной жизни. Исполнение данных технических спецификаций обеспечивалось ISTAT посредством внедрения веб-приложения (Starlac), которое применялось для сбора и проверки данных, передаваемых муниципалитетами. Starlac позволяет проверять данные на два вида ошибок на этапе ввода информации: а) выборочный контроль на несоответствие требуемым техническим стандартам; б) систематический контроль на несоответствие правовым нормам касательно регистров населения (Regolamento Anagrafico)².

9. Только те регистры населения, которые полностью соответствовали техническим спецификациям и имели ошибки типа б) в пределах области допустимых значений, далее проверялись на однородность и правильность переменных и обрабатывались в следующем порядке:

- а) Определение единиц переписи (домохозяйства или сожительства);
- б) Подстановка глав домохозяйств (в случае отсутствия такового или наличия нескольких глав домохозяйства);
- в) Стандартизация и геокодирование адресов в соответствии с кодами зон действия системы регистрации;
- г) Выборка домохозяйств из выборочных зон действия системы регистрации для получения длинного переписного листа;
- д) Выбор модели переписного листа (_6р или _3р), исходя из размера домохозяйства;
- е) Присвоение уникального идентификационного номера переписного листа;
- ж) Создание пароля для доступа к онлайн-овому переписному листу;
- з) Распечатка и рассылка переписных листов;
- и) Загрузка данных в SGR.

10. Итоговая база данных включала примерно 61 млн. человек, 25 млн. домохозяйств и около 300 тыс. сожительства. Более 93% всех муниципалитетов успешно передали данные с первой попытки. Лишь один из 8 094 муниципалитетов не направил LAC, в то время как 77 муниципалитетов были вынуждены отправить данные повторно в связи с ошибками. Во время работ по сбору и обработке данных муниципалитетам оказывалось содействие со стороны специальной группы, оказывающей поддержку по устранению проблем по телефону.

11. Для проведения всех необходимых подготовительных работ, указанных выше, муниципалитеты должны были передать свои регистры населения в ISTAT в период с января по середину февраля, которые содержали информацию по всем проживающим гражданам по состоянию на 31 декабря 2010 г. Для обновления SGR в части всех демографических изменений, таких как рождения, смерти или изменения адреса, произошедших в период между концом 2010 г. и датой переписи (9 октября 2011 г.) ISTAT выработал три разные стратегии: 1) муниципалитетам с численностью населения менее 5 тыс. человек разрешалось вручную загрузить все

² В *Regolamento Anagrafico* (D.P.R. n.223 del 30 maggio 1989) установлены стандарты учета в LAC. Вследствие разных темпов перехода муниципалитетов от бумажных к электронным регистрам, процесс гармонизации все еще не закончен.

демографические изменения, произошедшие до 8 октября, непосредственно в SGR; 2) муниципалитетам с численностью населения от 5 тыс. до 20 тыс. чел. также предоставлялась возможность ручной загрузки, а также другая альтернатива – представить полный архив повторно; 3) муниципалитеты с численностью населения свыше 20 тыс. человек должны представить полный обновленный регистр. Около 5600 муниципалитетов загрузили изменения вручную непосредственно в SGR. Около 2 400 муниципалитетов (50 млн. человек и 20 млн. домохозяйств) направили регистры повторно на 8 октября 2011 г. Для этих муниципалитетов обновление SGR было проведено ISTAT. Около 100 муниципалитетам пришлось повторно пересылать данные вследствие следующих ошибок: а) файлы данных по-прежнему были датированы 31/12/2010 г.; 2) файлы данных содержали лишь демографические изменения; с) ошибки в кодировании или формате некоторых основных ключей связей (например, код семьи или адрес). Как и после представления данных в январе, данные рассматривались на предмет содержания ошибок.

III. Интеграция LAC: Вспомогательные регистры населения

A. *Liste Integrative Anagrafiche Comunali*

12. В данном разделе описывается роль вспомогательных муниципальных регистров населения (*Liste Integrative Anagrafi Comunali* или LIAC) в автоматизированных процедурах, которые были задействованы в целях содействия проведению 15-й переписи в Италии.

