



Европейская экономическая комиссия

Конференция европейских статистиков

Восемнадцатое совещание

Женева, 28–30 сентября 2016 года

Пункт 7 предварительной повестки дня

**Возможные виды использования новых источников
данных (например, «больших данных») в целях переписей**

Новые источники данных и новые технологии в традиционной методологии переписи населения: планы Всероссийской переписи населения раунда 2020 года

**Записка представлена Федеральной службой государственной
статистики (Росстат), Российская Федерация**

Резюме

Опыт последней Всероссийской переписи населения 2010 года диктует потребность в модернизации некоторых подходов к проведению переписи населения, без которых уже невозможно достичь главной ее цели – получения полной и достоверной картины современного общества. Перепись должна применять современные методы сбора сведений о населении и снижать нагрузку на население

При переписи 2020 года будут применяться наряду с традиционными бумажными опросными листами электронные вопросники на планшетных компьютерах и Интернет-перепись. Исследованы все имеющиеся административные базы данных о населении и определена подходящая для целей переписи Единая система идентификации и аутентификации населения, она будет использоваться как точка входа на портал Интернет-переписи. Кроме того, рассматривается возможность использования больших данных геоаналитики – агрегированных данных операторов мобильной связи – для контроля полноты охвата населения переписью..

I. Введение

1. Опыт последней Всероссийской переписи населения 2010 года диктует потребность в модернизации некоторых подходов к проведению переписи населения, без которых уже невозможно достичь главной ее цели – получения полной и достоверной картины современного общества.
2. Перепись должна применять современные методы сбора сведений о населении и снижать нагрузку на население. Это связано с развитием информационных технологий, широким использованием Интернета в повседневной жизни, высказанным уже в 2010 году желанием жителей страны заполнить переписные листы самостоятельно в удобное для них время, большими сложностями в привлечении временных переписных работников, особенно в крупных городах.

II. Общие положения.

3. При переписи 2020 года будут использованы переписные электронные и бумажные вопросники (идентичные по составу вопросов) и 3 способа сбора сведений о населении:
 - - самостоятельное заполнение респондентами электронных вопросников в сети Интернет;
 - - опрос и заполнение переписчиком электронных вопросников на планшетных компьютерах с предустановленным специализированным программным обеспечением;
 - - опрос и традиционное заполнение переписчиком бумажных вопросников.
4. Сбор сведений в течение всего срока проведения переписи будет осуществляться также в специальных помещениях – стационарных участках – на электронных или бумажных вопросниках.
5. При ВПН-2020 предполагается ввести электронную государственную услугу прохождения переписи населения путем заполнения электронного вопросника самими респондентами на государственном портале переписи. Возможность пройти Интернет-перепись будет предоставлена всему населению. На основе опыта переписей населения других стран можно ожидать от 25 до 60% охвата населения этих городов Интернет-переписью.
6. По окончании периода Интернет-переписи переписчики будут обходить все помещения счетных участков, фиксировать подтверждение прохождения переписи по Интернету и проводить опрос населения, не прошедшего Интернет-перепись (с использованием планшетного компьютера или на бумажных переписных листах).
7. Планшетные компьютеры с электронными вопросниками могут использоваться переписчиками крупных населенных пунктов.
8. В населенных пунктах в труднодоступных и отдаленных районах, а также в небольших по числу жителей сельских населенных пунктах, когда нецелесообразно использование планшетных компьютеров будет использоваться метод заполнения переписчиком бумажных вопросников. При этом часть труднодоступных в 2010 году территорий может перейти в категорию обычных территорий за счет переписи по Интернету (например, в районах, где есть Интернет-связь, но нет возможности физического доступа переписчиков).

9. Во время работы переписчиков планируется использовать современные технологические решения (GPS-модуль планшетного компьютера, другие электронные устройства передачи информации о местонахождении переписчика) для контроля качества работы переписного персонала.

10. Для контроля полноты охвата переписью всех жилых помещений, в которых проживает (или может проживать) население, будет использовано электронное хранилище адресов респондентов, созданное нами при последней переписи населения 2010 года.

11. На основе данных этого хранилища будут составляться организационные планы переписи и проводиться мониторинг полноты охвата населения Интернет-переписью, будет отслеживаться факт участия в переписи респондента любым из предоставляемых способов - в сети Интернет, на стационарном участке или дома при обходе переписчиком.

III. Новые источники данных

12. Росстат провел изучение возможностей использования имеющихся в стране административных данных о населении для целей проверки непротиворечивости собранных данных переписи населения. Административные базы данных созданы для ведомственного учёта населения в Фонде обязательного медицинского страхования, Пенсионном фонде, Федеральной налоговой службе, Паспортно-визовой службе. Все эти источники имеют методологические особенности и определенную ограниченность или избыточность накапливаемых сведений.

13. Так, Федеральная налоговая служба осуществляет учёт физических лиц по месту их жительства, а также по месту нахождения принадлежащих им недвижимого имущества, транспортных средств и другого имущества. Поэтому физическое лицо может быть учтено одновременно в налоговых органах различных территориальных единиц.

