



ОРГАНИЗАЦИЯ ОБЪЕДИНЕННЫХ НАЦИЙ

ECE/CES/2012/40

ЭКОНОМИЧЕСКИЙ И СОЦИАЛЬНЫЙ СОВЕТ

27 апреля 2012

Распространение: Общее

Оригинал: английский

Европейская экономическая комиссия ООН

Конференция европейских статистиков

60-е пленарное заседание

Париж, 6-8 июня 2012

Пункт 3(b) повестки дня

Будущее переписей и их роль в национальных статистических системах

К новой парадигме непрерывного использования регистров и геокодированных данных

Записка Национального института статистики Италии (ИСТАТ)

Резюме

В данном документе описаны новые подходы и методы, применяемые Национальным институтом статистики Италии в ответ на необходимость предоставлять более оперативные данные переписи с более высокой периодичностью. Для этого необходимо пересмотреть всю стратегию проведения переписи. В документе описано, как были изменены информационные инфраструктуры, процедуры, методы и организационные решения для проведения переписи в Италии, и как новые методы были объединены с текущим рабочим процессом. Новая стратегия состоит в переходе от одного источника данных к смешанным способам сбора данных, от одномоментной переписи к непрерывной переписи, к более широкому использованию пространственных данных и согласованных обследований.

I. Почему переписи должны измениться?

1. Пользователям статистических данных требуются более оперативные и частые данные переписи. Участники политической и общественной жизни, такие как политики, администраторы, исследователи и предприниматели заинтересованы в получении актуализированных и гармонизированных данных демографической, социальной и экономической статистики, которые, по возможности, имели бы географическую привязку с высокой степенью разрешения. Высокая стоимость традиционных переписей и нагрузка по их проведению заставляют двигаться по пути эффективного использования административных источников и растягивания полевой работы по проведению переписей на несколько лет. Нагрузка по проведению переписей влияет, в основном, на качество данных, собираемых в крупных муниципалитетах, которым в течение короткого времени приходится нанимать большое количество трудовых ресурсов – переписчиков и дополнительных сотрудников – которые должны быть обучены и заняты в полевых работах. Нагрузка на респондентов в традиционной переписи также, вероятно, негативно влияет на качество данных.

2. Принимая во внимание то, что почти ежедневно человек оставляет какой-то след в

административных информационных системах, существует огромный потенциал для использования пространственных данных для статистического анализа. Однако эти данные обычно не обновляются, и им могут быть свойственны ошибки охвата, возникающие в связи с использованием негармонизированных классификаций и определений, что может поставить под угрозу возможность из использования.

3. Следовательно, правильный выбор будет состоять в движении к новой парадигме переписи, где будет очень важно объединить данные из множества административных источников и выборочных обследований, специально предназначенных для непрерывного пополнения и обновления этих источников.

4. Для того чтобы произвести эти перемены, необходимо существенно изменить всю стратегию проведения переписи. В связи с этим в настоящее время происходит пересмотр информационной инфраструктуры, методов и организационных решений, использованных при проведении последней переписи в Италии, и объединение их с новыми подходами в рамках инновационной структуры, которая лучше отражает новые стратегические цели.

5. Ниже вкратце изложены основные моменты новой стратегии:

(а) От одного способа сбора данных - к нескольким. Переход от традиционной переписи домохозяйств к сочетанию разных видов сбора данных с использованием многих источников необходимо для сокращения расходов ИСТАТ и снижения нагрузки на респондентов. Кроме того, это действительно может улучшить качество и оперативность данных;

(b) От проведения переписи каждые десять лет к непрерывным переписям. Для того чтобы административные ресурсы стали полезными для целей статистики, мероприятия должны быть распределены во времени, а не проводиться только один раз в десять лет. Поэтому важно принять стратегию проведения постоянного контроля качества на основе выборочных обследований, целью которых является контроль охвата регистра и полноты включенных в него записей;

