

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

Двадцать первое совещание

Женева, 18–20 сентября 2019 года

Пункт 4 предварительной повестки дня

**Исследования по вопросам использования
административных данных в целях переписи****Новаторские подходы для применения в ходе переписи
населения Канады 2021 года и в последующий период****Записка Статистического управления Канады****Резюме*

Перепись населения в Канаде проходит каждые пять лет, и проведение очередной переписи запланировано на май 2021 года. Канада по-прежнему применяет традиционный подход, в основу которого заложена в первую очередь «методология волн», обеспечивающая поступление значительной доли ответов через Интернет. Несмотря на недавние успешные случаи применения этого подхода, Статистическое управление Канады изыскивает возможности для внедрения в будущем комбинированной модели проведения переписи, которая опирается прежде всего на информацию, полученную при наличии возможности из статистических регистров, и дополняется традиционными подходами к сбору данных, когда наличие административных источников (для развернутой формы) или качество регистров являются удовлетворительными. В настоящем документе описано проводимое исследование по развитию статистических регистров, цель которого – содействовать внедрению комбинированного подхода, вероятно, для проведения цикла 2026 года. Кроме того, в нем кратко изложены методы сбора данных, которые планируются к применению для переписи 2021 года, в частности рассмотрен вопрос о том, как можно усовершенствовать функционирование статистических регистров наряду с получением информации из других источников для содействия мероприятиям по повторному опросу на местах в 2021 году.

* Документ подготовили Сильвен Делиль, Марк Амель и Софи Лефевр.



I. Введение

1. В соответствии с законом о статистике Статистическому управлению Канады поручено каждые пять лет проводить перепись населения; очередная перепись состоится в 2021 году. Программа переписи населения в Канаде включает в себя краткий вопросник, в котором содержится порядка десяти вопросов по базовым демографическим характеристикам, и развернутый вопросник для сбора более подробной социально-экономической информации на основе выборки, составляющей 25% домохозяйств. Эта программа рассчитана на комбинированный подход, предусматривающий сбор данных в домохозяйствах об отдельных лицах и о единицах жилья, производимый в конкретный момент времени, и получение некоторой информации из административных записей. Приведем пример: получение переменных дохода производится из данных отчетности о подоходном налоге, а информация о категории иммигранта заимствуется из иммиграционной продольной базы данных, которая ведется Канадской службой по делам иммиграции, беженцев и гражданства.

2. Несмотря на успешное проведение переписи 2016 года, в ходе которой были отмечены рекордно высокие показатели ответов через Интернет (68,4%) и самостоятельного заполнения вопросников (88,8%), ожидается, что при проведении последующих переписей Канада столкнется с некоторыми проблемами. Во-первых, нынешняя методика со временем вряд ли сохранит свою стабильность, как это обычно происходит с любым применяемым подходом по мере дальнейшей эволюции общества. Успех нынешнего подхода объясняется прежде всего «методологией волн», которая предусматривает обращение к домохозяйствам по почте¹. Более того, в случае применения электронного вопросника как основного способа участия в переписи требуется использование почтовой системы, рассылающей каждому домохозяйству надежный код доступа. Граждане Канады могут постепенно отказаться от такого способа коммуникации, поскольку все более широкое распространение получают цифровые средства связи. Проведенные на сегодняшний день испытания альтернативных вариантов доставки этой информации респондентам, таких как электронные и цифровые почтовые ящики, не дали положительных результатов. Во-вторых, канадцы выразили обеспокоенность в связи с нагрузкой по предоставлению данных, которая налагается на них государственными учреждениями различных уровней². В определенной степени ожидается, что перед проведением непосредственного сбора информации программа переписи должна быть нацелена на получение данных, уже собранных населением и даже частными организациями. Наконец, нельзя обходить вниманием нынешние международные тенденции и исследования в сфере подходов к проведению переписей, которые частично или полностью опираются на административные данные.

3. Чтобы преодолеть упомянутые вызовы, Статистическое управление Канады приступило к работе по новому исследовательскому направлению – Проекту преобразования программы переписи (ППП). Цель этого проекта – обеспечить в ходе переписи населения более пристальное внимание интересам пользователей и респондентов, чтобы перепись позволила успешнее удовлетворить растущие информационные потребности, уменьшить нагрузку на канадцев и сократить объем затрат. По опыту предшествующих инициатив – таких, как Проект стратегии проведения переписи 2016 года³, – был сделан вывод о том, что проведение административной переписи с опорой исключительно на данные регистров (как это наблюдается, например, в странах Северной Европы) не является для Канады возможным в кратко- или среднесрочной перспективе, поскольку в стране нет условий, необходимых для внедрения такого подхода.

