



Европейская экономическая комиссия

Конференция европейских статистиков

Группа экспертов по переписям населения и жилищного фонда

Восемнадцатое совещание

Женева, 28–30 сентября 2016 года

Пункт 6 предварительной повестки дня

Инновации в методологии и технологии переписи и результаты тестирования

Создание базового регистра территориальных единиц в целях повышения качества и географической привязки данных переписи

Записка Национального института статистики Италии¹

Резюме

На статистические управления оказывается все большее давление, с тем чтобы они готовили высококачественные данные с географической привязкой. Качество интеграции данных из многих различных источников может повыситься благодаря использованию географической привязки и пространственного анализа. В настоящем документе описаны преимущества использования строящейся на географической привязке комплексной системы базового регистра для расчета результатов переписи.

¹ Подготовлено Фабио Кресенци.



I. Введение

1. Значительные расходы на проведение традиционных переписей и связанное с ними бремя оперативной работы создают стимулы для объединения всех имеющихся источников данных, в том числе обследований, статистических и административных регистров, для расчета результатов переписи. Перепись населения, проводимая раз в десятилетие исключительно методом сплошной регистрации, является слишком сложным и уже исчерпавшим свой потенциал мероприятием. Кроме того, такие заинтересованные лица, как политики, администраторы и исследователи, нуждаются в регулярно обновляемых и согласованных демографических, социальных и экономических данных с географической привязкой при максимальной географической детализации.

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

3. Физические 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, при этом они страдают погрешностями охвата из-за несогласованности классификаций и определений, что может ухудшать их пригодность для использования.

4. В рамках новой итальянской переписи – *censimento permanente* (непрерывная перепись) – основная задача состоит в том, чтобы перейти к более интенсивному использованию административных и статистических источников при ограниченном дополнительном использовании специальных выборочных обследований для обеспечения расчета результатов переписи на ежегодной основе. Целью этого является переход на новый вид переписи, при котором 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. ИСТАТ тщательно изучает методы, которые позволяют распространять методику использования одного источника, и концептуальные вопросы, связанные с одновременным использованием различных источников.

5. Первостепенное значение по-прежнему придается ключевому принципу получения подробной статистики при максимальной географической детализации. 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.

6. В настоящее время эти приемы, методы и организационные решения пересматриваются в рамках новой структуры, в составе которой они больше отвечают стратегическим целям, с учетом следующего: i) необходимости более частой и своевременной подготовки статистических данных; ii) необходимости снизить расходы, связанные с переписью; iii) нежелания населения участвовать в переписи; iv) расширения технических возможностей для работы с источниками данных.

7. В Плане модернизации ИСТАТ поставлена задача перейти на интегрированную систему статистики, опирающуюся в первую очередь на статистические методы и новые технологии, которые позволяют более эффективно задействовать административные и статистические источники данных в интересах сведения к минимуму нагрузки на респондентов и затрат. Основными компонентами такой интегрированной системы, охватывающей экономическую, экологическую, демографическую и социальную сферы, являются интегрированные и согласованные между собой базовые регистры и основы выборки экономических единиц и физических лиц. Необходимо разработать согласованную систему выборки статистических единиц для различных обследований и составить специальные вопросники для целевых групп населения.
8. Экологические и территориальные данные находятся в центре этих изменений, поскольку необходимо удовлетворять рост спроса на данные, имеющие географическую привязку, со стороны пользователей на международном, национальном и местном уровнях. Спрос на территориальные данные учреждений и ассоциаций требует, с одной стороны, увеличения объема данных для инфраструктуры пространственных данных и, с другой стороны, повышения актуальности, своевременности и качества выпускаемых данных.
9. В Италии статья 45 президентского указа от 30 мая 1989 года (№ 223) гласит, что каждый муниципалитет составляет и обновляет перечень улиц, как это было рекомендовано Национальным институтом статистики. Кроме того, законом от 17 декабря 2012 года (№ 221) предусматривается учреждение национального архива улиц и адресов (ANNCSU), который создается и обновляется ИСТАТ и Кадастровым управлением Национального налогового управления.
10. Проект ANNCSU, разработанный для удовлетворения потребностей справочной базы данных для всей страны, содержит информацию об улицах и номерах домов в цифровом формате, геокодированных по переписным участкам геопространственных баз данных.
11. В рамках процесса обработки данных пятнадцатой переписи населения были установлены связи и осуществлена перекрестная проверка списков адресов, содержащихся в данных, направленных ИСТАТ муниципалитетами по поводу проведения переписи, при этом было отмечено несоответствие разных массивов данных. По согласованию с Кадастровым управлением Национального налогового управления в январе 2014 года отделениям муниципалитетов было предложено проверить эти несоответствия и приложить усилия, в случае необходимости, к тому, чтобы исправить, интегрировать и подтвердить данные в рамках деятельности, которая в настоящий момент завершена на 90%.
12. Этот архив имеет важнейшее значение для геопривязки данных в стандартной форме в рамках всей системы на основе сведений, содержащихся в административных файлах и статистических обследованиях.
13. Кроме того, Национальное агентство по цифровым базам данных считает этот архив основной базой данных в связи с тем большим значением, которое он имеет для перехода к Национальному регистру постоянного населения (ANPR), а также для многих других целей, представляющих общественный интерес.