(с) Компромисс между пространственным разрешением и детализацией. Традиционные переписи предоставляют данные для малых областей с большой степенью пространственной детализации, но эти данные формируются только один раз в десять лет. При новом подходе можно будет обеспечить разные стандарты качества для разных совокупностей, пространственных разрешений и гиперкубов данных;

(d) Использование междисциплинарной информации. Наличие интегрированных систем административных источников данных об отдельных лицах, домохозяйствах и хозяйствующих субъектах может помочь преодолеть обычную дихотомию между социальной и экономической статистикой, позволяя разрабатывать скоординированные обследования по социальной и экономической тематике;

6. Поскольку интеграция основных источников административных данных в Италии значительно усилилась в последние годы и на концептуальном, и на физическом уровне, пришло время для выполнения среднесрочной цели по производству интегрированных актуализированных пространственных данных о домохозяйствах, отдельных лицах и хозяйствующих единицах. В то же время должен быть осуществлен переход к «непрерывной» переписи в целях сдерживания роста расходов, снижения нагрузки на респондентов, повышения оперативности для пользователей и улучшения планирования. Конечно, предпосылкой для настоящего взаимодействия государственных информационных систем является гармонизация методов и форматов данных для осуществления обмена данными между различными административными органами.

II. Опыт переписи 2011 года

7. Стоит отметить основные нововведения переписи населения и жилого фонда 2011 года. ИСТАТ разработал национальный регистр улиц и адресов (ANSC) при сотрудничестве с “Agenzia del Territorio” - итальянским кадастром. В муниципальных образованиях с населением более 20 000 человек данные регистра ANSC были проверены и обновлены в ходе полевого контрольного обследования (RNC) за шесть месяцев до начала переписи, а списки адресов в муниципальных образованиях меньшего размера проверялись в ходе переписи.

8. Списки домохозяйств для направления вопросников были составлены на основе учетных

документов муниципалитетов, обновленных по 31 декабря 2010 года; были проведены нормализация, распознавание и геокодирование адресов. Вопросники были персонализированы благодаря указанию фамилии и имени адресата и информации о том, куда они должны были быть возвращены после заполнения.

9. Использование 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 того чтобы получить информацию о людях, не отраженных в муниципальных учетных записях, и для того чтобы пространственная информация была доступна на уровне единиц наблюдения – физических или юридических лиц – 20 административных и/или статистических источников информации, включающих более 400 миллионов индивидуальных записей, были объединены в Интегрированную систему микроданных (SIM). Одним из основных источников, использованных в этом процессе, был реестр разрешений на пребывание жителей из стран не членов Евросоюза. Данные из разных источников были увязаны через использование уникальных налоговых кодов. Вопросы, включенные в SIM: характеристики домохозяйств, место обычного проживания (местоположение работы, школы, колледжа или университета), статус занятости, образование, жилье и обеспеченность жильем и пр.

11. В результате этих нововведений была достигнута фактическая интеграция административных и статистических источников данных об отдельных лицах, домохозяйствах и хозяйствующих субъектах на детальном территориальном уровне. Эта инфраструктура, которая будет совершенствоваться и обновляться ежегодно, будет основным ресурсом не только для проведения нормализации и геокодирования адресов, но также для распространения статистической информации по малым областям, например по номерам домов.

12. Другим важным инструментом, примененным в 2011 году при проведении переписи в Италии, явилась информационная система для организации полевых работ по переписи (SGR – система управления обследованием). SGR представляет собой инструмент для работы в режиме онлайн, при помощи которого муниципалитеты и ИСТАТ могут управлять проведением переписи. Наличие системы SGR позволило проводить проверки в реальном времени и осуществлять мониторинг ежедневной работы переписчиков, а также контролировать процент ответов по каждому каналу сбора данных. Система также дает возможность проводить сравнение данных переписи и данных регистра.

