

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

Женева, 30 сентября – 2 октября 2015 года

Пункт 2 б) предварительной повестки дня

Инновации, запланированные для раунда**переписей 2020 года, и результаты тестирования****Комбинированное использование множественных
источников данных в рамках переписи населения****Записка Национального института статистики Италии***Резюме*

Новая стратегия проведения итальянской переписи населения и жилищного фонда предполагает более широкое использование имеющихся административных и статистических источников при ограниченном использовании выборочных обследований в соответствии с предусмотренной планом модернизации Национального института статистики Италии (ИСТАТ) задачей по переходу на интегрированную систему регистров. Ожидается, что уменьшение масштабов, но сохранение постоянного характера сбора данных на местах позволит поступательно совершенствовать методологию и накопить опыт. Кроме того, ожидается, что это позитивно скажется на финансировании; потребности в государственных финансовых ресурсах со временем уменьшатся и будут рассредоточены во времени. Кроме того, непрерывное составление данных позволит намного улучшить и расширить взаимодействие с пользователями.

В настоящем документе рассматриваются некоторые наиболее важные методологические, технические и организационные требования для получения таких результатов. ИСТАТ предлагает наладить более тесное сотрудничество между странами в области методов, связанных с использованием множественных источников, с уделением особого внимания следующим вопросам: i) модели, позволяющие комплексное использование данных из множественных источников; ii) проверки охвата регистров населения для расчета статистических показателей; iii) объединение данных имеющихся обследований; iv) оценка качества данных, получаемых путем интеграции различных источников.



I. Введение

1. Комбинирование имеющихся источников данных, в том числе данных обследований, переписей населения и административных регистров, в следующем раунде переписей станет, скорее, необходимостью, нежели просто одним из возможных вариантов. Перепись населения, проводимая раз в десятилетие исключительно методом сплошной регистрации, является слишком сложным и уже исчерпавшим свой потенциал мероприятием, а такие заинтересованные лица, как политики, администраторы и исследователи, нуждаются в регулярно обновляемых и согласованных демографических, социальных и экономических данных с географической привязкой на самом высоком уровне географической детализации. Значительные расходы на проведение традиционных переписей и связанное с ними бремя оперативной работы создают стимулы для более эффективного использования административных источников данных.

2. Если последний раунд переписей характеризовался многообразием методов сбора данных, то характерной особенностью следующего раунда будет, скорее всего, задействование множественных административных и статистических источников, и именно в русле этих вопросов следует развивать конструктивное сотрудничество между странами.

3. Лица постоянно оставляют информацию о себе в административных регистрах, которые потенциально содержат гигантский объем данных с географической привязкой, доступных для использования в статистических целях. Такие данные, однако, могут не обновляться и страдать погрешностями охвата из-за несогласованности классификаций и определений, что может ухудшать их пригодность для использования.

4. В пересмотренном варианте Принципов и рекомендаций Организации Объединенных Наций альтернативные методологии проведения переписей представлены в виде матрицы, в строках которой описываются методы сбора данных путем сплошной регистрации, а в столбцах указываются способы использования административных регистров и/или регистров населения в качестве источников данных. В столбцах матрицы приводится множество типов регистров: административные регистры, которые создаются и ведутся в основном в административных целях за пределами национальных статистических органов, статистические регистры и базовые регистры.

5. В рамках новой итальянской переписи – *непрерывной переписи* – основная задача состоит в том, чтобы перейти к более интенсивному использованию административных и статистических источников при ограниченном дополнительном использовании специальных выборочных обследований, с тем чтобы обеспечить расчет результатов переписи на ежегодной основе. Целью этого является переход на новый вид переписи, при которой с помощью статистических методов и новых технологий результаты переписи будут рассчитываться ежегодно на основе последовательности тщательно спланированных операций по интеграции данных, начиная с имеющихся административных и статистических источников. Мы занимаемся тщательным изучением методов, которые позволяют распространять методику использования одного источника, и концептуальных вопросов, связанных с одновременным использованием различных источников.

