

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

Двадцать шестое совещание

Женева, 2–4 октября 2024 года

Пункт 2 1) предварительной повестки дня

**Пересмотр рекомендаций Конференции европейских
статистиков по проведению переписей населения
и жилищного фонда для раунда 2030 года:
распространение дезагрегированных данных переписи****Разработка рекомендаций по распространению
и смежным вопросам: безопасность, документация,
метаданные и архивирование****Записка Целевой группы Конференции европейских статистиков
по распространению дезагрегированных данных переписи****Резюме*

Настоящий документ содержит проект главы «Распространение информации и смежные вопросы: безопасность, документация, метаданные и архивирование» Рекомендаций Конференции европейских статистиков (КЕС) по проведению переписей населения и жилищного фонда для раунда 2030 года, а также краткое изложение изменений, внесенных по сравнению с Рекомендациями для предыдущего раунда 2020 года. Основная цель документа заключается в получении от

* В состав Целевой группы по распространению дезагрегированных данных переписей Конференции европейских статистиков входят следующие члены: Фабиан Бах (Европейская комиссия — Евростат, председатель), Глен Хольманн (Канада), Дэвид Прайс (Канада), Берни Кейси (Ирландия), Клара Шалев (Израиль), Александра Сильва (Экономическая комиссия Организации Объединенных Наций для Латинской Америки и Карибского бассейна), Бактыбек Кайназаров (Фонд Организации Объединенных Наций в области народонаселения в Кыргызстане), Эндрю Смит (Статистический отдел Организации Объединенных Наций), Ала Негрута (Региональный офис Структуры «ООН-женщины» для стран Европы и Центральной Азии), Лара Кливленд (IPUMS). Высказанные мнения принадлежат исключительно авторам и ни при каких обстоятельствах не могут рассматриваться как изложение официальной позиции Европейской Комиссии, которая является председателем целевой группы, подготовившей настоящий документ.
ПРИМЕЧАНИЕ: Обозначения, используемые в настоящем документе, не означают выражения со стороны Секретариата Организации Объединенных Наций какого бы то ни было мнения относительно правового статуса страны, территории, города или района, или их властей или относительно делимитации их границ.



национальных экспертов по переписям комментариев и рекомендаций по предложенному тексту, чтобы обеспечить отражение в нем потребностей и приоритетов национальных статистических управлений.

I. Введение

1. Каждые десять лет Конференция европейских статистиков (КЕС) публикует Рекомендации, которыми должны руководствоваться страны при проведении переписей населения и жилищного фонда. Эти рекомендации разрабатываются целевыми группами экспертов под эгидой Руководящей группы КЕС по переписям населения и жилищного фонда.
2. Работа этой Целевой группы была организована по трем тематическим направлениям (общие аспекты распространения, статистический контроль для предотвращения идентификации, материалы с географической привязкой). После завершения Обследования раунда переписей 2020 года ЕЭК ООН Целевая группа провела три совещания по телефону и обмены различными письменными материалами. Целевая группа проанализировала ответы на вопросы обследования и разработала предложения по обновлению частей текста Рекомендаций КЕС в рамках своей тематической области.
3. К темам высокого уровня, по которым были предложены обновления, относятся следующие: современные технологии и инструменты распространения информации, включая генераторы таблиц или визуализации данных, новые разработки в области методов статистического контроля для предотвращения идентификации, включая методы, основанные на шумах, и последствия, связанные с сеточными материалами или гибкими инструментами распространения, а также последние разработки в области географии функциональных выходных материалов, включая сетки.
4. В разделе II данного документа кратко изложены изменения, внесенные в содержание, касающиеся распространения, по сравнению с Рекомендациями для предыдущего, 2020 года, раунда. В разделе III дается общий обзор проекта главы.
5. В разделе IV представлен проект главы о распространении (предлагаемой в качестве отдельной главы под названием «Распространение и смежные вопросы: безопасность, документация, метаданные и архивирование») для Рекомендаций КЕС по проведению переписей населения и жилищного фонда для раунда 2030 года.
6. Основная цель документа заключается в получении от национальных экспертов по переписям комментариев и рекомендаций по предложенному проекту текста, чтобы обеспечить отражение в нем потребностей и приоритетов национальных статистических управлений.