13. В муниципальных образованиях с численностью более 20000 человек были использованы выборочные методы. Обследование проводилось с помощью вопросника двух видов – полного и сокращенного. Были определены переписные области, которые состояли из соседних переписных участков, с общей численностью населения от 13000 до 18000 человек и которые использовались для проведения выборочного обследования по набору социально-демографических переменных из полного вопросника. В отношении будущей стратегии проведения непрерывной переписи, переписные участки могут рассматриваться как части муниципальных образований для ротации выборки через определенное количество лет для муниципальных образований с численностью населения выше 20 000 человек.

III. Планы после 2011 года: непрерывная перепись

14. Несмотря на улучшения, новые подходы и методы, примененные при проведении переписи 2011 года в Италии, недостаточны для того, чтобы достичь стабильного и длительного баланса между затратами на проведение переписи и ее результатами. Затраты остаются высокими и очень сконцентрированными во времени, а использование административных данных все еще недостаточно по сравнению с возможностями в Италии.

15. Кроме того, данные переписи быстро устаревают, а производство детализированных пространственных данных по-прежнему происходит только один раз в десять лет. В связи с этим представляется необходимой разработка совершенно нового подхода.

16. Лесли Киш (Leslie Kish) говорил, что «Производство годовых данных с подробной пространственной привязкой по разным экономическим и социальным показателям, а не только по численности населения, будет основной целью непрерывных переписей во многих странах».

17. Следуя этому предложению, в последнее десятилетие Бюро цензов США стало проводить Обследование американского общества (American Community Survey (ACS)). ACS использует подход

скользящей выборки для того, чтобы обновлять результаты десятилетних переписей посредством проведения ежегодных обследований больших выборочных совокупностей, объединенных для соседних лет, получения оценок на основе скользящих средних на разном уровне пространственной детализации. Во Франции также принят подход непрерывной переписи, использующей подход, отличный от ACS: в течение семи лет проводятся обследования 60% населения и обновляют данные по численности населения каждый год на основе скользящих средних.

18. Основным преимуществом непрерывных переписей является повышение эффективности оценок, благодаря аккумулярованию данных во времени. Это может быть использовано для получения хороших оценок для каждого цикла на заданном уровне географической детализации.

19. Итальянский подход к непрерывной переписи состоит в объединении использования административных источников с выборочными обследованиями с ротацией выборки по многолетним периодам. Для того чтобы преодолеть основной недостаток традиционных переписей, который состоит в слишком большом объеме одновременных мероприятий и слишком больших невозвратных издержках, следует двигаться по пути к новой концепции переписи: не автономные операции, а последовательность операций и специальных обследований, разработанных для создания полной информационной системы для формирования определенных показателей переписи в заданные периоды времени.

20. Для этого нужно будет усовершенствовать дизайн с усилением уже существующей связи между источниками данных и полевыми операциями.

