



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**

Отдельные элементы комплексной переписи в Словакии, влияющие на использование административных источников данных

Записка Статистического управления Словацкой Республики*

Резюме

Перепись 2021 года в Словакии имеет новую структуру. Она представляет собой сочетание данных из нескольких административных источников с данными, полученными на местах. Помимо обеспечения доступа к источникам, которые были определены как актуальные для целей переписи 2021 года, ее подготовительный этап включал в себя также работу с источниками, оценку их качества, анализ ошибок, предложения по исправлению и систему их интеграции. В документе представлены основные выводы и рекомендации с процедурной точки зрения.

* Авторы г-жа Людмила Иванчикова и г-жа Зузана Подманицка.



I. Введение

1. До настоящего времени все переписи населения в Словакии проводились традиционным способом. Запланированная на 2021 год перепись населения и жилищного фонда в Словакии основана на новой концепции, и ее структура включает в себя несколько новых элементов по сравнению с предыдущими переписями. Сама концепция базируется на комбинированном методе переписи. В ней сочетается использование административных источников данных (далее называемых «АИД») и данных о населении, собранных на местах (сплошная перепись).
2. Использование сплошной переписи стало результатом национального требования иметь подробные данные на региональном уровне (муниципалитеты – местные административные единицы (МАЕ), уровень 2), а также требования получать данные на уровне сетки для целей Евростата согласно соответствующему имплементационному регламенту. Аналогичным образом в области распространения данных определенную роль сыграло требование о сопоставимости результатов с предыдущими переписями и о производстве гиперкубов для Европейского союза (ЕС), а также «датакуба» для Статистического управления Словацкой Республики (СР).
3. Одним из ключевых элементов структуры является интеграция статистических данных. Это предполагает объединение данных из нескольких АИД друг с другом, а также совмещение данных из АИД и данных, полученных от населения. Как будет показано в настоящем документе, такая интеграция носит многоуровневый характер.

II. Законодательная основа использования АИД

4. Модель переписи, представленная на национальном уровне в законодательном порядке в 2018 году и впоследствии отраженная в законодательной форме в проекте Закона о переписи населения и жилого фонда 2021 года, является первой комплексной переписью в истории Словакии. Данная перепись освобождает население от обязанности заполнять сложные бумажные формуляры в порядке самоисчисления. Статистическое управление СР будет собирать непосредственно у населения только те данные, которые не могут быть получены из АИД. Население вообще не будет предоставлять данных о домах и жилых помещениях.
5. Хотя использование административных данных базируется на международных рекомендациях относительно переписей, Статистическое управление СР еще не внедрило подобный вид применения административных данных в статистическую практику в сфере социальной статистики. Что касается статистики предприятий, то такая практика уже внедрена, хотя и в ограниченных масштабах. Таким образом, первым шагом стало выявление существующих источников актуальных данных, а также руководителей, ответственных за ведение и содержание реестров, которые можно было бы использовать для целей переписи. В целом в 2016–2018 годах Статистическое управление СР поддерживало связь с 17 руководителями АИД, и первоначально было выявлено 12 источников.
6. Отдельным руководителям АИД были направлены просьбы о сотрудничестве в форме двусторонних соглашений о предоставлении данных. На первом этапе данные проверялись на предмет их пригодности для целей переписи. Несмотря на то что заключение соглашений заняло много времени и явилось результатом ряда переговоров, этот процесс оказался успешным, особенно в плане реестров, актуальных для сбора данных о населении. В связи с предстоящим вступлением в силу Общего регламента по защите данных (ОРЗД), который в Словакии был введен в действие специальным законом о защите персональных данных, некоторые руководители АИД подняли вопрос о законности такой просьбы. Оговорки руководителей АИД высветили необходимость четкого определения ответственности за проведение

переписи и актуальности использования данных для тестирования и проверки и отразили отсутствие соответствующего национального законодательства¹.