14. В базах данных Пенсионного фонда индивидуальный лицевой счёт открывается застрахованному лицу один раз. При изменении места жительства застрахованного лица его индивидуальный лицевой счёт остаётся на учете в месте открытия счета и не передаётся в другую территориальную единицу. Актуализация данных лицевого счёта застрахованного лица ведется по месту открытия счета, а не по месту фактического жительства человека.

15. В Фонде обязательного медицинского страхования учёт застрахованных лиц по страховым медицинским организациям ведется независимо от перемены места жительства и регистрации по месту пребывания. В учетных данных паспортно-визовых служб лица, выехавшие в другие населенные пункты, также не подлежат снятию с регистрационного учета согласно действующему законодательству.

16. Таким образом, в силу специфики характера существующих ведомственных данных о населении, они не могут использоваться как альтернативный или вспомогательный источник при статистическом учете населения и переписи населения.

17. На сегодняшний день в России действует Единый портал государственных услуг, который является средством доступа к электронным услугам различных государственных ведомств для населения России, зарегистрированного в Единой системе идентификации и аутентификации (ЕСИА). В ЕСИА на сего-

дняшний день зарегистрировано около 25 млн человек (почти 17% населения страны). К 2020 году ожидается увеличение числа зарегистрированных пользователей и за счет объединения с муниципальными и ведомственными системами идентификации населения, и за счет расширения списка предоставляемых государственных услуг.

18. ЕСИА содержит персональные данные «подтвержденных» пользователей, предоставленные ими при первичной регистрации и проверенные в ведомственных базах данных Пенсионного фонда и Федеральной миграционной службы. Часть этих персональных данных соответствует и собираемым при переписи населения характеристикам - пол, дата и место рождения, гражданство, место проживания, данные о детях (пол и дата рождения) и контактная информация. Организация Интернет-переписи посредством добавления этой услуги на Единый портал государственных услуг сделает возможным при он-лайн заполнении переписных листов использование респондентом уже имеющихся логина и пароля, а также данных своей учетной записи - персональной информации, содержащейся в ЕСИА. При этом мы должны учитывать административный временной интервал актуализации персональных данных на портале и обеспечить соблюдение принципов самоопределения при переписи населения. Поэтому планируется предоставление возможности респонденту изменить в переписном листе данные по сравнению с содержащимися в ЕСИА.

19. Еще одно направление развития и нововведений к переписи населения 2020 года - использование «больших данных» для повышения охвата и качества итогов переписи. Большие данные представляют собой источник новой информации, которую официальная статистика не может игнорировать.

20. Как известно, применение «больших данных» подразумевает работу с информацией огромного объема, разнообразного состава и неорганизованной структурой, часто поступающей в режиме реального времени и происходящей из разных (чаще всего, частных) источников.

21. В качестве первого шага в этом направлении мы организуем в ближайшее время анализ существующих в России источников нестатистических данных («больших данных») о численности и/или составе населения и об адресах жилых помещений, в которых проживает или может проживать (пребывать) население в каждом из регионов России, а также попытаемся провести оценку их доступности и стоимости использования для целей переписи населения.

22. Один из таких «нестатистических» источников использует средства геоаналитики – современного подхода в области получения агрегированных массивов данных о населении от операторов мобильной связи с помощью технологических возможностей «Big Data», с последующим исследованием и статистическим анализом этих данных. В основе геоаналитики лежит метод, заключающийся в том, что за счет анализа технологических событий сети операторов мобильной связи и построения математических, предиктивных моделей он позволяет получать информацию о населении, исключая использование и обработку персональных данных абонентов. С его помощью можно исследовать численность и динамику перемещений населения на больших территориях во времени.

23. В качестве доказательства того, что исследование численности и динамики перемещений населения на территории с помощью метода геоаналитики представляет собой эффективный и востребованный источник информации, приводится то, что данные геоаналитики уже задействованы в своей работе

многими государственными и муниципальными организациями, в частности в сфере градостроительства.

24. В рамках подготовки к Всероссийской переписи населения 2020 года Росстатом предполагается для контроля полноты охвата населения переписью рассмотреть возможность использования следующих показателей больших данных геоаналитики: *численности населения*, проживавшего на заданной территории в отчетном месяце; и *среднего количества человек*, находившихся на заданной территории у себя дома (на работе, на даче) в ночь на дату переписи.

25. Росстат рассматривает большие данные как потенциальный элемент информационной статистической системы и как один из возможных источников данных по различным статистическим направлениям, в том числе по демографии. Использование больших данных согласуется с реализуемой нами программой модернизации российской статистики для снижения респондентской нагрузки и экономии бюджетных средств. Для интеграции источников больших данных в информационную систему, прежде всего, необходимо решить проблему нормативно-правовой базы доступа к большим данным и их использования при формировании официальной статистической информации.