II. Нововведения в области сбора данных с использованием множественных методов в рамках переписи 2011 года

6. Для проведения переписи 2011 года по сравнению с предыдущими переписями был использован принципиально отличный подход. Для рассылки вопросников семьям использовались административные регистры населения (Anagrafi). Респонденты могли по своему усмотрению воспользоваться одной из нескольких возможностей для заполнения и возврата вопросника:

- по Интернету с использованием пароля, указанного в вопроснике;
- в любом почтовом отделении Италии;
- в одном из муниципальных центров переписи, где можно было получить помощь специалиста;
- непосредственно муниципальному переписчику.

7. Для измерения некоторых представляющих интерес переменных использовались методы выборки. Это обеспечивалось путем использования вопросников двух видов – сокращенного, предназначенного для двух третей семей в более крупных муниципалитетах, и полного, предназначенного для остальных семей.

8. Была разработана интернет-система для управления предоставлением ответов различными способами и оказания поддержки счетчикам по случаям непредоставления ответов. Эта система была разработана для автоматизации обработки информации и позволила контролировать статус каждого индивидуального вопросника практически в реальном времени. Кроме того, она позволяла составлять отчеты о ходе переписи, распределять переписные участки среди счетчиков и контролировать их работу, адресно выявлять лиц, не представивших ответы, и незарегистрированных лиц, сопоставлять результаты переписи с учетными муниципальными документами и вести сравнительный учет найденных и не найденных лиц. Наличие постоянно обновляемой информации о статусе каждого вопросника позволяло направлять переписчиков только в те домашние хозяйства, в которые вопросник был направлен, но из которых он еще не был получен, и тем самым систематически выявлять случаи непредоставления ответов.

9. В самых крупных муниципалитетах имелся дополнительный вспомогательный список, основанный на результатах предшествующей переписи обследования адресов и содержащий информацию о потенциально заселенных жилищных единицах, по которым в муниципальных учетных документах не было указано соответствующих им семей. Использование других вспомогательных списков из центральных и местных административных источников, содержащих информацию о наличии лиц, не зарегистрированных в муниципальных учетных документах, позволило вести поиск таких лиц целенаправленным и систематическим образом.

10. Для получения сигналов о лицах, не внесенных в муниципальные учетные документы, и обеспечения доступности пространственных данных на уровне отдельных единиц, будь то физических или юридических лиц, административные и статистические источники были интегрированы в Единую систему микроданных (SIM). Данные из различных источников, включая коммерческие, налоговые и образовательные регистры, регистры занятости и другие соответствующие регистры, были увязаны с помощью индивидуальных уникальных кодов. Одним из основных источников, использовавшихся в ходе этого процесса, был перечень видов на жительство, выдаваемых негражданам Европейского союза. В SIM были включены, в частности, такие признаки, как: характеристики домашнего хозяйства, место обычного жительства, адрес места работы, адрес места учебы

(школы, колледжа или университета), статус занятости, образовательные характеристики, жилье и условия проживания и т.д. Для решения проблемы потенциально неполного охвата населения использовались данные из таких источников, как данные налоговых органов или архивы сведений о видах на жительство иностранных граждан, на основании которых были составлены перечни лиц, не включенных в муниципальные регистры населения, но потенциально проживающих в каждом из муниципалитетов.

11. Несмотря на применение в ходе переписи 2011 года значительных нововведений, стабильного и устойчивого баланса между издержками и выгодами переписи достичь пока не удалось. Напротив, издержки остаются высокими и избыточно сконцентрированными во времени, а использование административных данных в их нынешнем виде пока не позволяет удовлетворять спрос на данные, аналогичные данным переписи. Кроме того, данные переписи быстро устаревают, а получение высокодетализированных географических данных возможно по-прежнему лишь раз в десять лет. Этими причинами объясняется необходимость разработки иного подхода.

