

**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 b) предварительной повестки дня

**Инновации, запланированные для раунда переписей
2020 года, и результаты тестирования****Планы в отношении канадской переписи 2016 года,
касающиеся тестов****Записка Статистического управления Канады***Резюме*

Канада проводит перепись населения каждые пять лет. Следующая перепись состоится в мае 2016 года. Такая периодичность создает широкие возможности для ускоренного внедрения инноваций применительно ко всем аспектам программы и одновременно порождает трудности, связанные с необходимостью отводить достаточно времени на проведение тщательных тестов. Один из используемых подходов состоит в том, чтобы в ходе подготовки к следующему раунду переписи разрабатывать контролируемые тесты, которые будут осуществляться в реальных условиях в процессе переписных работ.

В настоящем документе кратко описывается методология проведения переписи 2016 года. Далее в нем приводится информация о тестах, проведенных в рамках подготовки к переписи, и некоторые выводы по их результатам. В нем также описываются тесты, запланированные для проведения в реальных условиях в 2016 году. В заключение в нем приводится краткая информация о мероприятиях, запланированных в целях более широкого использования данных из административных источников в рамках программы переписи, а также о некоторых из запланированных исследований и тестов для оценки предполагаемого применения таких источников.



I. Введение

1. Статистическое управление Канады в соответствии с Законом о статистике обязано проводить перепись населения каждые пять лет. Канадская программа переписи включает в себя собственно перепись населения при помощи переписных листов, состоящих из примерно десяти вопросов, касающихся базовых демографических сведений, а также Общенациональное обследование домашних хозяйств, в ходе которого собирается более подробная социально-экономическая информация по выборке домашних хозяйств. В основе программы лежит комбинированный подход, в соответствии с которым производится сбор данных от физических лиц и данных о жилищных единицах по состоянию на определенный момент времени и некоторая часть информации – из административных источников (так, переменные доходов составляются на основе данных налоговой службы). Проведение переписи каждые пять лет позволяет более оперативно оценивать инновационные методы и принципиальные изменения, затрагивающие любые элементы программы. Для этого также необходимо планомерно обеспечивать своевременное проведение тестов, необходимых для внедрения таких инноваций. По сути дела, планировать более масштабные изменения на будущий раунд необходимо уже в ходе текущего раунда.

2. В соответствии с обычной практикой Статистического управления Канады все значимые изменения в методологии и подходах, затрагивающие важные статистические программы, тщательно тестируются перед их внедрением. В случае программы переписи населения тесты, касающиеся содержания и организации переписи, обычно проводятся в течение двух или трех лет, непосредственно предшествующих сбору данных. Для целей программы 2016 года тест содержания (ПТ 2) был проведен в мае–июне 2014 года, а тест операционных изменений (ПТ 3) – в октябре 2014 года. Качественное тестирование (ПТ 1) содержания и инструментария вопросника проводилось перед ПТ 2.

3. Для программы переписи также характерно и то, что новые подходы иногда тестируются в контролируемых условиях уже в ходе переписных мероприятий, в рамках подготовки к следующему раунду. Такие тесты в настоящее время разрабатываются и для переписи 2016 года.

4. В настоящем документе представлено краткое описание тестов, проведенных или запланированных в связи с программой переписи 2016 года. По результатам уже проведенных тестов излагаются основные выводы. В отношении предстоящих тестов описываются допущения исследователей. Помимо этого, Статистическое управление Канады в настоящее время проводит ряд исследований в целях более активного использования административных данных в рамках программы переписи населения. В настоящем документе кратко описываются некоторые из связанных с этим исследовательских проектов и запланированных тестов, имеющих отношение к программе 2016 года.

II. Поведенческие тесты для раунда 2016 года

5. В течение последних нескольких лет Статистическое управление Канады занимается реорганизацией своей бизнес-архитектуры. Основная цель реорганизации состоит в том, чтобы рационализировать инвестиции в инфраструктуру и общеорганизационные процессы благодаря разработке глобальных подходов в отличие от локальных подходов для отдельных программ. Эти инициативы затрагивают и программу переписи, особенно учитывая, что необходимые для нее крупные инвестиции можно с большей отдачей использовать в интересах реализации общеорганизационных инициатив. Так, системы сбора данных на местах,

которые используются для целей переписи, можно адаптировать с учетом требований всех остальных программ обследований, уменьшив объем необходимых общеорганизационных инвестиций в эти системы, а также сократив со временем количество требующих поддержки систем и платформ. Чтобы общеорганизационные инструменты могли использоваться в целях программы переписи в 2016 году, программа переписного тестирования была пересмотрена таким образом, чтобы основное внимание в ней уделялось изменениям по сравнению с последней переписью 2011 года, нежели проведению полномасштабной «генеральной репетиции». Стратегия тестирования была разделена на более мелкие поведенческие тесты (ПТ) для проверки поведения респондентов и персонала на местах в связи с использованием новых систем и подходов.