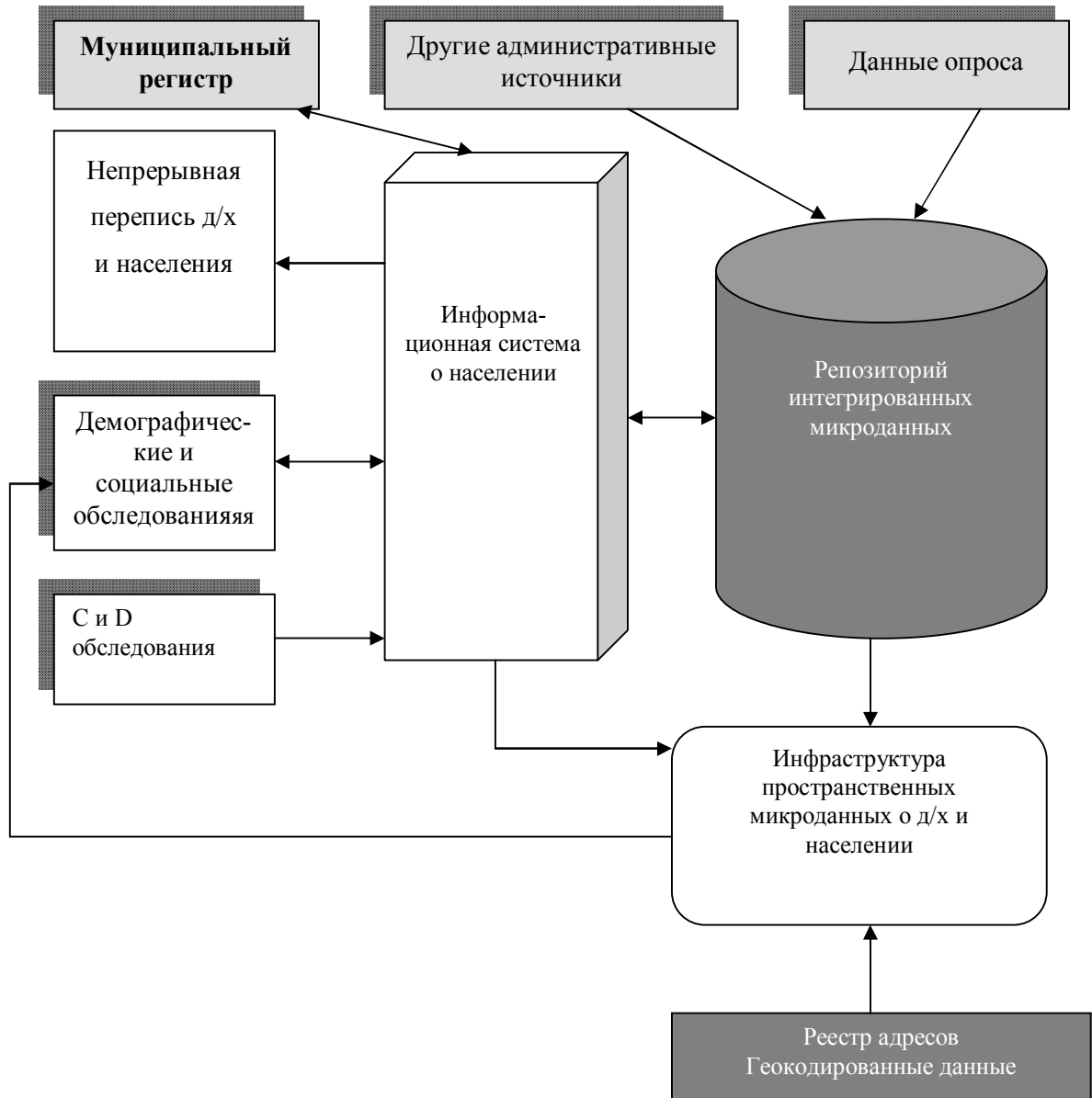
21. Основным источником административных данных будут списки из местных регистров населения LAC (Liste Anagrafiche Comunali). В соответствии с законодательством о регистрах населения (Regolamento Anagrafico) каждое муниципальное образование должно вести местный регистр, который называется Anagrafe. Каждый человек, имеющий обычное место жительства в данном муниципальном образовании, должен быть внесен в регистр с отражением информации о членах домохозяйства и адресе места обычного проживания. Кроме того в регистре должны отражаться все события жизни, такие как рождение и брак, и каждое миграционное движение как внутреннее для данного муниципалитета, так и внешние – внутри страны и из-за рубежа и за рубеж.

22. Стратегия скользящего обновления карт для переписи будет принята для проведения непрерывной переписи. Вся страна будет поделена на пять территориальных блоков, каждый из которых будет включать 20% муниципальных образований с численностью населения меньше 20000 человек, 20% переписных участков с населением 20 000 человек и более. Блоки будут сформированы таким образом, чтобы быть сбалансированными по размеру, географическому распределению и статистической репрезентативности муниципалитетов. Картографирование каждого блока для целей переписи, разбивка на счетные и переписные участки и другие суб-муниципальные образования будет обновляться по очереди каждые пять лет.