III. После 2011 года

12. На основе обобщения опыта предыдущего раунда переписей населения и жилищного фонда мы в настоящее время планируем использовать новые методологии для получения переписных данных. Первостепенное значение по-прежнему придается ключевому принципу получения подробной статистики на наименьшем географическом уровне. В целях разработки подробной статистики населения и жилищного фонда по малым районам изучается возможность использования регистров, а также применения с этой же целью методологии непрерывных обследований.

13. Внедренные в рамках раунда 2011 года приемы, методы и организационные решения в настоящее время пересматриваются и объединяются с новыми приемами, методами и решениями в инновационную систему, в составе которой они больше отвечают стратегическим задачам нового уровня.

14. Существуют многочисленные аргументы в пользу разработки такого альтернативного подхода. В их числе можно назвать: i) необходимость более частой и своевременной подготовки статистических данных; ii) ограниченность бюджета переписи; iii) нежелание населения участвовать в переписи; iv) расширение технических возможностей для работы с источниками данных.

15. Информация об отдельных лицах и домашних хозяйствах собирается из имеющихся административных или статистических источников, а именно регистров различного рода, из которых первостепенное значение имеют: регистры лиц, домашних хозяйств и жилищ, а также имеющиеся данные обследований, увязанные на уровне лиц. Специальные выборочные обследования используются для получения информации по признакам программы переписи, которые не представлены в административных источниках, либо для корректировки низкокачественных регистровых данных.

16. Статистические методы будут использоваться для решения двух основных задач переписи населения, а именно:

- регистрации постоянных жителей для составления ключевых данных о демографической структуре населения и домашних хозяйств (задача C);
- получения социально-экономических данных переписи (задача D).

17. Что касается задачи С, т.е. расчета ежегодных оценок численности населения в разбивке по демографической структуре (пол, возраст, семейное положение, гражданство), то наиболее серьезные проблемы в ближайшие годы будут связаны с подсчетом пожилых лиц и мигрантов.

18. Что касается задачи D, то в целях выполнения национальных и международных требований к оценке социально-экономических данных о домашних хозяйствах и отдельных лицах основная цель будет заключаться в сборе с помощью обследований данных только по тем переменным, которые подлежат измерению в соответствии с правилами ЕС и не могут быть оценены путем комбинирования информации, имеющейся в регистрах населения, административных источниках и данных статистических обследований. В настоящее время мы занимаемся изучением возможности замещения переписных данных социально-экономическими данными из административных источников. Цель заключается в том, чтобы заменить данные, традиционно собираемые в ходе переписи, оценками, полученными на основе имеющейся информации: административных источников и сводных данных уже существующих выборочных обследований. Для этого необходимо активизировать текущие исследования по изучению местных и центральных административных источников. Мы будем удовлетворять потребности пользователей, чаще представляя обновленные данные и избегая тем самым снижения их точности на протяжении десятилетия. Следует отметить, что лица и домохозяйства в первую очередь выиграют от использования статистическими органами уже имеющейся информации, которую иначе пришлось бы собирать от них повторно при помощи дорогостоящих и избыточных обследований. Кроме того, с помощью административных данных и больших данных можно увеличить объем данных по малым районам и разработать новые показатели для целей политики. Наконец, более активное использование административных данных расширит возможности государственных органов по сбору фактологических данных, на основе которых можно будет оценивать эффективность их программ и политики. Будут предприниматься усилия для определения областей, которые можно более эффективно охватить при помощи новой информации.