6. Содержание и инструментарий переписного листа проходили проверку с использованием поведенческого теста № 2 (ПТ 2) в мае и июне 2014 года. В частности, в целях тестирования использовался новый общеорганизационный инструмент в виде электронного вопросника, заменивший собой специальное приложение для проведения переписи, применявшееся начиная с переписи 2006 года. Тестовые возможности переписи использовались для тестирования применимости электронного вопросника на различных технологических платформах, таких как настольные компьютеры, планшеты и смартфоны. Учитывая запланированный показатель получения 65% ответов через Интернет и стремительные изменения в технологии, было необходимо удостовериться в том, что электронный вопросник будет надлежащим образом работать на всех платформах, которыми могут воспользоваться респонденты в 2016 году. Тест подтвердил, что запланированный показатель получения ответов через Интернет выбран правильно и что вопросник функционирует заданным образом на всех платформах. Еще одно тестирование платформ будет проведено в меньшем масштабе в конце 2015 года для проверки работоспособности электронного вопросника при использовании новейших технологий.

7. Поведенческий тест № 3 (ПТ 3) был проведен осенью 2014 года для проверки применения новых систем и процессов регистрации персоналом на местах, а также проверки соответствующих учебных материалов. В частности, нами была протестирована новая общеорганизационная Комплексная система сбора данных и проведения работ (КСДР). Эта система заменяет собой систему сбора данных на местах, которая использовалась в ходе переписи 2011 года, и обеспечивает интернет-интерфейс для получения местным персоналом рабочих заданий, учета выполнения работ и связанной с этим информации о выплатах и расходах. В данных системах также предусмотрен ряд вспомогательных функций, таких как обмен сообщениями, составление информационных отчетов для руководства и т.д. ПТ 3 проводился отдельно от ПТ 2 и позже по времени, в отличие прошлой практики подготовки к переписи, когда больше времени отводилось на доработку необходимых для использования в ходе переписи функциональных возможностей КСДР.

8. Тест КСДР продемонстрировал, что некоторые функции, разработанные в качестве универсальных (т.е. предназначенных для использования как обычными счетчиками в рамках любых обследований, так и переписчиками), слишком сложны для условий переписи, в рамках которой объем необходимого/предоставляемого обучения целесообразно ограничивать по финансовым соображениям, а также по той причине, что данная работа рассчитана на выполнение в течение короткого периода времени и должна быть освоена сравнительно быстро. С учетом результатов этого теста некоторые из универсальных функций в настоящее время упрощаются, чтобы их можно было использовать в ходе переписи 2016 года. Примером этого служит множество функций, разработанных для администраторов на местах в целях выдачи заданий переписчикам, в том числе

исходя из среднего числа попыток посещения каждого жилища, близости адресов к месту проживания счетчика и т.д. Такие функции оказались слишком сложными в контексте переписи.

9. На основании результатов ПТ в системы вносятся ряд корректировок и изменений. Чтобы удостовериться в том, что эти изменения позволят эффективно устранить недостатки, отмеченные в ходе тестирования, Статистическое управление Канады проведет еще один маломасштабный проверочный тест систем осенью 2015 года. Для проведения этого теста будут привлекаться неспециалисты, чтобы убедиться в том, что в контексте проведения переписи, когда для выполнения работ в течение шести–десяти недель будет нанято и обучено более 25 000 человек, использование систем будет интуитивно понятным.

III. Тестирование расширенного использования административных данных

10. В 2014 году Статистическое управление Канады начало реализацию программы исследований для изучения возможности более активного использования административных данных в рамках переписи населения. Одна из целей этого исследования заключается в изучении возможности использования данных из административных источников вместо традиционной регистрации фактической численности населения и базовых демографических характеристик. Канада использует относительно традиционный подход к проведению переписи населения, поскольку у нее не имеется всеобщих регистров населения или жилищ, а также какой-либо системы уникальных личных идентификаторов (Rouse, 2013).