IV. Какой должны быть новая модель интеграции регистров и переписей?

22. На рис. 1 показано общая схема отношений между входными и выходными данными с учетом построения всеобъемлющей системы информации о населении в Италии на основе объединения данных из административных источников и обследований. Поскольку успех этого предприятия зависит от эффективной интеграции различных типов статистических единиц и различных тематических областей, движение всех данных из административных источников – прямо или косвенно - происходит в общей среде под названием «Репозиторий интегрированных микроданных». Все процессы по увязке учетных записей и концептуальной интеграции источников происходят в этой среде.

Рис. 1. Взаимосвязи между входными и выходными данными в информационной системе о населении Италии



23. Особая роль в этой структуре отведена данным из муниципальных регистров, поскольку они имеют особое значение со статистической и административной точек зрения по сравнению с другими источниками данных о населении. Они подпадают под специальное регулирование о регистрах населения, которое имеет большое значение для муниципальных образований. По этой причине данные из муниципальных регистров подвергаются предварительной обработке в системе «Информация о населении», прежде чем они объединяются в Репозиторий интегрированных микроданных. В этой общей структуре статистические обследования выполняют функции контроля качества; С-обследование используется для проверки и исправления ошибок охвата в регистрах населения, а D-обследование – для улучшения информации по отдельным темам.

24. Важную роль в системе играет ANSC – упомянутый выше национальный регистр улиц и адресов с его системой географической привязки «улиц» и «адресов», который представляет собой третий столп системы. Такая территориальная база данных важна для определения местоположения единиц, содержащихся в информационной системе. Это существенные элементы для развития интегрированной системы и для управления непрерывной переписью в части компонента картографирования.

25. Что касается картографирования для этой третьей системы, то обновление территориальной

базы (место, переписные участки, счетные участки) должно соответствовать требованиям непрерывной переписи. Будет необходимо постоянно обновлять данные регистра улиц и номеров домов (ANSC): это будет делать ИСТАТ в сотрудничестве с территориальным агентством, отвечающим за территориальный аспект при проведении 15-й переписи населения и жилого фонда в Италии. Поскольку инструменты по учету землепользования, на основе которых сформирована инфраструктура для статистических целей, очень важны для понимания территории, они будут усовершенствованы, и к ним будет обеспечен доступ как к общему информационному ресурсу.

26. Две тематические системы обеспечивают «статистическую» интеграцию посредством использования стратегий и методов объединения различных источников, а также посредством проведения контрольных обследований и обследований, посвященных усовершенствованию информации. В модели особые функции приданы выборочным обследованиям, которые называются С и D обследования. С-обследование нацелено на статистическую корректировку ошибок охвата в административных источниках, а целью D-обследования является сбор информации, отсутствующей в административных источниках.

27. В стратегический план ИСТАТа на 2013-2015 годы включены мероприятия, необходимые для разработки интегрированных пространственных данных с той целью, чтобы начать первый пятилетний цикл в 2017 году и продолжать его до достижения 2021 года – года переписи, когда ИСТАТ представит результаты переписи в соответствии с регламентом ЕС.

V. Выборочные обследования С и D для исправления ошибок охвата и пополнения данных

28. Два разных выборочных обследования (С и D) будут проводиться на важных этапах процесса, разработанного для достижения двух основных целей переписи:

- (a) Перепись постоянно (обычно) проживающих и получение основных данных по демографической структуре населения и домохозяйств;
- (b) Построение гиперкубов социально-экономических данных переписи.

