

**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****Изучение административных источников и наборов  
геопространственных данных для использования  
в ходе переписи населения Канады 2021 года****Записка Статистического управления Канады\****Резюме*

При подготовке программы переписи населения Канады 2021 года были проведены широкие консультации по ее содержанию. В ходе этих консультаций был выявлен ряд пробелов в данных, которые потребовали дополнения или пересмотра существующего содержания. В целях уменьшения нагрузки профильным группам было поручено изучить административные источники, которые могли бы заменить данные, получаемые путем опроса, добавить новое содержание или улучшить качество имеющегося без соответствующего увеличения нагрузки на респондентов и расходов. В пользу такого подхода свидетельствует тот факт, что в 2016 году вместо вопросов о доходах успешно использовались данные из административных источников. Несколько профильных групп провели исследования для изучения возможности замены и дополнения существующего содержания путем интеграции других административных источников данных, включая данные по иммиграции и жилью. В дополнение к этим исследованиям ведется работа по использованию административных источников данных для дальнейшего уточнения таких понятий, как поездки на работу, а также для расширения информации о лицах в местах коллективного проживания, в том числе на военных базах, в домах престарелых и тюрьмах. В настоящем документе описывается проведенное в Канаде исследование по использованию административных данных в целях замены, дополнения и улучшения качества элементов существующего содержания для применения в ходе переписи населения Канады 2021 года.

\* Автор: Хизер Драйберг.



## I. Введение

1. При подготовке программы переписи населения Канады 2021 года были проведены широкие консультации по ее содержанию. В ходе этих консультаций был выявлен ряд пробелов в данных, которые потребовали дополнения или пересмотра существующего содержания.
2. В целях уменьшения нагрузки профильным группам было поручено изучить административные источники, которые могли бы заменить данные, получаемые путем опроса, добавить новое содержание или улучшить качество имеющегося без соответствующего увеличения нагрузки на респондентов и расходов. В пользу такого подхода свидетельствует тот факт, что в 2016 году вместо вопросов о доходах успешно использовались данные из административных источников.
3. В настоящем документе описывается проведенное в Канаде исследование по использованию административных данных в целях замены, дополнения и улучшения качества существующего содержания для применения в ходе переписи населения Канады 2021 года.

## II. Административные данные

4. В рамках предметных областей переписи проводятся исследования для оценки качества и изучения преимуществ и недостатков замены определенных элементов содержания вопросника данными из административных источников. Профильные эксперты изучают возможности получения из административных источников информации о статусе иммигрантов и годе иммиграции, а также о жилье.
5. Помимо работы по дополнению содержания, проводится оценка потенциала использования административных источников данных для уточнения таких понятий, как поездки на работу, а также для расширения информации о лицах в местах коллективного проживания, в том числе на военных базах, в домах престарелых и исправительных учреждениях.

### A. Иммиграция

#### 1. Замена данных, получаемых с помощью вопросов, касающихся иммиграции, в рамках переписи населения 2021 года

6. Канада располагает значительными массивами административных данных по иммиграции, что позволяет рассмотреть возможность их использования вместо опроса респондентов в ходе переписи. В ходе переписи населения 2016 года данные по таким переменным, как *категория въезда* и *тип заявителя* позволили собрать информацию об условиях, на которых иммигранты получили право на проживание в Канаде. Эти переменные были получены из административных записей об иммигрантах в Канадском управлении по вопросам гражданства, иммиграции и беженцев (УГИБ). Административные данные по этим переменным имелись в наличии только в отношении иммигрантов, прибывших в период начиная с 1980 года. В результате только те лица, которые, отвечая на вопрос о годе иммиграции в переписном листе (вопрос 15), заявили, что они иммигрировали в Канаду в 1980 году или позже, оказались в пределах диапазона, подлежащего замене данными из административных источников. Недостающие значения по переменной «категория въезда» были рассчитаны методом импутации в отношении иммигрантов, по которым отсутствовали данные в административных источниках.
7. В рамках подготовки к переписи 2021 года Канада проводит тестирование замены годе иммиграции и статуса иммигранта, используя аналогичную методологию и источники данных. В настоящее время Статистическому управлению Канады доступны следующие административные данные УГИБ:

- а) подробные данные по иммиграции с 1980 года по настоящее время;

- b) ограниченные данные по иммиграции в период с 1952 по 1979 год;
- c) данные по непостоянным резидентам с 1980 года по настоящее время.