II. Краткое описание изменений

7. Данная Целевая группа (ЦГ) пересмотрела следующие разделы Рекомендаций КЕС 2020 года (РКЕС):
 - a) «Конфиденциальность и безопасность», пункты 140–148 в главе I (Методология);
 - b) «Распространение, документация, метаданные и архивирование», пункты 292–318 в главе III (Полевая и другая оперативная деятельность).
8. На основе анализа ответов на вопросы обследования 2023 года ЦГ выделила три основных тематических аспекта, которые необходимо рассмотреть в рамках пересмотра РКЕС:
 - a) *Современные технологии и инструменты распространения:* Обследование подтвердило, что цифровые онлайн-форматы в настоящее время являются наиболее популярными методами распространения информации среди стран Европейской экономической комиссии Организации Объединенных Наций (ЕЭК ООН). Интерактивные онлайн-инструменты, такие как инструменты построения таблиц и картографические инструменты Географической информационной системы (ГИС), являются относительно новыми среди наиболее популярных форматов, в то время как удобные и доступные функции визуализации данных считаются ключевым элементом во многих странах. Это является ключевой областью,

в которой страны ЕЭК ООН продолжают внедрять инновации. ЦГ пересмотрела текст РКЕС в нескольких местах, чтобы сделать эту общую тенденцию более заметной и подчеркнуть ее различные преимущества, такие как степень удовлетворенности пользователей и более эффективное распространение;

b) *Новые разработки в области статистического контроля для предотвращения идентификации (СКПИ)*: Обследование показало, что многие страны ЕЭК ООН внесли существенные обновления в свои системы статистического контроля для предотвращения идентификации (СКПИ), в первую очередь благодаря появлению более эффективных методов, основанных на шуме, которые позволяют, например, избегать подавления ячеек, а также устранять новые риски, которые могут возникнуть при использовании более подробных и разносторонних продуктов распространения. Примечательно, что для значительного числа стран ЕЭК ООН становятся актуальными дополнительные риски, связанные с интерактивными инструментами распространения, напрямую взаимодействующими с микроданными, и риски географической дифференциации (связанные с внедрением сеток). ЦГ пересмотрела текст РКЕС в нескольких местах и предложила несколько новых специальных пунктов, чтобы охватить эти новые изменения и подчеркнуть связь между современными технологиями распространения и новыми рисками идентификации;

c) *Последние разработки в области географии функциональных выходных материалов*: Сбор информации с точной географической привязкой (с помощью координат) о населении и зданиях становится стандартом в регионе ЕЭК ООН, что дает различные преимущества для распространения информации, например свободный/настраиваемый выбор единиц географических выходных материалов. В этом контексте сетки стали ключевым географическим уровнем выходных материалов в регионе ЕЭК ООН. Пользователи получают выгоды, например, благодаря высоко гибким интерактивным картографическим инструментам. ЦГ пересмотрела текст РКЕС в нескольких местах, чтобы учесть эти последние изменения и подчеркнуть основные преимущества с точки зрения сопоставимости между странами и степени удовлетворенности пользователей.

9. Наконец, в рамках общих усилий нескольких Целевых групп по реорганизации различных разделов главы III РКЕС 2020 года в более последовательную структуру, Целевая группа 12 предлагает объединить вышеупомянутые разделы в новую специальную главу под названием «Распространение и смежные вопросы: безопасность, документация, метаданные и архивирование». Заметные сопутствующие выгоды заключаются в том, что конфиденциальность и безопасность становятся более заметными, а касающиеся безопасности и архивирования аспекты доступа к закрытым переписным данным объединяются.

III. Обзор главы

10. Распространение является одним из важнейших аспектов успешного проведения переписи, поскольку оно определяет взаимодействие между всеобъемлющими микроданными, составленными переписными органами, и продуктами данных, с которыми работают пользователи. Таким образом, аспекты распространения охватывают не только конечные продукты, но и то, как эти продукты получают из полных микроданных. Как отмечалось в разделе II, последние разработки в области технологии распространения привели к более сложным взаимозависимостям между безопасностью персональных данных и характеристиками современных продуктов (интерактивность, гибкость, географическая детализация). Поэтому в данной главе основное внимание уделяется распространению и различным важным смежным вопросам, а именно: безопасности, включая СКПИ, документированию процесса переписи и ее результатов, метаданным, сопровождающим продукты переписных данных, и долгосрочному архивированию данных переписи, включая доступ к историческим переписным данным. Ниже приводится краткое описание ее разделов.