7. Решение было найдено путем внесения поправки в Закон № 540/2001 из Сборника законов о государственной статистике, в статье 15 которого говорится, что «перепись населения, домов и жилых помещений представляет собой особый вид статистического обследования, проводимого Статистическим управлением СР, которое является частью государственной статистики, и содержание, объем и порядок осуществления которого регулируются конкретным законодательством. Для целей подготовки и проведения переписи Статистическое управление СР использует также данные из административных источников в соответствии со статьей 13». С помощью этой поправки Статистическое управление СР обеспечило себе доступ ко всем соответствующим АИД еще до начала самой переписи. С методологической точки зрения это позволило осуществить их подробный анализ, разработать и проверить методы оценки качества, провести работу с метаданными, создать систему проверок обработки данных, а также заняться проектированием и тестированием процесса интеграции статистики.

8. Проект Закона о переписи населения и жилого фонда 2021 года предусматривает, что по письменному запросу Статистического управления СР орган государственного управления и лицо, которое является руководителем или оператором информационной системы государственного управления, обязаны предоставлять Статистическому управлению СР данные, информацию, метаданные и показатели, касающиеся населения, включая персональный идентификационный номер, используемый Статистическим управлением СР для интеграции данных из информационных систем государственного управления и других АИД, которыми они управляют или которые функционируют под их контролем.

9. В этом законе также урегулирован вопрос, возникший в связи с тем, что два руководителя АИД, которые по договоренности должны были предоставить данные для проверки в ходе подготовки к переписи, сделали это только один раз, обусловив предоставление дополнительных данных осуществлением Закона о переписи. Кроме того, Закон определяет предоставление АИД в рамках переписи и послеперисного обследования и, что особенно важно, определяет также предоставление данных для такого обследования.

10. В Законе дается перечисление руководителей АИД. В общей сложности будут задействованы девять организаций (Министерство внутренних дел СР, Министерство образования СР, Министерство охраны окружающей среды СР, Управление геодезии, картографии и кадастра СР, Агентство социального страхования, Финансовое управление СР, Служба тюремной и судебной охраны, Главное управление труда, социальных дел и по вопросам семьи и Управление охраны здоровья), которые должны будут обеспечивать, чтобы руководитель соответствующего АИД в установленные сроки предоставил данные АИД до решающего момента переписи Статистическому управлению СР. В нем предусматривается также, например, что руководитель АИД бесплатно предоставляет данные в электронном виде на основании письменного запроса со стороны Статистического управления СР, заранее согласованного с руководителем АИД. Предмет заявки и технические детали предоставления данных из АИД будут определены соглашением, при этом соответствующие требования будут установлены Статистическим управлением СР в виде законодательного постановления.

¹ Перепись населения и жилого фонда является единственным статистическим обследованием в Словакии, проводимым в соответствии с отдельным законом, а не с Законом № 540/2001 из Сборника законов о государственной статистике.

III. Использование административных источников данных

11. Для целей переписи 2021 года будут использоваться следующие административные источники данных, включая реестры:

- a) реестр физических лиц;
- b) реестр адресов;
- c) центральный реестр держателей полисов государственного медицинского страхования;
- d) реестр обвиняемых и осужденных, который ведется Службой тюремной и судебной охраны;
- e) информационные системы государственного управления, находящиеся в ведении Министерства образования СР;
- f) информационная система Агентства социального страхования;
- g) информационные системы Управления геодезии, картографии и кадастра СР;
- h) информационная система Финансового управления СР;
- i) информационная система Главного управления труда, социальных дел и по вопросам семьи;
- j) информационная система Министерства охраны окружающей среды СР.

12. При налаживании отдельных процессов в рамках Оперативного плана проекта переписи 2021 года Статистическое управление СР вполне обоснованно предположило, что подготовительный этап переписи 2021 года займет больше времени, чем при проведении традиционной переписи, и что, помимо ряда исследований, он будет включать в себя анализ АИД.

13. Анализ отдельных источников на подготовительном этапе основывается на единообразном подходе. Поскольку некоторые источники анализируются неоднократно, то процесс их совершенствования также проверяется. Тем не менее следует отметить, что, несмотря на осведомленность руководителей АИД о выявленных несоответствиях и на их согласие с ними, во многих случаях характер самих реестров и АИД (и соответствующего законодательства) не позволяет корректировать наборы данных даже для целей переписи.

14. Анализ охватывает технические характеристики предоставляемых наборов данных, применимость содержания данных, степень полноты отдельных элементов, анализ классификаторов и возможность их преобразования в обязательные статистические классификаторы, анализ отношений и логических связей между переменными, выявление ошибок и возможности для оптимального устранения пробелов или логических ошибок.