14. В рамках этого проекта ИСТАТ: а) определяет источники, правила и методологии периодического обновления ANNCSU; б) на основе имеющихся источников производит надлежащую оценку координат точек при геопривязке для каждого номера дома; в) присваивает уникальный код улицам и номерам улиц; г) осуществляет географическое проектирование улиц и номеров улиц в географической системе Института; д) обеспечивает оперативную совместимость с другими национальными и местными базами данных; е) готовит документацию об источниках, методах геопривязки, используемых номенклатурах и классификациях.

15. В частности, наличие координат точек будет иметь большое значение для пространственного анализа и разработки статистических данных на регулярной сетке, как того требует стратегия модернизации европейской статистики.

16. Управление базовым статистическим регистром географических/территориальных единиц и соответствующими спутниковыми регистрами будет включать сопоставление различных компонентов и определение соответствующих единиц, в том числе адресов, зданий и географических единиц.

17. С тем чтобы обеспечить практическое управление пространственными элементами базовой системы регистров будут разрабатываться специальные проекты со следующими целями:

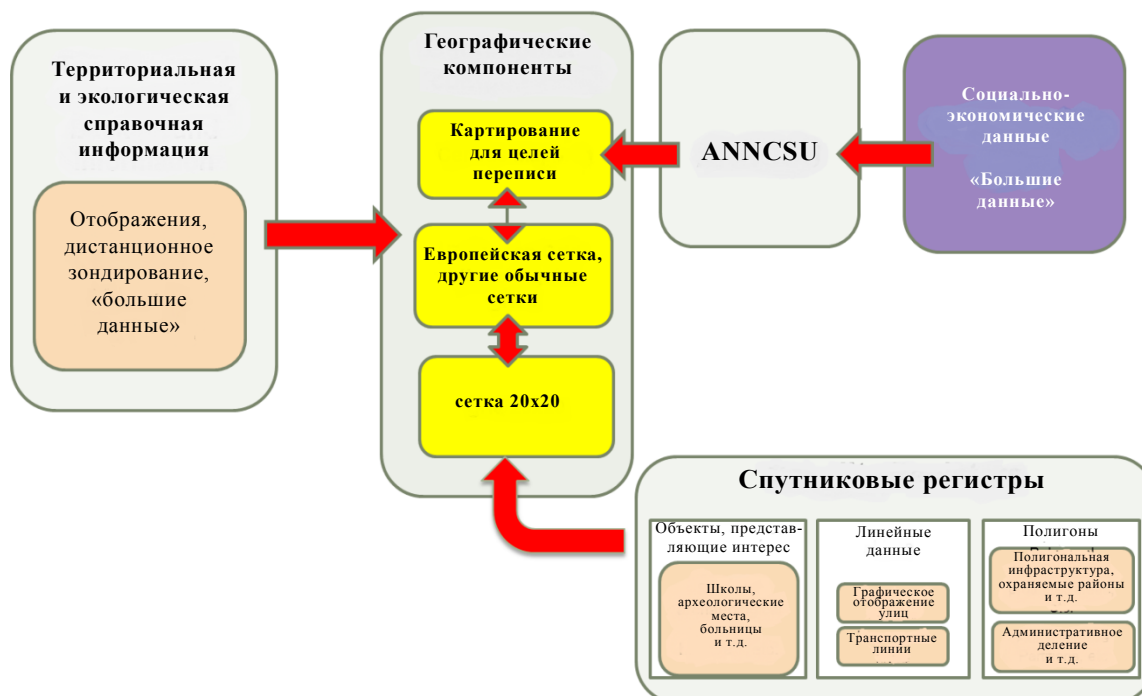
- определить стратегию для обеспечения обновления географических единиц в базовых регистрах, в том числе в связи с потребностями постоянных переписей;
- определить качественные показатели географического расположения единиц системы базовых регистров;
- расширить использование ГИС Института для статистических целей.