11. Раздел «Конфиденциальность и безопасность» состоит из двух подразделов:

а) в разделе «Принципы конфиденциальности» изложены основные принципы, определяющие меры по обеспечению физической и логической безопасности персональных данных, обрабатываемых в рамках переписи. К их числу относятся обеспечение неприкосновенности частной жизни и гарантирование общественности того, что персональная информация, обрабатываемая переписными органами, защищена на всех этапах и используется только для конкретных статистических целей переписи. Для обеспечения такой гарантии необходимо принять достаточные внутренние меры, например, в отношении инфраструктуры информационных технологий (ИТ) и управления доступом сотрудников;

б) Статистический контроль в целях предотвращения идентификации описывает процесс защиты статистических данных таким образом, чтобы они могли публиковаться без разглашения конфиденциальной информации о каком-то конкретном лице или субъекте. Основная цель любой методологии СКПИ состоит в обеспечении достаточного уровня защиты с минимальной утратой информации и, следовательно, максимальным сохранением полезности данных. Для защиты распространения табличных и микроданных предусмотрены несколько мер СКПИ. Подчеркиваются связи между новыми технологиями распространения и дополнительными рисками разглашения информации.

12. Раздел «Распространение» охватывает основные аспекты и рекомендации, касающиеся стратегии распространения и последующих продуктов, а именно:

а) одновременного распространения информации среди всех пользователей в соответствии с графиком, что имеет принципиальное значение;

б) описания различных способов предоставления результатов переписи пользователю, в том числе, например, в виде статичных публикуемых отчетов или таблиц, в виде интерактивных продуктов данных, дополнительно оснащенных средствами визуализации данных или картографическими функциями; в виде заказных или сформированных по заданным параметрам выходных материалов; или в виде микроданных;

в) затратоэффективности стратегии распространения с упором на потенциальное повышение эффективности за счет использования новых технологических решений (например, автоматизация методов СКПИ, интерактивные инструменты, которые могут частично заменить услуги по требованию);

г) удобного и бесплатного доступа к продуктам, включая технические стандарты для взаимодействия с онлайн-продуктами данных и обслуживания современных информационных каналов, таких как социальные сети;

д) взаимосвязей между передовыми технологическими разработками в проведении переписи и более мощными инструментами распространения, такими как технология географических информационных систем (ГИС), облегчающая составление карт и разработку средств географической визуализации.

13. Раздел «Документация и метаданные» включает в себя, в частности, рекомендации, касающиеся:

а) всеобъемлющего набора во вспомогательной документации и метаданных, способствующих объяснению, уточнению и повышению полезности статистических материалов;

б) четко структурированной системы метаданных, опирающейся на международные стандарты и соответствующей конкретным национальным требованиям;

в) стандартных элементов системы метаданных;

г) публикации методологических отчетов, когда методология переписи изменилась по сравнению с предыдущим раундом;

е) консультаций с заинтересованными сторонами из разных сфер для улучшения метаданных.

14. В разделе «Архивирование и доступ к переписным данным», в частности, содержатся рекомендации, касающиеся:

а) переписи как уникального источника данных, что возлагает на переписные органы большую ответственность за сохранение исторической картины общества для будущего;

б) соответствующих целей, контекста и обязанностей по обеспечению безопасности данных в тех случаях, когда переписные органы рассматривают возможность предоставления пользователям доступа к историческим переписным данным по истечении определенного периода (рекомендуется, чтобы он составлял не менее 100 лет);

с) в случае архивирования переписных данных важно обеспечить сохранность и удобный доступ ко всем метаданным и материалам о процедурах/ операциях.

IV. Распространение и смежные вопросы: безопасность, документация, метаданные и архивирование¹

A. Конфиденциальность и безопасность

1. Принципы конфиденциальности

15. В рамках переписей осуществляется сбор информации о каждом лице и каждом домохозяйстве страны. Перепись может проводиться путем прямого сбора данных, путем формирования информации из административных регистров или путем сочетания сбора данных из регистров и обследований. В центре ее внимания находятся не конкретные сведения об отдельных людях, а статистические данные о сообществе и группах внутри сообщества в целом. Исходя из этого, общественность имеет право рассчитывать и должна быть уверена в том, что конфиденциальность персональной информации будет соблюдаться. Имена, адреса и персональные идентификационные номера (ПИН) должны как можно скорее обособляться от других данных в процессе переписи и не разглашаться во избежание появления в выходных данных переписи личных идентификаторов. Требование конфиденциальности, обычно подкрепленное специальными законами о защите статистических и персональных данных, охватывает весь процесс переписи, начиная с обеспечения безопасности любой обработки персональной информации (например, заполненных переписных листов или персональной информации из регистров населения или административных источников данных) и заканчивая защитой информации, содержащейся в выходных материалах и доступной общественности.