15. Подготовленная (готовая к использованию) система проверок, преобразований и логических правок будет верифицироваться, модифицироваться и настраиваться в рамках системы обработки электронной информации.

16. АИД будут использоваться на стадии территориальной подготовки, при переписи домов и жилых помещений, в ходе электронного сбора данных о населении и на этапе обработки.

17. В целях обеспечения территориальной полноты переписи ее процесс основывается на «территориальной подготовке». Территориальная подготовка для целей переписи 2021 года означает пространственную и основанную на базах данных идентификацию всех объектов статистической территориальной единицы, в которой может проживать население СР на момент переписи, независимо от того, имеет ли оно там (какое-либо) зарегистрированное место жительства или нет.

18. Для территориальной подготовки будут использоваться следующие АИД: реестр адресов, находящийся в ведении Министерства внутренних дел СР, информационные системы Управления геодезии, картографии и кадастра (Кадастр недвижимости и ZBGIS®) и информационная система Министерства охраны окружающей среды СР. Результатом территориальной подготовки станет перечень адресов, который будет представлять собой максимальный набор всех возможных населенных и жилых объектов (семейные дома, многоквартирные дома, жилища, садовые домики, школы и т. д.), полученных из АИД, и географический слой адресных точек этих объектов на базовых картах.

Рис. I

Административные источники данных для территориальной подготовки переписи 2021 года



19. Перепись домов и жилых помещений будет проводиться на основе территориальной подготовки к переписи. Как уже упоминалось, предпринимаются усилия по снижению нагрузки на население, связанной с представлением отчетности. В силу этого для целей переписи домов и жилых помещений имеющиеся в распоряжении Статистического управления СР статистические данные из АИД (а именно из кадастра недвижимости) будут использоваться для предварительного заполнения формуляра на каждый жилой объект в соответствии с территориальной подготовкой, а затем муниципалитеты будут дополнять недостающие данные с помощью реестров (в основном муниципальных реестров, реестров управляющих многоквартирных домов и реестров управляющих компаний). Данная процедура избавляет отдельных жителей СР от необходимости участия в сборе данных о характеристиках жилых помещений и домов, поэтому перепись домов и жилых помещений будет передана органами государственного управления муниципалитетам (МАЕ 2).

Рис. II
Электронный сбор данных о домах и жилых помещениях в рамках переписи 2021 года



20. Ответственным за перепись домов и жилых помещений будет лицо, уполномоченное муниципалитетом. Перепись домов и жилых помещений будет проводиться только в электронном виде через веб-портал Статистического управления СР.

21. Сбор данных начнется 1 июня 2020 года, когда через портал Статистического управления СР все муниципалитеты получат доступ к модулю «COLLECTION_Census of Houses and Dwellings_2021» («ZBER_SDB_2021»), содержащему электронные формуляры на дома и жилые помещения, предварительно заполненные в максимально возможной степени. Доступ каждого муниципалитета будет осуществляться с помощью пароля и имени, составленных из статистического кода муниципалитета в соответствии с обязательным классификатором. Имя доступа и пароль будут переданы муниципалитетам по электронной почте в соответствии с пунктом 12 Закона № 305/2013 из Сборника законов. Лица, уполномоченные муниципалитетом, будут редактировать записи о домах и жилых помещениях на основе рекомендованных источников данных. Данные будут собираться до 12 февраля 2021 года; решающим моментом переписи станет полночь с четверга 31 декабря 2020 года на пятницу 1 января 2021 года.

22. В процессе обработки собранных данных база данных будет пополняться расчетной переменной (например, количество людей в жилом помещении) и производными переменными (например, владение домом в разбивке по видам собственности на жилое помещение). Такой статистически обработанный набор данных о домах и жилых помещениях, включая расчетные и производные переменные, будет предоставлен каждому муниципалитету для его кадастровой территории после окончания всей переписи населения, домов и жилых помещений. Это означает, что муниципальные наборы данных будут предоставлены каждому муниципалитету; они будут предоставляться в различных форматах с учетом специфики импорта данных в информационные системы отдельных муниципалитетов. В то же время эта база данных позволит муниципалитетам постоянно обновлять данные и, таким образом, обеспечивать устойчивость решения и его использование для удовлетворения местных потребностей муниципалитетов, а также национальных требований в период после переписи населения 2021 года.