18. Что касается показателей качества, то проводятся мероприятия как концептуального, так и экспериментального характера по тематике оценки качества данных геокодирования. Это будет сделано путем оценки и принятия необходимых мер в отношении: а) пробелов в охвате географических районов, б) семантического качества и позиционной точности данных, в) качества временных обновлений данных.

19. Цель заключается в том, чтобы создать возможности для геопривязки данных в стандартной форме в рамках всех компонентов системы, включая сведения, содержащиеся в административных файлах и статистических обследованиях.

Рис. 1

Базовый регистр географических единиц; взаимосвязь между данными и географической принадлежностью



II. Системы местной рабочей силы

20. Географическая структура систем местной рабочей силы (LLS), формирующаяся в результате ежедневных поездок до места работы, с хорошей степенью приближения описывает функционирование городских систем в дневное время, т.е. выявляет те места, в которых сосредоточено большинство действий и ежедневных перемещений в том, что касается физических лиц и хозяйствующих субъектов.

21. В декабре 2014 года ИСТАТ выпустил четвертое издание LLS, подготовленное с использованием данных о потоках перемещения «место жительства/место работы» (ежедневные поездки), сведения о которых были собраны в ходе переписей населения. Это дало возможность внести принципиальные изменения в территориальные вопросы.

22. Использование LLS для целей новой «регионализации» территории Италии позволило выйти за рамки классификации по таким традиционным признакам, как секторы, размеры и административные единицы, которые более не являются достаточными для отражения и описания социально-экономической системы Италии, отличающейся значительной разнородностью.

23. LLS, именно в силу того, что они являются «спонтанными» областями и зависят от собранных данных, при этом они (в основном) не зависят от выбора и действий отдельных лиц, позволяют пользователям наблюдать «реальную страну» в ее разнообразии и особенностях, более точно определяя периметры отношений, сети, обмены и потоки, которые характеризуют те или иные места.

24. Таким образом, данная географическая структура позволяет описывать систему, зачастую существенно отличную от той, которую дает традиционное деление территории. ИНСТАТ будет и впредь вкладывать средства в эту функциональную географию посредством дальнейшего анализа и подготовки данных.

25. Можно будет изменить методы построения систем местной рабочей силы в связи с началом постоянной переписи, начальным внедрением системы регистров и матриц ежедневных поездок на административной основе. Воспроизведение LLS может проводиться чаще, чем раз в 10 лет, при этом могут быть созданы новые статистические продукты. В целом это придаст новый импульс тематической классификации территории с особым акцентом на степень урбанизации и землепользования.



26. Переход Института к системе, основанной на регистрах и использовании новых аналитических инструментов (средств для оценки и прогнозирования, моделирования и оценки последствий государственной политики) и на связи между данными, будет способствовать повышению качества и согласованности в плане единого восприятия данных пространственных и экологических обследований с данными обследований других секторов. Особое внимание необходимо будет уделить следующему:

- придать новый импульс разработке как методологии, так и организационной структуры обследований состояния ресурсов окружающей среды и управления ими, обследований культурных и земельных параметров, обследований выбора населения и предприятий в плане землепользования и использования источников энергии и ресурсов окружающей среды; обследований туризма, транспорта и доступа к услугам, мобильности и перемещения физических лиц и товаров по территории страны;
- повысить согласованность и последовательность в том, что касается обследований структурных и экономических характеристик фермерских хозяйств и предприятий, участвующих в сельскохозяйственной деятельности, а также их экономических и управленческих характеристик;
- повысить согласованность и последовательность в том, что касается статистических данных городских систем, параметров «разумных городов» и жилищных условий в городах и внутренних районах;
- подготовить всеобъемлющую информационную основу, содержащую последовательные сведения о зданиях и жилищных единицах.

III. Выводы

27. С тем чтобы дать заинтересованным сторонам возможность принимать более обоснованные решения, необходимо будет укрепить производство своевременной территориальной статистической информации, пригодной для территориального и секторального планирования и предварительной и последующей оценки государственной политики. Статистические управления министерств, регионов, провинций и муниципалитетов должны быть обеспечены как можно более свежей и подробной информацией, отражающей основные переменные на муниципальном и более низком уровнях. Следует предоставить более широкий доступ к статистической информации для целей научных исследований, обеспечивая при этом конфиденциальность респондентов. Важное значение для понимания динамики явлений будет также иметь доступ к динамическим рядам данных.