11. Одной из основных инициатив, подкрепляющих исследование административных источников, является разработка Статистическим управлением Канады прототипов для Канадской базы данных демографической статистики – КБДДС (Cug and al., 2013). Построение этой базы данных начинается с данных федеральной налоговой службы, затем добавляются данные о рождениях и вычитаются данные о смертях, а также учитываются некоторые подгруппы населения с помощью альтернативных источников, таких как Регистр индейского населения и продольный файл данных об иммиграции. Первые два прототипа КБДДС были разработаны для описания состава населения за базисный 2011 год и позволяли получать демографические данные, менее чем на 1% расходящиеся с общенациональными оценками численности населения за тот же период. Вместе с тем более существенные пробелы в этих прототипах были обнаружены на малых географических уровнях или по некоторым категориям населения, выделяемых по признаку возраста или пола.

12. Статистическое управление Канады пока не определило, какими должны быть критерии успеха этой работы или сроки ее завершения. Пока этого не сделано, результаты, получение которых может быть задано в КБДДС, будут оцениваться в течение следующих двух лет. Чтобы отвечать условиям оценки, они не обязательно должны иметь такой же глобальный уровень охвата. Например, при доступности своевременной базовой демографической информации по большому числу жилищ такую информацию можно было бы теоретически использовать вместо дорогостоящих повторных обходов в ходе проведения переписи. Она потенциально пригодна для подтверждения статуса заселенности жилищ по состоянию на дату переписи, для использования вместо повторных обходов в связи с непредставлением ответов, в отношении с трудом поддающихся регистрации районов, в целях исследований по проблемам охвата и т.д.

13. Создание прототипа базового массива данных о населении (КБДДС) сопряжено с целым рядом сложностей. Одна из них касается своевременности, а именно возможности составить массив данных, охватывающий большую долю населения, своевременно и в сроки, совпадающие с проведением переписи. Сложность в данном случае связана со временем, которое необходимо для получения и сопоставления файлов определенного возраста, так как задействуемая информация должна быть как можно более актуальной. Еще одна сложность в условиях Канады связана с возможностью правильного установления жилища, т.е. основного места жительства лиц, выявленных по административным источникам. В большинстве административных файлов, имеющих в распоряжении Статистического управления Канады, указания такой адресной информации не требуется. Приведенные адресные сведения могут соответствовать второму месту жительства или, если речь идет о налоговых данных, даже адресу бухгалтера, заполнявшего налоговую декларацию.

14. Для проверки применимости КБДДС в поддержку переписных мероприятий изучается возможность проведения двух наборов тестов. Возможные последствия замены прямой регистрации на использование административных данных можно оценить, смоделировав функционирование КБДДС в рабочих условиях на основе информации переписи 2016 года. Будут разработаны контрольные тесты для определения того, какие результаты могли бы быть получены в 2016 году благодаря использованию КБДДС вместо повторного обхода некоторых жилищ, не предоставивших ответы. Например, можно установить правило, согласно которому после x попыток посещения конкретных жилищ дальнейшие попытки их посещения должны быть прекращены. Такое правило может быть соблюдено при условии, что в КБДДС имеется качественная информация о тех, кто проживает в отобранных для теста жилищах. Учетный период данных КБДДС должен быть максимально близок к дате переписи 2016 года (т.е. 10 мая). Тест призван продемонстрировать, что вместо прямой регистрации лиц, находящихся в таких жилищах, можно полагаться на информацию из КБДДС без существенного ущерба для получаемых в результате данных на уровне малых районов. Поскольку КБДДС не будет содержать всех переменных, регистрируемых в переписном листе (например, будет отсутствовать информация о языке), отсутствующая информация будет вменяться. Итоговые данные в новом файле затем необходимо будет сопоставить с официальными результатами переписи 2016 года для проверки их точности.

15. Одним из основных преимуществ замены повторного обхода использованием административных данных может стать значительная экономия расходов. Такая экономия возможна благодаря уменьшению числа обходов жилищ, не предоставивших ответы, что является одним из самых дорогостоящих мероприятий в ходе переписи. Частично экономия затронет постоянные расходы в связи с более быстрым сворачиванием инфраструктуры на местах, поскольку в некоторых районах сбор данных будет заканчиваться раньше.