13. Во-первых, в целях содействия процессу была разработана реляционная база данных. Для каждого объекта (лица, домохозяйства и сожительства) были созданы две таблицы: первая таблица с исходными данными муниципальных регистров населения, а вторая – с выходными данными и показателями. Максимальный результат в отношении однородности данных был достигнут путем применения методов очистки и проверки. На этапе сбора применялись методы стандартизации (например, четко определенные форматы для всех областей) и единые правила (например, в отношении обработки исключений в связи с отсутствующими данными). В частности, применялись тщательные средства контроля в отношении тех атрибутов, которые должны были использоваться в качестве ключей связи между двумя выпусками муниципальных регистров населения, таких как индивидуальный идентификатор, налоговый код (*codice fiscale*), код домохозяйства/сожительства, код топонима.

14. Во-вторых, после прохождения формальных средств контроля, данные подвергались проверке на непротиворечивость на нескольких уровнях: отдельная запись, вид проживания, муниципалитет. Для каждой ошибки на каждом уровне рассчитывался показатель в соответствии со следующей таксономией: а) показатели на уровне записи отдельного лица, касающиеся ошибок, связанных с кодами идентификаторов; б) показатели на уровне записи отдельного лица, касающиеся ошибок, связанных с другими индивидуальными атрибутами; с) показатели на уровне домохозяйств; и d) показатели на уровне муниципалитета.

15. Таксономия изменений между двумя выпусками регистров населения с их распространенностью представлена в Таблице 1:

Таблица 1: Характер и масштаб обновления LAC

Тип изменений	Домохозяйства	Лица
Новые домохозяйства	662 931	1 099 563
Домохозяйства, у которых добавлен, как минимум, один компонент	356 034	851 870
Домохозяйства, у которых исключен, как минимум, один компонент	947 764	1 183 497
Домохозяйства, у которых изменился адрес проживания	232 416	380 372
Исключенные домохозяйства	523 911	523 911
Домохозяйства с новой главой	139 840	263 771

16. Процесс объединения был организован поэтапно в соответствии с логической последовательностью. Каждый шаг является подготовкой к следующему шагу и обеспечивает детальный протокол по результатам его обработки. Структура протокола такова, что позволяет проводить как простые виды автоматического анализа, так и сложные статистические исследования в целях принятия решения о продолжении процесса каждый раз по завершению определенного шага. Кроме того, каждый шаг включал сравнения с информацией, полученной из признанных внешних источников, в целях проведения количественных проверок на непротиворечивость на основании пороговых значений, установленных в зависимости от численности населения муниципалитетов.

17. На первом этапе изменения определялись на уровне лиц: основным ключом для привязки записи был индивидуальный код, за которым следовал налоговый код и затем один за другим до соединения нескольких атрибутов. Для определения изменений по проживанию домохозяйств и сожительств была принята более сложная процедура, с использованием как исходных и нормализованных адресов, так и кода топонима.

18. Эффективность каждого ключа связи оценивалась и фиксировалась для каждого муниципалитета в отдельности в соответствии с результатами процедурами сверки учетных данных³. Полученные LIAC были сохранены в двух таблицах, одна – для домохозяйств, а вторая – для сожительств, которые включали все типы изменений, как на уровне отдельных лиц, так и на уровне домохозяйств/сожительств. Такие таблицы содержат ключи связи с двумя справочными LAC, а также все атрибуты регистра. В заключение, основные сложности возникли в связи со средствами контроля по адресам проживания. Однако этот аспект должен улучшиться после выравнивания, которое будет осуществляться Управлениями регистров в соответствии с рядом правил, установленных ISTAT.

В. LIFA и RNC

19. Несмотря на то, что в соответствии с законодательством LAC должны постоянно обновляться для того, чтобы в любое время можно было получить точную картину о населении, проживающим в пределах муниципалитета, не столь редки случаи проживания домохозяйств или отдельных лиц, не зарегистрированных в

³ Ожидается, что эти меры эффективности будут полезными для будущих обследований, особенно учитывая идею «скользящей переписи», рассматриваемой в настоящее время ISTAT.

муниципалитете, и/или их проживания не по адресу, по которому они обычно проживают. Может оказаться, что домохозяйства или лица проживают в муниципалитете, при этом в соответствующем регистре нет соответствующей записи. С другой стороны, имеют место случаи, когда домохозяйства или лица регулярно регистрируются в LAC, но при этом проживают в другом муниципалитете. Причины таких отклонений могут быть разными. Например, для неитальянцев, бюрократия и трудности, связанные с получением регистрации, обычно отбивают желание подавать заявление. С другой стороны, для итальянцев выбор места, где они фактически не проживают, в качестве официального адреса, зачастую служит для того, чтобы избежать местного налогообложения.