8. Эти массивы содержат данные о фактическом административном статусе иммигрантов и год иммиграции, когда это применимо, в отношении соответствующей категории иммигрантов и непостоянных резидентов. Данные УГИБ отражают административную регистрацию всех прибывающих в Канаду иммигрантов и непостоянных резидентов, однако они не обновляются и не учитывают случаи смерти или эмиграции. Поэтому эти данные не могут использоваться как самостоятельный источник для оценки численности иммигрантов, проживающих в Канаде в настоящее время.

9. С использованием методов, аналогичных тем, которые применялись в предыдущих переписях (Biernot, 2017; Brennan, 2013; Brennan, 2011), планируется провести тестирование интеграции данных, полученных от респондентов в ходе переписи, и административных данных УГИБ сразу же после пробного сбора данных переписи в 2019 году и до этапа редактирования и импутации недостающих данных. Для интеграции данных из двух источников будут использоваться все доступные общие признаки, такие как имя, дата рождения, адрес и т. д. В целях тестирования информация, полученная в ответ на вопросы о статусе иммигранта и годе иммиграции, будет заменяться значением из административных источников и затем сопоставляться. В отношении иммигрантов, по которым отсутствуют данные в административных источниках, недостающие значения будут рассчитаны методом импутации.

10. Вопрос о гражданстве неразрывно связан с обоими элементами (год иммиграции и статус иммигранта), поэтому ответы на вопросы о гражданстве будут использоваться для определения того, в каких случаях требуется условный расчет значения. Население Канады разделено на четыре подгруппы, как показано в таблице 1 ниже:

Таблица 1

**Иммиграционный статус и гражданство, Канада**

<i>Гражданский статус</i>	<i>Иммигрант</i>	<i>Неиммигрант</i>
Гражданин Канады по рождению		Гражданин Канады по рождению
Гражданин Канады в результате натурализации	Гражданин Канады в результате натурализации	
Не гражданин Канады	Постоянный резидент	Непостоянный резидент

11. Иммигрантом не может быть лицо, являющееся гражданином Канады по рождению. И наоборот, любое лицо, являющееся канадским гражданином в результате натурализации, должно быть иммигрантом. Лица, не имеющие канадского гражданства, являются либо постоянными, либо непостоянными резидентами. В некоторых случаях респонденты могут указать, что они являются гражданами Канады по рождению, но в административных данных могут проходить как иммигранты – такие противоречия будут устраняться в ходе обработки данных.

12. В ходе исследования также будет проанализировано, как замена данных, получаемых с помощью вопросов, касающихся иммиграции, может повлиять на сопоставимость данных за прошлые периоды. С целью сопоставления различий в качестве данных, полученных путем опроса и из административных источников, была проведена оценка качества используемых в ходе переписи вопросников. В докладе, который будет представлен в сентябре 2019 года, будут полностью отражены преимущества и недостатки обоих методов вместе с рекомендациями для правительства по содержанию переписи населения 2021 года.

13. Дополнительная административная информация, содействующая совершенствованию иммиграционной статистики, которая может быть использована в ходе переписи 2021 года и основана на данных по иммиграции УГИБ, включает в

себя: год прибытия; тип временного резидента (лица, имеющие разрешение на работу, разрешение на учебу, ходатайствующие о предоставлении статуса беженца, лица, имеющие другие разрешения); пункт назначения; последнюю страну постоянного проживания. В настоящее время проводятся консультации с заинтересованными сторонами с целью определения окончательного перечня переменных, оценки качества переменных и учета практических соображений.