4. В настоящем документе приводится краткое описание исследований, проведенных в связи с перспективой применения комбинированного подхода в ходе будущих канадских переписей. В нем также рассмотрены возможности для частичного

¹ Hamel, 2012.

² Statistics Canada, 2019.

³ Howatson-Leo, Trépanier, 2011.

выполнения работы в этом направлении с целью содействия операциям по сбору данных в рамках будущей переписи 2021 года.

II. Разработка статистических регистров

5. С 2010 года Статистическое управление Канады изучает вопрос о том, как можно обеспечить дальнейший сбор административных данных в рамках переписи населения Канады. Некоторая информация, подготовленная для целей переписи, была получена в ходе последних трех циклов из административных источников (например, в 2016 году – данные о доходах и категориях иммигрантов). Перед внедрением того или иного смешанного подхода был рассмотрен «эффект воздействия» различных способов или источников представления ответов на вопросы переписи; результаты показали, что качество полученной информации существенно не пострадало⁴. В целом же более качественными обычно считаются данные, получаемые из административных источников и онлайн-вопросников.

6. Административная перепись в том виде, в котором она проводилась, например, в странах Северной Европы (в Дании в 1981 году, в Финляндии в 1990 году и в Норвегии и Швеции – в 2011 году) или в Нидерландах, нередко характеризуется как самый передовой способ использования административных данных или регистров для получения информации о численности населения и единиц жилья. Тем не менее к канадским условиям эта модель не подходит, поскольку в Канаде отсутствуют ее основные элементы – единая система административных регистров и личный идентификационный номер для каждого постоянного жителя страны.

7. Первоначальное исследование темы об использовании административных источников в ходе переписи было сосредоточено на построении статистического регистра населения (Канадской базы данных демографической статистики (КБДДС)) путем компилирования информации, полученной из различных административных источников, в первую очередь из налоговых и иммиграционных данных, а также информации из регистров рождения и смерти. Получив первые обнадеживающие результаты исследования по КБДДС, Статистическое управление Канады приступило к разработке статистического регистра населения (СРН) и к преобразованию адресного регистра (АР) в статистический регистр строений (СРС). Эти регистры являются статистическими регистрами, т. е. они конструируются на основе различных источников административных данных и будут использоваться только для содействия Статистическому управлению Канады в выполнении своих функций по производству защищенных обезличенных статистических продуктов. СРН будет обеспечивать постоянно обновляемый список лиц, проживающих в Канаде по состоянию на какой-либо конкретный момент времени, наряду с их основными демографическими характеристиками, такими как возраст, пол при рождении и место проживания.

8. СРС станет отражением эволюции адресного регистра Статистического управления Канады и будет обеспечивать точный список всех имеющих в Канаде строений, включая коллективные жилища, заселенные или незаселенные частные жилища и нежилые здания.

9. СРН и СРС в сочетании с имеющимся коммерческим регистром (КР) и потенциально возможным регистром активности (РА) сформируют инфраструктуру основных взаимосвязанных статистических регистров, которая будет использоваться для всех программ Статистического управления Канады. Например, СРН можно было бы применять в рамках программы оценки численности населения для проведения оценочных расчетов в периоды между переписями или в качестве основы выборки для программ социальной статистики.

⁴ Grondin, Sun, 2008.