IV. Тестирование регистрации личных идентификаторов

16. Увязка записей в ходе изучения административных источников в Статистическом управлении Канады всегда осуществляется исходя из сочетаний имен респондентов, их возраста или даты рождения, пола и адреса. Хотя в Канаде не применяется уникальных личных идентификаторов, использование которых упростило бы такую увязку, подходящей заменой им может служить номер социального страхования (НСС). Большинство канадцев имеют НСС, и, кроме того, этот номер содержится в некоторых административных файлах, имеющих в распоряжении Статистического управления Канады, в частности данных налого-

вой службы. Сложность заключается в том, что в настоящее время Статистическое управление Канады не имеет прямого доступа к информации об НСС.

17. Регистрация номеров социального страхования в ходе непосредственного опроса респондентов была опробована в рамках ПТ 2. Цель этой части теста заключалась в том, чтобы проверить готовность канадцев указать свой НСС в переписном вопроснике, их способность точно назвать номер при регистрации в опосредованном режиме (переписной лист обычно заполняет одно лицо от имени всего домашнего хозяйства), а также возможное влияние соображений, связанных с конфиденциальностью и другими факторами, на предоставление ответов и возврат вопросников как в ходе переписи, так и в ходе Общенационального обследования домашних хозяйств. Данное влияние оценивалось в отношении режимов предоставления ответов через Интернет и с помощью бумажных вопросников и сопоставлялось с данными по контрольным группам, в которых НСС не регистрировался. Переписной компонент теста проводился также с соблюдением положений Закона о статистике об обязательном воспроизведении реально существующих условий переписи.

18. В ходе теста было установлено, что большинство лиц готовы указать свой НСС и что по эти данные в основном имеют высокое качество. Чуть более 80% респондентов в ходе теста представили свой НСС и лишь 1,6% указали неправильный номер. В ходе анализа результатов респонденты, не имеющие НСС, не выделялись в отдельную группу, поэтому фактическая доля респондентов без НСС неизвестна. В ходе теста были также отмечены факторы, оказывающие влияние на организацию работы. Нагрузка на респондентов при заполнении вопросников с вопросом об НСС увеличивалась как минимум на 20%, если исходить из дополнительного времени, затраченного на заполнение вопросника в Интернет. Было отмечено также уменьшение показателей возврата вопросников – примерно на 2% в случае Общенационального обследования домашних хозяйства и на 2–3% в случае переписи на этапах второй и третьей волны (когда неотвечившим лицам направляются напоминания). Более медленные темпы возврата вопросников необязательно приводят к снижению коэффициента предоставления ответов по завершении процесса сбора данных, однако могут означать увеличение рабочей нагрузки на начальном этапе повторных обходов неотвечивших лиц, потенциально увеличивая расходы на сбор данных на местах.

19. В ходе дальнейшего анализа было установлено, что наличие НСС не обеспечивает существенного повышения коэффициента увязки данных с данными налогового учета (увеличение составило менее 1%), что подтверждает достаточную эффективность подхода, применяемого в Статистическом управлении Канады в настоящее время. Тем не менее наличие НСС, несомненно, ускорило бы процесс увязки. Учитывая неоднозначный характер результатов теста, в рамках переписи 2016 года информация об НСС собираться не будет.

V. Тестовые испытания в реальных условиях, запланированные для реализации в ходе проведения переписи 2016 года

20. В рамках переписи 2016 года Статистическое управление Канады проведет ряд тестов в процессе сбора данных. Эти тесты будут проведены в целях накопления знаний, имеющих значение для планирования следующей переписи 2021 года. Они включают в себя сбор данных об адресах электронной почты в качестве контактной информации, регистрацию как почтового, так и учетного адреса (если они различаются), использование цифровых каналов доставки при-

глашения заполнить переписной лист и более широкую рассылку вопросников по почте в некоторых районах.

21. Респондентам переписи 2016 года будет предложено указать в вопроснике свой адрес электронной почты. Эта информация будет добавлена в основу выборки Обследования домашних хозяйств, служащей источником адресов для целей переписи. Адреса электронной почты собираются в рамках изучения альтернативных способов установления контактов с канадскими домашними хозяйствами как для целей самой переписи, так и любого обследования, выборочной совокупностью которого служит основа выборки Обследования домашних хозяйств. В ходе теста будет оцениваться выполнение инструкций (доля респондентов, предоставивших адрес электронной почты), вероятность корректной регистрации такой информации и возможные факторы, имеющие значение для порядка регистрации данных при заполнении респондентами бумажных вопросников. В последнем случае, в частности, предстоит оценить, насколько технология оптического распознавания позволяет автоматически считывать и распознавать адреса электронной почты.