## 2. Ограничения при использовании административных данных по иммиграции

14. Замена данных, получаемых с помощью двух касающихся иммиграции вопросов в переписном листе, сопряжена с тремя ограничениями, обусловленными охватом данных УГИБ.

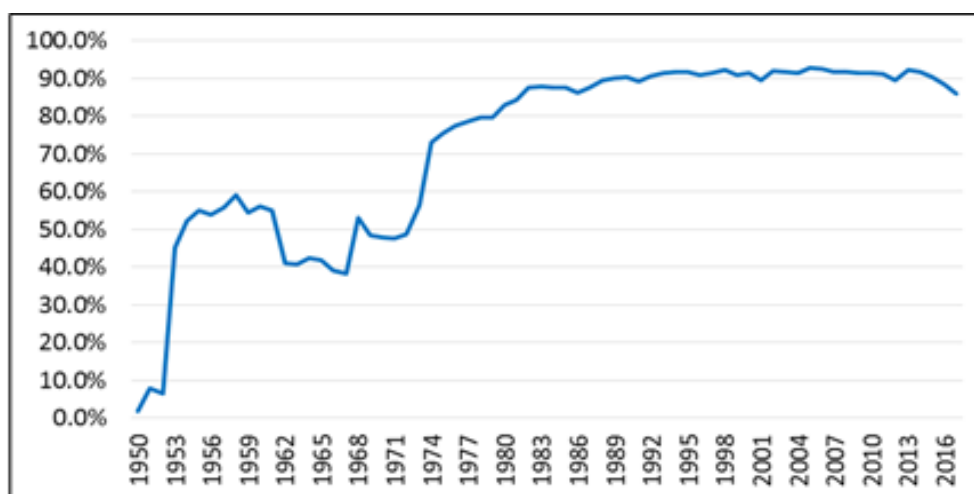
15. Во-первых, отсутствуют данные по иммиграции до 1952 года. Хотя численность иммигрантов, прибывших до 1952 года, сокращается и представляет меньший интерес для исследователей, по данным переписи населения 2016 года, эта группа насчитывала 111 000 иммигрантов (или 1,5% всех иммигрантов), проживающих в Канаде. Из них 86 285 человек (или 1,2% от общего числа иммигрантов) прибыли в страну с 1947 по 1951 год, а остальные (составляющие 0,3% от общего числа иммигрантов) иммигрировали до 1947 года.

16. Во-вторых, записи о непостоянных резидентах содержат информацию только о лицах, имеющих разрешения, и не включают данные о каких-либо сопровождающих членах семьи, не имеющих отдельного разрешения. Хотя некоторые приезжие отвечают на вопросы переписного листа, информация по ним не включена в административные данные УГИБ, интегрируемые с данными переписи 2016 года.

17. В-третьих, помимо проблемы охвата данных, имеется еще одно ограничение, связанное с качеством данных об иммиграции за период до 1980 года. В целом, они содержат меньше информации, пригодной для целей интеграции данных. В частности, данные за период 1961–1972 годов содержат неполную информацию о дате рождения, что снижает успех интеграции. Рис. I иллюстрирует то, как неполная информация о дате рождения влияет на степень успеха интеграции. В период 1961–1972 годов наблюдается заметное снижение этой степени.

Рис. I

**Процентная доля иммигрантов, участвовавших в переписи населения 2016 года, данные, успешно интегрированные с административными данными учета иммиграции, в разбивке по годам иммиграции, выделявшимся в рамках переписи 2016 года**



Источник: Statistics Canada, 2016 Census integrated with IRCC administrative immigration data since 1952.

18. Кроме того, переход непосредственно от опроса к использованию административных данных означает, что один источник данных не может быть использован для подтверждения данных из другого, как это было в случае переписи

2016 года. Замена опроса данными из административных источников может повлиять на сопоставимость с данными за предыдущие периоды. В частности, могут появиться смещения в распределении данных по годам иммиграции, когда вместо «считающегося» годом иммиграции, указываемого самим респондентом (например, года прибытия), будет использоваться фактический год иммиграции, взятый из административных данных. Например, некоторые иммигранты могли указать год своего прибытия, поскольку они, возможно, проживали в Канаде в качестве непостоянных резидентов до получения статуса постоянного жителя.