IV. Проекты, реализуемые ИСТАТ

19. В Плане модернизации ИСТАТ перед итальянской статистической системой поставлена задача перейти на интегрированную систему статистики, опирающуюся в первую очередь на новые статистические методы и новые технологии, позволяющие более эффективно задействовать административные и статистические источники данных в интересах сведения к минимуму нагрузки на респондентов и затрат. Основным компонентом такой интегрированной системы, охватывающей экономическую, экологическую и социальную сферы, являются интегрированные и согласованные между собой регистры и основы выборки коммерческих предприятий и населения. Будет разработана согласованная система выборок статистических единиц для различных обследований и составлены специальные вопросники для целевых групп населения. Уникальный идентификационный код, используемый для интеграции записей в SIM, будет также применяться в системе национального статистического управления для упрощения комбинирования данных из различных источников, их сопоставления во времени, а также поощрения многоцелевого использования данных. Для корректировки данных применяются статистические методы. Данные от лиц, домашних хозяйств или предприятий следует собирать только в тех случаях, когда их невозможно получить из базовых регистров, административных и имеющихся статистических источников.

20. Ключевая роль отводится регистру ANNCSU (итальянское сокращение, обозначающее национальный регистр улиц и адресов) с его системой географиче-

ской привязки улиц и номеров домов. Такая база географических данных действительно крайне важна для поиска единиц в информационной системе и служит незаменимым инструментом повышения качества геокодирования данных из административных источников. Номер каждого дома будет геокодирован с привязкой к установленным ИСТАТ переписным участкам.

21. Основные компоненты программы Италии в области цифровых технологий установлены законом, принятым в конце 2012 года. Помимо *непрерывной переписи* и ANNCSSU, она предусматривает ведение уникального национального регистра населения ANPR (национальный архив постоянного населения), в который будут включены все муниципальные регистры. Для составления данных по малым районам на основе регистра населения решающее значение имеет интеграция трех указанных компонентов.

22. Чтобы обеспечивать пригодность административных источников для целей статистики, крайне важно добиться их соответствия требованиям качества на основе стратегии непрерывного контроля качества и корректировки данных. ARCOLAIO является проектом, который был начат в ИСТАТ для обеспечения контроля качества при использовании административных источников в целях статистики. Кроме того, мы тщательно изучаем методы, которые можно было бы использовать для решения этих задач.

23. В соответствии с Европейским регламентом о переписях населения и жилищного фонда (Европейский парламент, 2008 год) целевой совокупностью переписи являются постоянные жители, определяемые как лица, которые проживали или намерены проживать в течение более 12 месяцев в месте своего обычного жительства. Поскольку намерения не регистрируются в административных источниках, а идентификация жителей, которые проживали в одном и том же месте на протяжении более года, весьма затруднительна, данное определение трудноприменимо в рамках нового подхода. Кроме того, "реальное" постоянное население может отличаться от зарегистрированного населения по причине не отражаемых в документах переездов или фиктивных регистраций по налоговым сообщениям.

24. ARCHIMEDE (архив экономических и социально-демографических микроданных) представляет собой инновационную инфраструктуру, которая будет производить микроданные, получаемые административными источниками и в ходе статистических обследований, после их проверки на качество и обработки. Геокодированные данные с привязкой к переписным участкам будут публиковаться с соблюдением требований конфиденциальности эффективным и транспарентным образом. Выходные материалы ARCHIMEDE будут определяться строго в соответствии с потребностями пользователей, чтобы обеспечить снабжение государственных органов, исследователей и пользователей наборами данных общего назначения, имеющих привязку к статистическим единицам базового уровня и географическим единицам (вплоть до переписных участков).

25. Что касается методологии, то мы занимаемся изучением прогнозного потенциала моделей, в которых обычное место жительства служит зависимой переменной, а данные из множественных административных и статистических источников – независимыми переменными. В частности, мы рассматриваем возможность применения моделей латентных классов (Biemer Paul P., 2011, p. 258) и нескольких генерализаций методов двойного охвата в целях работы с множественными списками. Добавление в систему интегрированных микроданных таких новых источников, как списки договоров коммунального обслуживания от газовых и электроэнергетических компаний и списки пользователей местных услуг, может значительно повысить прогнозный потенциал этих моделей.