22. Параллельно со сбором данных об адресах электронной почты Статистическое управление Канады в ходе этапа регистрации в 2016 году протестирует электронную технологию контактирования домашних хозяйств. В ходе этого теста с 50 000–100 000 домашних хозяйств контакты будут осуществляться с помощью цифрового почтового ящика (epost) – сервиса, предлагаемого Канадской почтовой службой. Данный метод будет использоваться в отношении тестовой группы подписчиков на услугу epost и заменит собой запланированную во всех остальных случаях связь по почте. Кроме того, нами будет отобрана контрольная группа из числа подписчиков epost для проверки достоверности результатов, полученных по тестовой группе. Предварительно служба epost обратится к домашним хозяйствам с просьбой дать согласие на то, чтобы Статистическое управление Канады имело полномочия направлять материалы в их цифровой ящик. Давшие согласие лица получают приглашение участвовать в переписи через свой цифровой почтовый ящик. По своему содержанию цифровое сообщение будет очень схожим с бумажным приглашением, рассылаемым домашним хозяйствам по традиционной почте. При помощи этого теста будут измерены показатели возврата вопросников и предоставления ответов в «цифровой» группе по сравнению с контрольной группой. Полученные результаты будут использованы для определения способов контактирования домашних хозяйств в 2021 год.

23. Статистическое управление Канады использует несколько способов контактирования домашних хозяйств для целей переписи. На данный момент предполагается, что в 2016 году примерно 82% жилищ будет сначала направлено приглашительное письмо с предложением проживающим в них лицам заполнить переписной лист в режиме онлайн. В отношении примерно 17% жилищ Статистическое управление Канады будет использовать методику раздачи и сбора переписных листов, предполагающий доставку вопросников непосредственно на дом, в первую очередь в сельских районах, поскольку почтовая служба не в состоянии доставлять в конкретные жилища почтовые управления, на которых не указано адреса. Жители этих районов получают почту на общие почтовые ящики, в то время как Статистическое управление Канады может адресовать письма только «жильцам» того или иного жилища.

24. Разграничение между районами, где переписные листы рассылаются по почте, и районами, где они раздаются и затем собираются, устанавливается отчасти исходя из практических соображений. Например, некоторые жилища, куда возможно организовать рассылку, оставлены в районах раздачи и сбора, чтобы избежать необходимости применения переписчиками смешанной методологии и

тем самым упростить процедуры. В предлагаемом тесте нами будет опробован метод «рассылки с дополнительной раздачей». В некоторых районах, по которым у Статистического управления Канады имеется большая доля действительных адресов, на эти адреса будут высланы приглашения, при этом переписчикам на местах будет выдан перечень известных жилищ, по которым у Управления не имеется действительного почтового адреса, для доставки по ним писем. Тест призван оценить осуществимость такого подхода, возможность сокращения расходов по сравнению с методом раздачи и сбора переписных листов, а также возможность обеспечения надлежащего охвата жилищ. Если этот подход окажется успешным, то в 2021 году Статистическое управление Канады может потенциально охватить рассылкой не менее 90% жилищ.

25. Наконец, по каждому жилищу Статистическое управление Канады будет собирать информацию как об учетном адресе, так и о почтовом адресе лиц. Регистр адресов содержит информацию об учетных адресах, которые не всегда совпадают с почтовыми, особенно в сельских районах. Установление соответствий между этими двумя адресами может сделать возможным более широкую рассылку вопросников по почте в ходе следующей переписи 2021 года. Это будет также способствовать исследованию административных источников, поскольку позволит корректно определять место жительства лиц, что является ключевой предпосылкой для составления базового массива данных о населении.

VI. Справочные документы

Boulet, C., Mathieu, P., Statistics Canada, 2015, Testing the addition of the Social Insurance Number to the Canadian Census of Population; Presented at the Joint Statistical Meetings of the American Statistical Association, Seattle, August 2015.

Royce, D. 2011. Preliminary Report on Methodology Options for the 2016 Census. Prepared for the 2016 Census Strategy, Census Management Office.

Cyr, A., Pignal, J. Trépanier, J., Statistics Canada, 2013 Canadian Statistical Demographic Database: A Research Project, Presented at the Statistics Canada Methodology Symposium.