19. Охват непостоянных резидентов является меньшим по сравнению с постоянными. Информация о непостоянных резидентах ограничивается данными о просителях убежища и временных резидентах, имевших действующее разрешение в 2016 году.

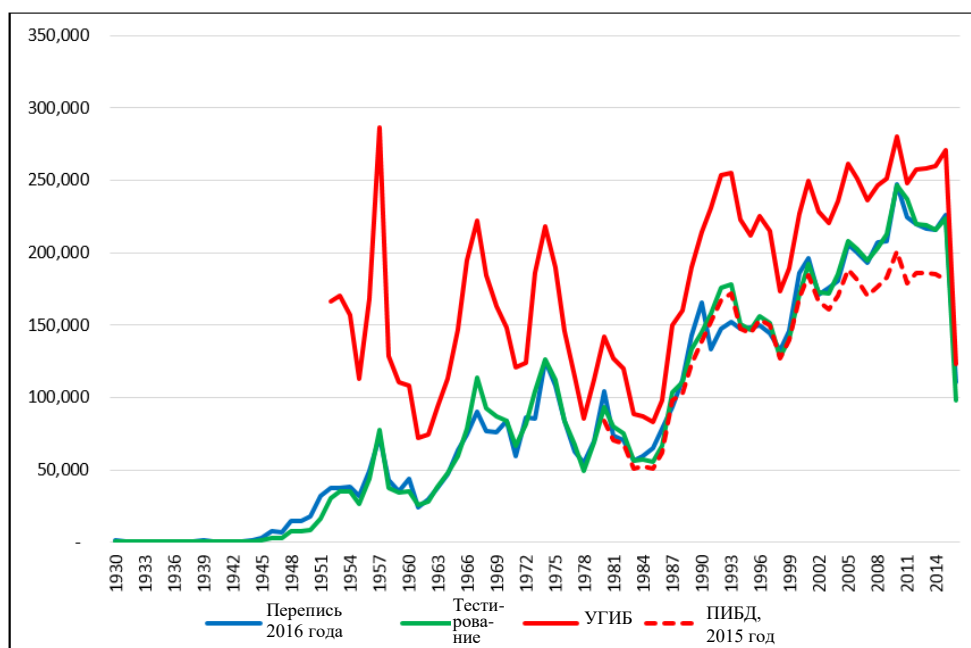
### 3. Исследование: как решить проблему недостающих для интеграции данных или несоответствий между интегрированными значениями и прочими ответами на вопросы переписи

20. В принципе, могут применяться аналогичные использовавшимся в прошлом методы редактирования и импутации недостающих данных. Однако в контексте использования интегрированных данных отсутствие значения, возможно, является следствием недостаточных для увязки данных (или эквивалентом отсутствия ответа по отдельным позициям) или указывает на респондентов вне диапазона, используемого для увязки. В настоящее время тестируется интеграция других источников данных с целью смягчения проблем, обусловленных недостаточным охватом административных данных.

21. В целом, как показано на рис. II, данные переписи населения 2016 года и предварительные результаты замены административными источниками являются схожими. Однако результаты использования административных данных все же демонстрируют некоторое расхождение между ответами иммигрантов и административными данными. Это наглядно иллюстрирует эффект отсроченной реакции, вызванный большим притоком просителей убежища в начале 1990-х годов: респонденты указали год прибытия вместо года иммиграции.

Рис. II

**Предварительные результаты в разбивке по годам иммиграции, результаты переписи 2016 года и тестирования**



*Источник:* Statistics Canada, 2016 census and administrative data replacement results, Census of Population and Longitudinal Immigration Database (IMDB) for tax year 2014; and Immigration, Refugees and Citizenship Canada (IRCC).