20. Эти недостатки побудили ISTAT рассмотреть использование во время переписи вспомогательных регистров из других источников в качестве дополнения к LAC и LIAC. Что касается LIAC, информация была использована для того, чтобы помочь счетчикам определить местонахождение домохозяйств или отдельных респондентов, подверженных риску оказаться исключенными из переписи. Для каждого муниципалитета было создано два альтернативных архива: а) перечень лиц (LIFA), которые были либо незарегистрированными, или в случае если были зарегистрированы, использовали второй адрес в другом муниципалитете; б) перечень действительных адресов (RNC), которые не были связаны с какими-либо зарегистрированными домохозяйствами.

21. LIFA являются результатом ряда действий во привязке записей между следующими архивами: а) LAC по состоянию на 31/12/2010 г.; б) налоговый регистр (AT) по состоянию на 30 апреля 2011 г., который использовался, в основном, для получения вторых адресов (*domicilio fiscale*) для граждан Италии, а также других граждан Европы, проживающих в Италии; в) регистр разрешений на пребывание (PS) по состоянию на 30 ноября 2010 г. для лиц с неевропейским паспортом и постоянным разрешением на проживание в Италии⁴.

22. Архив PS был получен после процесса очистки данных, исключения избыточных данных, геокодирования и геопривязки, примененного примерно к 4 млн. отдельных записей, собранных Департаментом государственной безопасности (ДГБ) и Департаментом гражданских прав (ДГП) Министерства внутренних дел. Исходные данные хранятся в 11 различных и частично совпадающих базах данных в зависимости от этапа обработки каждого отдельного заявления⁵. В конце процесса было определено и исключено около 700 тыс. дублирующих записей в соответствии

⁴ В случае если какое-либо лицо оказывалось в LIFA как из PS, так и AT, более актуальная запись имела из последнего из указанных двух источников, которая сохранялась после окончательного исключения избыточных данных.

⁵ На момент получения, базы данных ДГБ содержали информацию в отношении порядка 185 тыс. вновь заполненных заявлений, находящихся на этапе предобработки; 2 600 тыс. держателей постоянного разрешения; 500 тыс. заявлений, находящихся на рассмотрении; 9 500 отклоненных заявлений; 755 тыс. истекших разрешений; 39 тыс. несовершеннолетних, зарегистрированных по разрешению своих родителей; 57 тыс. разрешений на работу, выданных лицам, приехавшим впервые; и 235 тыс. разрешений, выданных в целях воссоединения семьи. ДГП отвечает за три дополнительных архива, включающих информацию по около 250 тыс. ранее «незаконных» мигрантов, подавших заявления на «окончание учебного заведения» держателям постоянных разрешений; 131 тыс. заявлений на воссоединение семьи; и 122 тыс. заявлений на продление/замену разрешений на работу.

с согласованным протоколом правил исключения, в результате чего в окончательный архив PS вошло около 3,3 млн. записей⁶.

23. Налоговый регистр AT собирает базовую персональную информацию по всем лицам, которым присвоен налоговый код, включая сообщаемый ими *domicilio fiscale*⁷, и содержит около 80 млн. записей. После исключения всех записей, носящих временный характер (например, зарубежные артисты, приезжающие на выступление и затем уезжающие из страны) или ссылающихся на уже умерших лиц, оставшиеся записи подверглись процессу проверки, в ходе которого записи сохранялись в архиве только в случае, если информация подтверждалась другими административными регистрами, такими как архивы студентов, регистры предприятий, регистры родительных отделений, архивы пенсионеров и другие.

24. В целом, привязка записей между AT и LAC (RL1) явилась детерминированным слиянием с использованием налогового кода в качестве ключа сравнения на национальном уровне⁸. Привязка записей между PS и LAC (RL2) потребовала более сложной стратегии, как детерминированной, так и вероятностной, принимая во внимание более низкое качество имеющихся ключей сравнения для неевропейских резидентов. С учетом количества задействованных записей, привязка обычно осуществлялась на уровне региона (*regione*). Однако, для густо населенных местностей было необходимо снижать уровень проведения анализа до уровня провинции (*provincia*) или даже до муниципального уровня. LIFA был получен путем добавления непарных (несовпадающих) оставшихся записей из RL1 и RL2 и удаления повторяющихся данных в рамках завершающего шага по устранению избыточных данных⁹. В окончательный архив вошло чуть более 2,1 млн. записей, равномерно разделенных между PS и AT. Из них порядка одной трети – это лица, зарегистрированные в LAC, но использующие второй адрес, в то время как другие две трети были представлены лицами, не зарегистрированными в каком-либо LAC.