28. Быстрый переход к информационному обществу, которое использует все больше и больше данных, производимых в том числе в результате растущей активности в Интернете, ведет к перемещению разнообразных данных из различных источников с такой скоростью и в таких объемах, которые всего лишь несколько лет назад трудно было себе представить. В этой связи руководители получают возможность проводить своевременные и эффективные действия по поиску новых источников, которые предстоит рассмотреть на предмет их приобретения и новых методов использования.

29. Кроме того, качество документации о характеристиках продукции и данных о качестве методов, используемых для их производства, приобретает все большее значение в связи с налаживанием диалога со всеми сторонами об их социальной значимости и их оперативным привлечением для переориентации производства с учетом потребностей пользователей.

30. Для успешного решения этих проблем и предложения надлежащих статистических продуктов и услуг мы должны использовать возможности, возникающие на передних рубежах технологического развития, стремиться наращивать инвестиции в передовые технологии и в человеческий капитал, с тем чтобы сократить бремя, повысить эффективность за счет повторного использования данных, использовать административные данные, задействовать новые источники данных («большие данные») и передовые статистические технологии получения и обработки многомерных данных, которые находятся на переднем крае в плане выявления, описания, визуализации и распространения данных. Процесс руководства должен быть структурирован на основе отраслевой модели, которая способствует повышению эффективности, тиражированию, повышению качества продукции, снижению издержек производства путем применения наилучших имеющихся технологий и методологий при соблюдении максимальных требований к своевременности, беспристрастности и объективности.

31. Необходимо будет предоставлять данные в открытых и пригодных для повторного использования форматах (связанные открытые данные) и услуги для приведения данных в соответствие с потребностями пользователя, с тем чтобы поставить его в центр этого процесса, осуществляя интеграцию производства данных и производства услуг с высоким содержанием информации. В перспективе в соответствии с европейской концепцией статистики эта система должна быть постепенно адаптирована путем включения в нее метаданных, которые будут служить ориентиром для стандартизации процессов и облегчат отслеживание статистических данных и доступ к ним. Она будет способствовать облегчению и расширению доступа исследователей к файлам макро- и микроданных, подпадающих под действие норм, которые регулируют неприкосновенность частной жизни, путем улучшения использования инструментов оперативной совместимости и обеспечения удаленного доступа в соответствии с передовым международным опытом.

32. Вся эта деятельность должна стать неотъемлемой составной частью процессов, направленных на достижение целей Повестки дня в области цифровых технологий на национальном и международном уровнях, при этом будут продолжены усилия по предоставлению новых технологических решений, методологического и статистического анализа для закрепления достигнутых результатов и дальнейшего расширения границ инноваций в партнерстве с другими итальянскими и международными субъектами. Предполагается также развивать деятельность в рамках европейских проектов и инициатив, уже осуществляемых Институтом, которые предусматривают совместное планирование и разработку некоторыми крупными национальными статистическими органами общих решений, пригодных для повторного использования, стандартизацию и перевод на отраслевую основу процессов в рамках статистических организаций.

33. В этом контексте предусмотрены следующие три основных компонента:

а) ключевая роль отводится регистру ANNCSU (национальный регистр улиц и адресов) с его системой географической привязки улиц и номеров домов. Такая база географических данных действительно крайне важна для поиска единиц в информационной системе. Она служит незаменимым инструментом повышения качества геокодирования данных из административных источ-

ников. Номер каждого дома будет геокодирован с привязкой к определенным ИСТАТ переписным участкам;

б) основные компоненты программы Италии в области цифровых технологий установлены законом, принятым в 2012 году. Помимо непрерывной переписи и ANNCSU, 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;

с) с целью обеспечить пригодность административных источников для статистики крайне важно добиться их соответствия требованиям качества на основе стратегии непрерывного контроля качества и корректировки данных. ИСТАТ в рамках своего плана модернизации приступил к осуществлению проектов, направленных на контроль качества статистического использования административных источников. Кроме того, ИСТАТ тщательно изучает методы, которые можно было бы использовать для улучшения качества географических компонентов.

Справочная документация

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.

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.

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*.