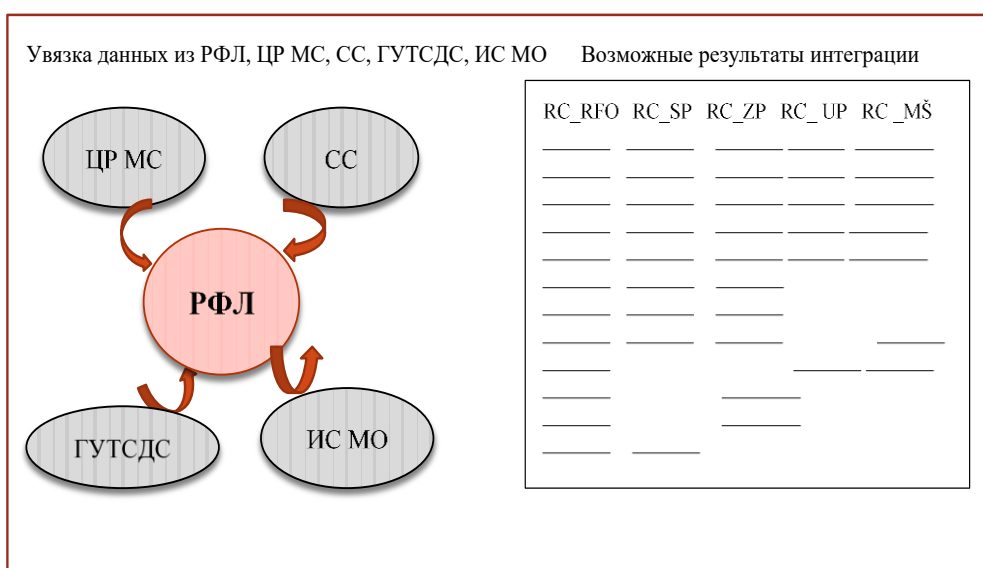
23. Для целей переписи населения и сбора данных о населении данные будут взяты из реестра физических лиц (РФЛ), реестра адресов, центрального реестра держателей полисов государственного медицинского страхования (ЦР МС), реестра обвиняемых и осужденных, информационных систем государственного управления, находящихся в ведении Министерства образования СР (ИС МО), информационной системы Агентства социального страхования (СС), информационной системы Финансового управления СР и информационной системы Главного управления труда, социальных дел и по вопросам семьи (ГУТСДС). Минимальный набор данных, отсутствующих в АИД, будет собираться непосредственно у населения и исключительно с помощью электронных средств.

24. Интеграция данных из АИД в рамках подготовки и обработки данных осуществляется в три этапа:

- a) первый этап – передача и хранение пакетов данных;
- b) второй этап – проверка и первичная обработка данных в конкретном источнике; например, в случае реестра физических лиц (РФЛ), пакет данных которого состоит из нескольких таблиц, они будут заключаться в их привязке через персональный идентификационный номер к полученной базе данных РФЛ. В то же время будут устранены потенциальные источники ошибок, выявленные в ходе первоначального анализа;
- c) третий этап – объединение отдельных АИД друг с другом с использованием уникального идентификатора и интеграционного классификатора.

Рис. III

Пример интеграции АИД о населении



25. Сбор данных у населения будет осуществляться полностью в электронной форме. Перепись по каждому жителю проводится с помощью электронного формуляра (e-form). Этот формуляр существует как в онлайн-овом, так и в оффлайн-овом формате. Оффлайн-овый формуляр будет использоваться только в исключительных обстоятельствах. Что касается характера собранных данных, то при заполнении электронного формуляра предпочтение отдается самоисчислению. Для несовершеннолетнего и для жителя, который не может предоставить данные самостоятельно, электронный формуляр заполняет его законный представитель, близкий родственник или лицо, проживающее в том же домохозяйстве.