29. Два обследования сильно различаются по масштабам и характеристикам: С-обследование будет специально разработано для оценки и ежегодной корректировки структуры населения, включенного в регистры, имея в виду, по крайней мере, демографические характеристики (пол, возраст, семейное положение, гражданство); D- обследование будет разработано для оценки гиперкубов социально-экономических данных домашних хозяйств и физических лиц, в целях выполнения национальных и международных требований. Основная цель заключается в сборе данных в соответствии со стандартами ЕС по основным характеристикам, не отраженным в регистрах населения; кроме того, D-обследование будет отвечать потребностям пользователей и обеспечит более частое обновление данных, устраняя снижение точности данных в течение десятилетия.

A. Характеристики С-обследования

30. С-обследование будет специально разработано для того, чтобы оценить недостаточный и избыточный охват регистра ЛАС в каждом муниципальном образовании; оно также позволит получить корректировочные факторы. Регистры могут включать лиц, уже не проживающих постоянно (обычно) в данном муниципальном образовании (избыточный охват); и, наоборот, в некоторых случаях лица, обычно проживающие в муниципальном образовании, бывают не включены в регистр (недостаточный охват). Кроме того, в Италии огромное увеличение регулярно проживающих иностранцев требует использования статистических методов для оценки такого вида ошибок. Оценки будут получены методом двойного охвата (*capture-recapture method*).

31. Полная перепись будет проведена посредством полевого сбора данных по кратким формам вопросника в выбранных переписных участках или по спискам адресов. Выборочное обследование должно быть строго независимо и по возможности должно проводиться полевым персоналом, не связанным с муниципалитетом.

32. С-обследование будет разработано для получения точных оценок на муниципальном (LAU2) и суб-муниципальном (переписные участки) уровнях каждый год (первый раз – в 2016 г.).

33. В ходе специальной волны выборочного обследования будут получены оценки для

зарегистрированного населения в год переписи (первый раз – в 2021 г.).

В. Характеристики D-обследования

34. Повышение качества оценок, полученных путем накопления данных во времени, позволит предопределить хорошие оценки в каждом цикле для разных уровней географической детализации. Для сбора социально-экономической информации будут использованы полные формы вопросников.

35. Для того чтобы определить, какие переменные использовать для сбора данных, сначала будет проведено исследование социально-экономической информации, содержащейся в регистрах. Цель состоит в том, чтобы заменить данные, собираемые посредством обследований, на данные, имеющиеся в административных источниках. Для этого необходимо провести более тщательное изучение административных источников, что уже было начато в ходе этой переписи.

36. Дизайн обследования позволит получить надежные годовые оценки. На национальном уровне и на территориальном уровне, соответствующем NUTS1, оценки будут доступны, начиная с 2017 года; они будут получены на основе ежегодных выборок. Более масштабное обследование на основе объединения выборок нескольких соседних лет необходимо для получения оценок на более низком территориальном уровне. Оценки на уровне регионов (NUTS2) и провинций (NUTS3) будут получены на основе объединения выборок за три последовательных года (первые результаты – в 2019 году для 2018 года); муниципальные (LAU2) и суб-муниципальные (переписные участки) оценки будут получены путем объединения выборок за 5 лет (первый выпуск в 2021 г., данные за 2019 г.).

37. Проведение D-обследования будет зависеть от результатов по численности населения и демографической структуре, которые будут получены при проведении выборочного обследования С.

38. Будет использована двухступенчатая выборка, где единицами отбора на первом этапе будут являться муниципальные образования, а на втором этапе – домохозяйства. Муниципальные образования с численностью населения меньше 20 000 человек не будут считаться авто-репрезентативными; они будут разбиты на 5 сбалансированных групп, которые будут наблюдаться в течение 5 ежегодных волн обследования; 509 муниципальных образований с численностью населения как минимум 20 000 человек, будут авто-репрезентативными, и в этих муниципалитетах отбор домохозяйств будет проходить каждый год.