25. RNC является базой данных действительных адресов с привязкой к местности с геокодами зон действия системы регистрации для всех муниципалитетов с численностью населения 20 тыс. человек и более. Несмотря на то, что первоначально RNC был спроектирован для использования в переписи зданий, он также использовался вместе с LIACs и LIFAs для определения местонахождения незарегистрированных домохозяйств и сокращения ошибок неполного охвата LAC. В этих целях в SGR было загружено и предоставлено счетчикам порядка 2,6 млн.

IV. Предварительные результаты

26. На момент подготовки данного доклада работа по проведению переписи подходит к завершению, но все еще продолжается во многих муниципалитетах, в

⁶ См. дополнительную информацию в Fortini *et al.* (2011) *Counting in non-Italian residents: The use of the "Permits to Stay" archive in the next population census*. В трудах второй конференции ITACOSM по методам обследований, 27-29 июня 2011 г., Пиза, Италия: Plus-Pisa University Press

⁷ Статья 43 Гражданского кодекса Италии определяет *domicilio fiscale* как местность, где лицо учреждает основную коммерческую деятельность или деятельность, приносящую доход.

⁸ Данная операция привязки записей, а также подготовка архива AT, описанная в предыдущем абзаце, была проведена коллегами Директората административных и статистических регистров (DCAR).

⁹ Частичное совпадение PS и AT было минимальным, учитывая менее 4 тыс. записей (около 0,2% консолидированного LIFA).

особенности в тех, где численность населения превышает 150 тыс. человек. Метаданные по возвращенным переписным листам являются окончательными только для муниципалитетов с численностью населения менее 20 тыс. человек, где работа по переписи была завершена 31 декабря 2011 г.

27. В таблице 2 представлены некоторые предварительные результаты по охвату и эффективности регистров, задействованных в переписи, разбитые по численности населения районов и муниципалитетов. В муниципалитетах с численностью населения менее 20 тыс. человек около 94% всех зарегистрированных домохозяйств/сожителей заполнили и вернули переписные листы. В более населенных муниципалитетах процент заполнения ниже в связи с тем, что счетчики все еще продолжают поиск зон действия системы регистрации и сбор переписных листов. В группе муниципалитетов с численностью населения менее 20 тыс. человек около 6% всех переписных листов, отправленных по почте, не были возвращены, поскольку домохозяйства не проживали по адресам, по которым были направлены переписные листы (неполный охват LAC). С другой стороны, около 4% всех возвращенных переписных листов были распространены счетчиками незарегистрированным домохозяйствам (неполный охват). Результаты показывают, что поиск неполного охвата осуществлялся преимущественно на основе LIAC. Неутешительно, что как LIFA, так и RNC, сыграли очень незначительную роль в дополнении LAC. Еще предстоит выяснить вызвано ли это быстрым устареванием этих регистров, недостаточной интеграцией LIAC с дополнительными архивами или их низкой степенью использования муниципалитетами во время переписи.

Таблица 2: Процент возврата переписных листов и охвата LAC (%)

Численность населения	Архивы				Избыточный охват LAC	Неполный охват LAC
	LAC	LIAC	LIFA+RNC	Все		
Всего	89,5	62,7	2,8	69,0		
<=20 000 чел.	93,9	80,4	2,8	72,1	6,1	4,1
>20 000 чел.	85,8	49,5	2,9	66,3		

V. Заключение

28. Впервые 15-я перепись населения в Италии проводилась с применением административных регистров. Муниципальным регистрам населения была придана новая роль, к примеру, они были использованы для рассылки по почте около 25 млн. переписных листов. Эти регистры дополнялись другими архивами в целях сокращения потенциальных недостатков и, соответственно, риска того, что значительное число незарегистрированных домохозяйств окажется неучтенными. При том, что перепись еще продолжается, предварительные результаты свидетельствуют в целом об успешности LAC в гарантировании высокой степени охвата. Общий вклад вспомогательных архивов, использованных в целях дополнения LAC, на сегодня не оправдал ожиданий и, возможно, по завершению переписи будет необходимо провести тщательный анализ затрат и выгод для обсуждения их использования в будущем.