III. Расширение комбинированного подхода для проведения последующих переписей

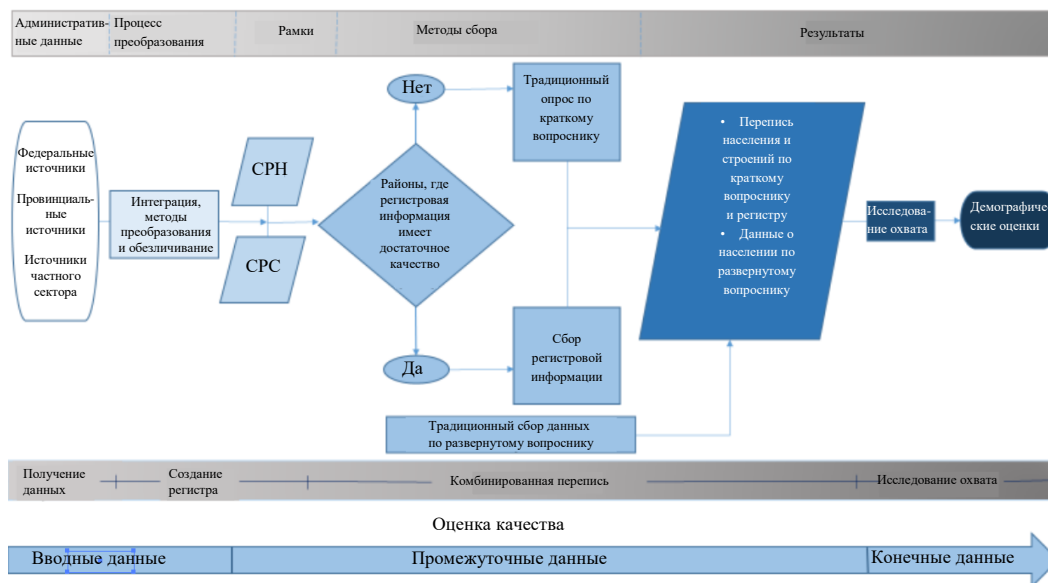
10. В таких странах, как Соединенное Королевство или Новая Зеландия, изучаются возможности внедрения такого подхода к проведению переписи, который частично или полностью опирается на административные данные даже при отсутствии надежной единой системы государственных регистров и личных идентификационных номеров. С 2016 года Статистическое управление Канады рассматривает вопрос об использовании информации, которая накапливается в потенциально возможном регистре СРН, в качестве вероятного нового способа ответов на вопросы при проведении переписи населения.

11. Основной вывод из проведенных на сегодняшний день исследований заключается в том, что необработанные административные данные, находящиеся сейчас в распоряжении Статистического управления Канады, обладают большим потенциалом для обеспечения всестороннего охвата канадского населения и в конечном счете могли бы использоваться для расчетов его численности. Нужно проделать дополнительную работу, поскольку исследования также указывают на необходимость усиления охвата на местном уровне посредством совершенствования нынешних административных записей или составления новых. Тем не менее информацию из статистических регистров, скорее всего, можно считать достаточно качественной для использования в обозримом будущем во многих областях программы проведения переписей. Такой результат можно обеспечить путем расширения применения нынешнего комбинированного подхода к проведению переписи.

12. В рамках расширенного комбинированного подхода к переписи в Канаде получение информации из статистических регистров будет сочетаться с традиционным сбором данных. По сути это предполагает, что генеральная совокупность переписи жилищного фонда Канады будет разделена на две части с использованием различных характеристик (охват и точность регистра, тип строения, география, практические аспекты, расходы и т. п.). В одной части результаты переписи будут рассчитываться не на основе прямого опроса, а с применением административных данных достаточного качества, взятых из СРН и СРС. Таким образом, СРН будет использоваться для определения показателя численности населения с базовыми демографическими данными, такими как дата рождения, пол при рождении и географическое местонахождение. Традиционный опрос с помощью краткого вопросника будет применяться во второй части, когда информация из регистров не может считаться достаточно качественной или точной. В дополнение к этому подходу будет продолжено составление репрезентативной выборки строений для представления подробных характеристик в форме развернутого вопросника, как это предусмотрено нынешней моделью. На рис. I показаны различные шаги от источников административных данных к демографическим оценкам.

Рис. I

От источников административных данных к демографическим оценкам при комбинированном подходе к проведению переписи



13. В настоящее время Статистическое управление Канады проводит ряд операций по моделированию этого подхода, сопоставляя ранние итерации CPN с базой данных переписи 2016 года и с результатами из Программы демографических оценок как для подтверждения качества статистических регистров, так и для выявления оптимальной методики применения этих регистров при комбинированном подходе. Аналогичные операции по моделированию будут проводиться параллельно с переписью 2021 года; по итогам этого исследования будет принято решение о возможности применения расширенного комбинированного подхода в 2026 году.