39. Первая волна стартует в 2017 г. В каждую волну будет обследовано от 600 000 до 900 000 итальянских домохозяйств (альтернативный подход состоит в том, чтобы проводить сбор данных только за две волны обследования в 2017 и в 2021 гг., при этом в каждую волну следует обследовать от 1 500 000 до 2 200 000 домохозяйств).

40. Недостатки, связанные с более высокими ошибками выборки при проведении обследования на низком территориальном уровне (муниципальном и суб-муниципальном) могут быть компенсированы более высокой периодичностью производства статистических данных.

41. После проведения отбора и составления списков всех домохозяйств для обследования на основе регистра LAC, можно будет применить некоторые подходы, апробированные в раунде 2011 с целью сокращения операционной нагрузки: i) отправлять вопросники по почте; ii) получать «автоматические ответы» по многим каналам (интернет, обратные почтовые отправления, муниципальный офис для приема заполненных вопросников) iii) выявлять неотвечивших лиц при помощи переписчиков; iv) использовать онлайн системы управления переписью (SGR) такого же типа, как была использована при проведении переписи в Италии в 2011 г.

VI. Инфраструктура пространственных микроданных по физическим лицам и домохозяйствам

42. Наконец, на основе инфраструктуры пространственных микроданных по отдельным лицам и домохозяйствам будут получены результаты в географической привязке по домохозяйствам, отдельным лицам и хозяйствующим субъектам с использованием данных, хранящихся в информационной системе. При помощи этой инфраструктуры ИСТАТ сможет предоставлять пользователям микро- и макроданные эффективно и прозрачно при обеспечении конфиденциальности. В основном, данные будут сформированы на основе LAC - муниципальных

регистров населения. Ко всем единицам будет применено геокодирование по переписным участкам.

43. Эта инфраструктура представляет собой инновационную центральную структуру производства специальных продуктов и характеризуется использованием пространственной информации по домохозяйствам, отдельным лицам и хозяйствующим субъектам; она была создана в результате анализа и сравнения потребностей внешних пользователей и оценки качества входящей информации. Инфраструктура использует на входе геокодированные данные на элементарном уровне из Интегрированной системы микроданных и позволяет получить результаты лонгитюдных и одномоментных обследований, которые важны для микро- и макро-анализа, при этом соблюдаются принципы конфиденциальности и анонимности. Эта система также позволит преодолеть историческую дихотомию между социальным и экономическим анализом. Возможность одновременно проводить анализ характеристик отдельного человека и экономической единицы, к которой он/она имеет отношение, предоставляет более широкое поле для анализа, что невозможно без этого инструмента.

44. Эта инфраструктура будет использована для предоставления государственным органам, исследователям и пользователям данных в привязке к интегрированным элементарным статистическим единицам и территориальным единицам (до переписных и счетных участков). Для обеспечения релевантности и прозрачности продуктов, предлагаемых в рамках данного проекта, будут проанализированы потребности пользователей в информации и предоставлена документация об источниках и классификациях, а также о методах интеграции данных и методах их обработки.

VII. Заключение

45. До сих пор, включая 2011 год, перепись проводилась один раз в десять лет. Новая стратегия в Италии будет состоять в проведении «скользящей (непрерывной)» переписи и в добавлении более широкого использования административных источников к выборочным обследованиям с ротацией в течение нескольких лет.

46. Проведение работы в непрерывном режиме приведет к существенному повышению эффективности полевых операций и к повышению качества данных. Постоянная работа на полевом уровне на местах позволит сохранять и развивать опыт и знания; ожидается, что менее интенсивная, но постоянная полевая работа позволит усовершенствовать методологию и накопить опыт.

47. Такой подход окажет положительное влияние на финансирование. Потребность в государственных финансах будет распределена во времени, и при непрерывных операциях контракты на оказание услуг могут оказаться более привлекательными и, возможно, более дешевыми, чем при «залповых» операциях. Производство данных на постоянной основе также будет способствовать более прочным и значимым отношениям с пользователями.