## **В. Жилье**

### **1. Использование больших данных для получения информации о коммунальных услугах**

22. В ходе предыдущих переписей населения канадцам задавали ряд вопросов об их жилищных условиях, в том числе о расходах на коммунальные услуги, такие как электроснабжение. Статистическое управление Канады изучает практическую возможность использования информации поставщиков электроэнергии вместо этого вопроса в переписном листе. В Канаде довольно много поставщиков электроэнергии. Статистическое управление Канады заключило соглашения о получении данных за платежные циклы от многих из них. На сегодняшний день несколько коммунальных предприятий могут предоставлять ежемесячные данные о потреблении и фактурировании Статистическому управлению Канады.

23. В рамках подготовки к переписи 2021 года Статистическое управление изучит эти данные, преследуя две конкретные цели. Во-первых, данные будут проанализированы на предмет того, помогут ли они Статистическому управлению Канады определить статус заселенности жилых помещений, не представивших ответы. Желаемый результат будет заключаться в сокращении последующих усилий по уточнению. Во-вторых, будет проведена оценка степени пригодности этих данных для замены вопроса о платежах за электроэнергию в ходе переписи.

24. Исследования по изучению возможности интеграции или замене данными поставщиков коммунальных услуг находятся на начальном этапе. В рамках первого этапа необходимо будет оценить, каким образом можно изменить переписной лист, чтобы учесть тот факт, что в настоящее время полный охват страны невозможен (к 2021 году доступ к своим данным предоставит только часть поставщиков). Как показала пробная перепись 2019 года, можно адаптировать электронный вопросник к конкретным областям и случайным подгруппам, чтобы представить ту или иную версию вопросника (с конкретными вопросами или без них). Предварительные результаты также указывают на сокращение времени заполнения в тех домохозяйствах, где вопрос об электричестве не задавался. Однако эти преимущества были частично сведены на нет путаницей, внесенной некоторыми респондентами, которые, возможно, отнесли платежи за электроэнергию к платежам за топливо или другие коммунальные услуги. Важное значение будет иметь оценка качества этих факторов.

25. Второй этап будет заключаться в сопоставлении информации по счетчикам электроэнергии с данными переписи населения по жилым помещениям и анализе расхождений. В большинстве случаев можно установить взаимно однозначное соответствие и провести сопоставление данных. Однако зачастую несколько счетчиков могут обслуживать одно жилище или, напротив, один счетчик – несколько. А другие не предоставляют достаточную информацию для надлежащей идентификации соответствующей переписной жилищной единицы. Эти более сложные проблемы касаются конкретных районов или типов жилья. На данном этапе будет проведена оценка того, как правильно распределить потребление между жилищными единицами.

26. На заключительном этапе будут осуществлены оценка влияния смешанного подхода на результаты и прогноз конечного эффекта такой интеграции данных.

### **2. Ограничения использования административных данных о жилье**

27. На интеграцию административных данных в перепись распространяются общие оговорки. Все они должны быть учтены или сведены до приемлемого уровня риска:

а) совместимость понятий (сообщенные платежи за электроэнергию совпадают с информацией из административных источников);

б) достаточный охват (пока охвачены не все провинции, а интеграция с использованием данных по счетчикам и жилью может привести к сокращению существующего охвата в определенных районах);

с) риск, связанный с предоставлением данных (требуется систематическое предоставление данных);

d) единообразии по всей стране (в данном случае понятия относительно просты);

e) единообразии методов (анализ продолжается: респонденты часто приводят округленные цифры).

28. Данная работа поможет уменьшить нагрузку на респондентов, так как это относительно сложный для предоставления ответа вопрос.

## С. Поездки на работу

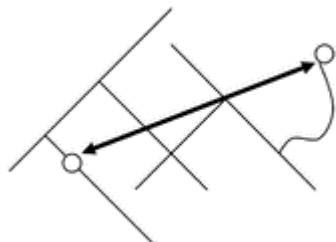
### 1. Исследования по использованию наборов пространственных данных и программного обеспечения для получения сведений о ежедневных поездках на работу

29. Одной из производных переменных, распространяемых в материалах переписи, является расстояние от дома до работы. В общих чертах, это расстояние по прямой («полета птицы») от «границы квартала», где находится место проживания, до «границы квартала», где расположено место работы (для имеющих обычное место работы). Эта производная переменная используется Статистическим управлением Канады в распространяемых материалах переписи населения на протяжении десятилетий.