IV. Использование статистических регистров для проведения переписи 2021 года

14. В рамках Программы переписи населения 2021 года Статистическое управление Канады вновь будет широко прибегать к повторному использованию тех подходов, которые успешно применялись по программе 2016 года. Однако возможности для совершенствования нынешнего подхода, ориентированного на максимально высокую долю самостоятельных ответов, теперь являются более ограниченными. Наиболее значимая возможность для повышения эффективности существующего подхода к проведению переписи заключается в использовании уже доступной информации, что позволяет снизить потребность в отслеживании жилищ, не представивших свои ответы. Использование информации, полученной в результате мероприятий по сбору, а также данных из новых статистических регистров и новых источников административных данных, имеющихся в распоряжении Статистического управления Канады, могло бы способствовать снижению издержек и нагрузки, позволяя в то же время поддерживать высокий уровень качества переписи. Стратегия, предлагаемая для проведения переписи в 2021 году, предусматривает три основных задачи:

a) сократить избыточность охвата генеральной совокупности переписи: снизить число жилищ, изъятых из эксплуатации, до 3,5% (с 4,5% в 2016 году) посредством сверки с данными из административных источников;

b) прогнозировать уровень занятости жилищ: выявлять изъятые из эксплуатации, незаселенные и заселенные единицы жилья в случаях, когда ответы не представлены;

с) прекратить отслеживание ряда случаев, когда ответы не представлены, и использовать данные из СРН в качестве опосредованного ответа.

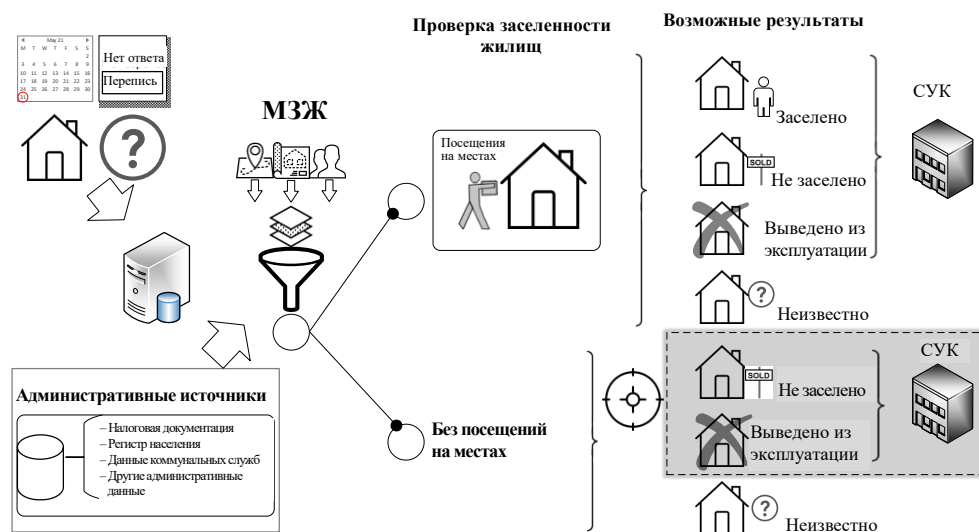
15. Сокращение избыточности охвата будет широко обеспечиваться благодаря поиску соответствий различным административным записям. В качестве генеральной совокупности на разных циклах переписи использовался адресный регистр (АР). Большинство обновлений в АР вносятся на основе административных источников. Основным источником данных для этой цели служат материалы из пунктов обращения граждан, которые курирует Почта Канады. Менее 20% адресов выверяются в ходе внутриведомственных процедур или через списки на местах.

16. Процесс новой постпереписной итерации АР будет ретроактивно охватывать и те позиции, которые были впервые исключены из выборки при проведении предыдущей переписи. Затем эти единицы жилья войдут в генеральную совокупность для следующей переписи исходя из предположения, что на тот момент они могли находиться в эксплуатации. Данный подход был изменен при первом обновлении данных АР после переписи 2016 года. Это обновление затрагивало только частное жилье, находящееся в эксплуатации со времени проведения переписи 2016 года; единицы жилья, исключенные из выборки в 2016 году, будут возвращены в нее лишь в случае, если они будут числиться в административных записях, используемых для обновления АР. Кроме того, разрабатывается метод выявления дублирующих регистровых записей, касающихся одного и того же жилья. Были разработаны другие внутренние методы для уточнения процессов выявления соответствий и выверки, цель которых – сводить к минимуму версию адресов.

17. Один из нынешних подходов к сбору информации состоит в том, что процедура переписи 2021 года будет частично усовершенствована благодаря использованию новых источников административных данных и расширению ее целей. Меры по проверке заселенности жилья (ПЗЖ), направленные на выявление тех единиц жилья, которые максимально близко к дню переписи являются незаселенными и неэксплуатируемыми, будут опираться на более точное составление выборки с применением новых источников административных данных.