26. Пилотная серия обследований в рамках *непрерывной переписи*, которая была начата в марте 2015 года, охватывает 150 муниципалитетов и 160 000 домашних хозяйств. Цель заключается в тестировании прогнозного потенциала вышеупомянутых статистических моделей и нескольких возможных вариантов организации переписей, ориентированных в первую очередь на новые технологии ввода данных и архитектуры ИТ, с уделением особого внимания:

- архитектуре ИТ и разработке интегрированной системы управления сбором данных множественными методами, позволяющей вести работу с использованием самых разнообразных инструментов (планшетов, мобильных устройств, ПК и т.д.), с подключением к Сети и без него, а также независимо от используемых аппаратных средств и операционной системы;
- прекращению использования бумажных вопросников для сбора данных в ходе обследований на местах в целях внедрения различных организационных решений, учитывающих местные тенденции возврата переписных листов и различия в использовании ИКТ;
- выверке названий улиц и номеров домов в регистре ANNCUSU.

V. Выводы

27. До настоящего времени, в том числе в 2011 году, перепись проводилась в Италии каждые 10 лет. Новая итальянская стратегия проведения переписей населения и жилищного фонда предполагает более широкое применение имеющихся административных и статистических источников при ограниченном использовании выборочных обследований.

28. Непрерывные расчеты позволят значительно повысить эффективность деятельности на местах и принесут множество выгод в плане качества. Непрерывная работа на местном уровне позволит удержать квалифицированные кадры и с течением времени повысить их квалификацию; уменьшение масштабов, но сохранение постоянного характера работы на местах, как ожидается, позволит поступательно совершенствовать методологию и накопить опыт.

29. Кроме того, это, как ожидается, позитивно скажется на финансировании; потребности в государственных финансовых ресурсах со временем уменьшатся, а постоянный характер работы может сделать контракты на обслуживание более привлекательными и, возможно, менее дорогостоящими, чем при проведении "разовой" операции. Кроме того, непрерывное составление данных позволит намного улучшить и расширить взаимодействие с пользователями.

30. В настоящем документе обсуждаются некоторые наиболее важные методологические, технические и организационные требования для получения таких результатов.

31. Мы предлагаем наладить более тесное сотрудничество между странами в области методов, связанных с использованием множественных источников, с уделением особого внимания следующим вопросам:

- модели, позволяющие комплексно использовать данные из множественных источников;
- проверки охвата регистров населения для расчета статистических показателей;
- объединение данных имеющихся обследований;
- оценка качества данных, получаемых путем интеграции различных источников.

Справочные материалы

- Biemer Paul P., 2011 Latent class analysis of survey errors, Wiley Series in Survey Methodologies
- Baffour, B., J.J. Brown, P.W.F Smith, 2013, An investigation of triple system estimators in censuses. *Statistical Journal of the International Association for Official Statistics*, 29, 53–68
- Bakker, B.F.M., Rooijen, J. van & Toor, L. van (2014). The system of social statistical datasets of Statistics Netherlands: an integral approach to the production of register-based social statistics. *Statistical journal of the United Nations ECE*, 30(4), 411–424
- European Parliament, 2008, Regulation (EC) No 763/2008 of the European Parliament and of the council of 9 July 2008 on population and housing censuses, In: *Official Journal of the European Union*, 13.8.2008, pp. L 218/14-L 218/20
- Schulte Nordholt, E. (2005), The Dutch Virtual Census 2001: a new approach by combining different Sources, *Statistical Journal of the United Nations Economic Commission for Europe* 22, 25–37
- ЕЭК ООН (2007), Статистика на основе регистров в Северо-европейских странах: обзор передовых методик с уделением основного внимания статистике населения и социальной статистике, издание Организации Объединенных Наций, ISBN 978-92-1-116963-8
- Wallgren, A. and Wallgren, B. (2007), Register-based statistics – administrative data for statistical purposes, John Wiley and Sons, Chichester
- Zhang, L-C. (2012), Topics of statistical theory for register-based statistics and data integration, *Statistica Neerlandica*