30. Основное признанное ограничение этой производной переменной состоит в том, что очень мало, кто ездит из дома на работу по прямой. И дорожные сети, и особенности рельефа (реки, горы и т. д.) меняют траекторию нашего пути на работу. В результате этот способ измерения в большинстве случаев недооценивает фактическое расстояние между домом и местом работы.

Рис. III

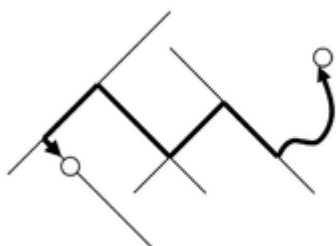
**Прямая линия, или евклидово расстояние**



31. В эпоху «Google Карт» и усовершенствованных наборов геопропространственных данных трудно объяснить продолжающееся использование способа измерения расстояния по прямой. В связи с этим Статистическое управление Канады изучает возможности использования наборов геопропространственных данных и инструментов для расчета наиболее вероятного маршрута следования из дома на работу («расстояние по сети»), с тем чтобы получить гораздо более точное представление о фактическом расстоянии от дома до работы.

Рис. IV

**Расстояние по сети**



32. В настоящее время проводится оценка различных наборов геопространственных данных, включая DMTI Route Logistics, National Road Network File (NRN), открытые данные управлений городского транспорта и интерфейс прикладного программирования «Google Карты» (ИПП). Инструменты, используемые для оценки этих наборов данных, включают ArcGIS, Python, Google Карты ИПП, Open Street Map и Azure Server.

## **2. Использование координат широты и долготы с переписными данными**

33. Идея заключается в том, чтобы вместо координат широты и долготы местоположения, определяемых Статистическим управлением Канады по границам квартала, использовать фактические координаты широты и долготы географического положения человека (жилища).

34. Применительно к поездкам на работу использование фактических координат широты и долготы места жительства и места работы обеспечит более реалистичную оценку расстояния до места работы по сравнению с использованием координат места жительства и работы, определяемых по границам квартала. При определении координат по границе квартала, особенно за пределами городских районов, не учитываются дорожные сети. К примеру, граница квартала проходит вдоль берега реки, а соответствующее здание находится на углу этого квартала и следующего, и в этой точке имеется выезд на дорогу. В этом случае способность компьютерных программ определить наиболее логичный маршрут от А до В может ограничиваться использованием (как в настоящее время) координат широты и долготы, привязанных к точке на берегу реки, а не к дорожной сети.

35. В целом, компьютерным программам сложнее определить наиболее логичный маршрут следования, когда не используются точные координаты широты и долготы.

36. Изучение этого вопроса все еще находится на начальном этапе.

## **3. Ограничения при использовании административных данных о поездках на работу**

37. Существует несколько ограничений при использовании этого нового подхода для расчета расстояния от дома до работы. Первое ограничение связано с нехваткой ресурсов как людских, так и финансовых. Потребуется получить данные, а это подразумевает создание новых рабочих мест и найма на них специалистов.

38. Во-вторых, расчеты маршрутов не 100% точны. Для уменьшения количества ошибок необходимо вводить дополнительные ограничения на расчет маршрутов.

## **D. Места коллективного проживания**

39. Под местом коллективного проживания понимается жилье коммерческого, институционального или коммунального типа. Примерами коллективного жилья могут служить меблированные дома или пансионаты, гостиницы, мотели, туристические заведения, дома престарелых, больницы, общежития для персонала, военные базы, трудовые лагеря, тюрьмы и интернаты.