18. Для прогнозирования статуса заселенности каждого жилого здания разрабатывается модель заселенности жилищ (МЗЖ). На основе информации из различных административных источников, таких как налоговая документация, данные коммунальных служб и статистические регистры, перед проведением операций на местах каждому жилищу будет присвоен прогнозный вероятный статус «заселено», «не заселено» и «выведено из эксплуатации». Этот процесс отражен на рис. II. Жилища, которые с высокой вероятностью являются незаселенными или неэксплуатируемыми (т. е. они вряд ли заселены), будут включены в выборку для проведения ПЗЖ. Выборка для ПЗЖ также составляется на основе практических критериев, таких как распределение жилищ по кластерам для контроля расходов, связанных со сбором данных. В процессе ПЗЖ статус заселенности проверяется сотрудниками на местах.

Рис. II
Процедура на основе модели заселенности жилищ



19. В зависимости от результатов применения МЗЖ перед определением выборки для ПЗЖ могут применяться другие методы обработки информации с полной отменой посещений на местах или сокращения их числа в период проведения последующей работы в связи с получением ответов (ПРНО):

а) жилые здания с высокой вероятностью того, что они не заселены, и с низкой вероятностью того, что они выведены из эксплуатации или заселены, будут классифицироваться как незаселенные;

б) жилые здания с высокой вероятностью того, что они выведены из эксплуатации, и низкой вероятностью того, что они не заселены или заселены, будут классифицироваться как выведенные из эксплуатации.

20. Третий элемент этой стратегии даст возможность прекратить последующие действия в связи с некоторыми не ответившими жилищами в процессе ПРНО благодаря использованию данных СРН в качестве опосредованного ответа. Стратегия, которая применяется в Канаде в настоящее время, направлена на получение ответов на уровне не менее 98% для каждой выборки объемом порядка 45 000 единиц, определенной для процедуры ПРНО. В этой связи данные СРН могли бы использоваться для предварительно установленного максимального числа единиц жилья для каждого переписного участка в тех районах, где уже были приложены активные, но безуспешные усилия с целью получения ответов от не ответивших жилищ, или даже для повышения доли ответов в переписном участке с уже достигнутой долей ответов, составляющей 98% и более.

21. К числу критериев, применимых для выявления единиц жилья, данные о которых будут фигурировать в СРН, относятся: высокая вероятность заселенности соответствующего жилья, определяемая на основе МЗЖ; данное жилье не относится к выборке для развернутого вопросника; сравнительно высокое качество данных из СРН; и установление предела для такой замены в рамках каждого переписного участка. Следует устанавливать пределы для использования такого типа ответов для целых домохозяйств во избежание воздействия на вариативность результатов исследований охвата переписи, которые сами основаны на данных из административных источников.

V. Заключение

22. Статистическое управление Канады активно продвигается в своем исследовании административных альтернатив сплошной переписи. С другой стороны, любой случай широкомасштабного применения таких подходов потребует

тщательного анализа. Данные переписи населения служат основой для избирательной системы Канады. Кроме того, ежегодные трансферы объемом в миллиарды долларов, осуществляемые федеральным правительством в провинции и территории (и, следовательно, в муниципалитеты), опираются на результаты переписи. Одной из ключевых характеристик любого решения о последующем расширенном применении комбинированного подхода должно стать качество данных.

23. С другой стороны, применение комбинированного подхода потребует определенных компромиссов. Например, сейчас некоторые результаты переписи обеспечиваются на основе краткой формы опроса (100% населения), включающей, например, информацию о семейном положении и образовании семьи, которую нужно будет преобразовать в развернутую форму (25% выборки), поскольку в административной документации такие сведения обычно не фигурируют. Вероятно, это не является большой проблемой, поскольку ценность этих переменных намного возрастает в сочетании с другими переменными в рамках развернутой формы. Тем не менее перед будущим расширением применения комбинированного подхода потребуются проведение масштабных консультаций с пользователями данных и заинтересованными сторонами.

Ссылки

Hamel, Internet data collection in the Canadian Census of Population, UNECE Expert Group Meeting on Censuses Using Registers, Geneva, 2012

Statistics Canada (2019). Modernization: a key to Statistics Canada's effort to reduce response burden, https://www.statcan.gc.ca/eng/about/smr09/smr09_100

L. Howatson-Leo, and J. Trépanier, (2011). Final Report on the 2016 Census Options: Proposed Content Determination Framework and Methodology Options, Statistics Canada. http://www12.statcan.gc.ca/census-recensement/fc-rf/reports-rapports/r2_index-eng.cfm

C. Grondin, L. Sun, 2006 Census Internet Mode Effect Study, Joint Statistical Meeting of the American Statistical Association, August 2008