26. Сбор данных у населения будет проводиться с 15 февраля 2021 года по 31 марта 2021 года. На всей территории СР решающим моментом переписи (датой, к которой относятся данные) является полночь с четверга 31 декабря 2020 года на пятницу 1 января 2021 года. Такая дата переписи была выбрана в отношении использования реестров и других АИД.

27. Сбор данных будет осуществляться таким образом, чтобы жители могли проводить самоисчисление в течение всего периода переписи (с 15 февраля 2021 года по 31 марта 2021 года). Кроме того, в течение всего периода они могут использовать контактные пункты, где также могут провести самоисчисление, или же это может сделать сотрудник контактного пункта.

28. Интеграция данных АИД и данных, полученных от населения, основана на объединении переменных в четыре группы. Для этой цели выделяются следующие четыре типа переменных:

a) переменные первого типа – переменные, которые поступают в совокупность электронных данных: эта совокупность заполняется ими, и в фоновом режиме проводятся или могут проводиться соответствующие проверки (персональный идентификационный номер, имя и фамилия, дата рождения, пол, адрес);

b) переменные второго типа – переменные, которые заполняются из АИД, но не будут использоваться для совокупности электронных данных (количество детей, некоторые виды экономической деятельности);

c) переменные третьего типа – заполняются исключительно из АИД (возраст, пол, экономическая активность, занятость);

d) переменные четвертого типа – собираются исключительно на местах (религия, гражданство, образование).

29. Набор переменных первого типа, созданный с помощью всех используемых АИД, будет служить отправной точкой для совокупности электронных данных.

30. Интеграция АИД с данными из совокупности электронных данных будет проходить в два этапа. На первом этапе будут использоваться только переменные первого типа: они будут выполнять функцию аутентификации в совокупности электронных данных. На втором этапе, после завершения электронного сбора данных, данные из совокупности электронных данных будут увязаны с обработанными данными из АИД.

31. Интеграция для получения дополнительных данных (например, данных о домохозяйствах или о населении в домах и жилых помещениях) будет обрабатываться статистически после завершения электронного сбора данных о населении.

32. Полученные из АИД данные, которые после переписи будут оценены как высококачественные и актуальные, могут стать отправной точкой для системы, которая будет создана после переписи 2021 года.

IV. Заключение

33. Как подтверждает опыт стран, проводивших комбинированные переписи и переписи на основе реестров, внедрение переписей подобного типа базируется на предпосылках вспомогательного законодательства, общественном одобрении, сотрудничестве с руководителями АИД и другими органами и организациями, участвующими в процессе переписи, существовании соответствующей системы реестров (которая в Словакии отсутствовала и сформировалась только после 2011 года) и наличии уникального персонального идентификационного номера². Исходя из этих предпосылок, в Словакии появилась возможность перехода от традиционной переписи к комбинированной «комплексной» переписи населения.

34. Необходимость анализа новых источников, оценки их качества и методов интеграции, а также использование этих способов и инструментов интеграции влекут за собой изменения в методах работы. В то же время они создают возможность организовать процессы, отвечающие требованиям (хотя они и сформулированы еще в достаточной степени в идеологическом ключе) для периода после переписи (и связанного с этим более частого производства части данных переписи) и для возможного перехода к переписи, полностью основанной на реестрах.

35. Эта концепция имеет как преимущества, так и недостатки. Теряется возможность сбора данных, которые отсутствуют в реестрах, недостаточно качественны или методически не соответствуют задачам переписи. Многие страны

² Schulte Nordholt, Eric, (2018) The usability of administrative data for register-based censuses, Statistical Journal of the IAOS 34:4, pp.487–498.

отказываются от некоторых переменных, единственным источником которых была перепись. Реестры создаются для нестатистических целей. Характер «учета» не всегда отражает нюансы социальных изменений, в частности состав нетипичных домашних хозяйств, как это было в рамках традиционной переписи. Данный аспект потребует более глубокого анализа и, возможно, поиска альтернативных источников, поскольку укрепление выборочных обследований представляется затруднительным. В то же время сокращение участия населения в переписи требует использования источников, которые доступны или могут стать доступны для разработчиков статистики. Их использование обогащает статистику как с точки зрения содержания, так и с точки зрения разработки новых методов. В то же время это создает некоторую зависимость статистических управлений от других субъектов, в частности от руководителей АИД.