40. Статистическое управление Канады использует ряд механизмов для снижения нагрузки на респондентов, задействуя существующие административные источники данных, такие как административные записи, для подсчета лиц в местах коллективного проживания. Подобные записи использовались всегда. Однако с 2016 года Статистическое управление Канады изучает новые возможности использования административных данных для сбора информации о местах коллективного проживания, включая базы Вооруженных сил Канады, дома престарелых и исправительные учреждения.



## 1. Базы Вооруженных сил Канады

41. Одна из категорий, в отношении которой изучается возможность использования административных данных, – это лица, проживающие в определенном типе жилых помещений: на базах Вооруженных сил Канады (включая различные здания на территории баз, такие как больницы), в казармах и караульных помещениях.

42. Данные, получаемые Статистическим управлением Канады, были изучены на предмет выяснения возможности осуществления их интеграции для подсчета проживающих на базах лиц. Однако на этапе предварительной работы было решено провести перепись на военных базах с выездом на места, как это делалось до переписи 2016 года, из-за опасений по поводу качества данных. Эти административные данные оказались недостаточно подробными для достижения первоначальной цели, но будут использоваться для разработки показателей заселенности частных жилых помещений в районах проведения переписи по почте в рамках подготовки к переписи 2021 года.

## 2. Государственные опекуны и попечители

43. Статистическое управление Канады также изучает возможность использования сведений, имеющихся у государственных опекунов и попечителей, для дополнения данных по некоторым лицам в местах коллективного проживания, в частности в домах престарелых, интернатах для пожилых и учреждениях с уходом. Информация такого рода будет использоваться для того, чтобы правильно определить местонахождение соответствующих категорий лиц.

44. В настоящее время Статистическое управление Канады узнает о том, что то или иное лицо, возможно, проживает в коллективном жилом помещении, когда в адресе этого лица указано учреждение общественного попечителя. Люди в местах коллективного проживания редко получают свою почту напрямую, и, если кто-то из членов семьи ведет их дела, существует вероятность ошибочного учета пожилых людей в этой семье. Таким образом, эта административная информация может помочь в некоторых случаях правильно определить местонахождение пожилых людей.

45. На данном этапе в Статистическом управлении Канады данные по этому признаку отсутствуют, а доступ к ним требует согласования в каждой провинции, поэтому мероприятия по сбору и использованию этих данных были перенесены на период подготовки к переписи населения 2026 года.

## 3. Информация об исправительных учреждениях

46. Статистическое управление Канады изучает возможность использования имеющихся административных массивов для получения информации о лицах, содержащихся в таких исправительных учреждениях, как федеральные исправительные учреждения, исправительные учреждения провинций/территорий и учреждения для несовершеннолетних правонарушителей. Ведется работа по изучению возможности замены очного обследования этих объектов административными данными в рамках переписи 2021 года. В ходе переписи населения 2016 года было зарегистрировано 813 единиц коллективных жилых помещений такого типа.

## III. Выводы

47. Задействование административных источников позволяет Статистическому управлению Канады повысить качество данных переписи, уменьшить нагрузку на респондентов и расширить содержание переписи. Тем не менее существуют проблемы, которые необходимо преодолеть для успешного использования административных данных в этих целях.

48. Одной из ключевых проблем является получение необходимых данных. Зачастую административные файлы ведутся органами управления иного уровня, с которыми нужно заключать соглашения о передаче информации. Иногда качество административных данных недостаточно высоко для замены данных опроса.

49. Тем не менее Статистическое управление Канады располагает несколькими примерами весьма успешной замены данных, включая данные о доходах и иммиграционную информацию. В рамках переписи населения 2021 года планируется расширить использование регистрационных данных по иммиграции для получения соответствующей информации, однако это будет зависеть от результатов окончательного тестирования и анализа.

50. Источники административных данных также открывают возможности для инноваций, о чем свидетельствуют исследования и разработка новой методологии Статистического управления Канады для получения сведений о поездках на работу. Аналогичным образом исследуются возможности для получения более качественной информации о лицах в местах коллективного проживания. Однако в случае таких коллективных домохозяйств важными факторами являются получение данных и качество административных записей.

---