16. Общественности должны быть предоставлены гарантии того, что вся персональная информация (например, имена, подробная адресная информация, государственные идентификационные номера или внутренние связующие ключи) будет обрабатываться со строгим соблюдением конфиденциальности переписными органами и всеми лицами, которые работают в переписных органах или оказывают им услуги в целях проведения переписи. Многие страны будут располагать внутренним законодательством, обеспечивающим защиту такой информации, в форме либо конкретного переписного законодательства, либо более общего законодательства,

¹ Перекрестные ссылки, содержащиеся в данной главе, даются в соответствии с нумерацией, используемой в настоящем документе. В случае перекрестных ссылок, выходящих за пределы главы, в квадратных скобках указаны номера пунктов в Рекомендациях 2020 года. В обоих случаях подразумевается, что в окончательном, опубликованном варианте полного текста Рекомендаций 2030 года номера пунктов с перекрестными ссылками будут отличаться.

касающегося статистической конфиденциальности или защиты персональных данных и свободы информации.

17. При работе с информацией, полученной в ходе любой части процесса переписи, следует руководствоваться следующими дополнительными принципами:

а) доступ к любой значимой для целей переписи полученной персональной информации должен предоставляться только лицам, подведомственным переписным органам или действующим от их имени;

б) защита неприкосновенности частной жизни должна быть ключевым требованием, позволяющим обрабатывать значимую для целей переписи персональную информацию, таким образом, чтобы она не раскрывала общественности никаких персональных данных. Кроме того, при проведении опроса члены домохозяйств должны, если они этого желают, иметь возможность предоставлять персональную информацию на отдельном переписном листе таким образом, чтобы она была недоступна другим членам их домохозяйства, или счетчику. В рамках переписей, проводимых с использованием информации из административных регистров, персональная информация должна храниться отдельно от государственных связующих ключей, конфиденциальным образом и не раскрываться при распространении результатов;

в) все члены переписной организации и внешние агенты, оказывающие услуги переписным органам в связи с проведением переписи, должны получить строгие инструкции и подписать правовые обязательства в отношении соблюдения конфиденциальности. Они должны быть предупреждены о том, что в случае любых нарушений будут преследоваться по закону;

г) должна строго обеспечиваться и, в случае необходимости, проверяться независимым образом физическая целостность любых документов или цифровых данных, хранящихся для процесса переписи и содержащих персональную информацию, во время нахождения их у переписных органов, полевого персонала и уполномоченных агентов;

д) любой способ доступа к персональным переписным данным или их обработки, включая ИТ-системы, телефонные приложения и т. д., должен иметь надежную защиту во избежание несанкционированного доступа к информации;

е) при публикации статистических данных, опирающихся на материалы переписи, необходимо принимать достаточные меры во избежание непреднамеренного разглашения информации о поддающихся идентификации лицах и домохозяйствах. Особые меры предосторожности, возможно, следует предпринять в отношении статистических выходных материалов по малым районам.

2. Статистический контроль в целях предотвращения идентификации

18. Статистический контроль в целях предотвращения идентификации (СКПИ) означает процесс защиты статистических данных таким образом, чтобы не допустить разглашения конфиденциальной информации, которая могла бы быть увязана с какими-то конкретными лицами или субъектами. Основная цель любой методологии СКПИ состоит в обеспечении достаточного уровня защиты с минимальной утратой информации и, следовательно, максимальным сохранением полезности данных. В литературе по СКПИ приводится целый ряд стандартных показателей риска и полезности, позволяющих систематически оценивать соотношение риска и полезности и обоснованно выбирать наиболее подходящий вариант СКПИ для программы публикации данных страны.

19. Меры в целях предотвращения идентификации на основе табличных данных могут включать в себя некоторые или все из нижеследующих процедур:

а) ограничение количества категорий на выходе, в которых может быть раскрыта переменная, например использование агрегированных возрастных групп, а не просто значений возраста, особенно для старших возрастов («глобальное перекодирование»);

b) в тех случаях, когда число лиц или домохозяйств на какой-то определенной территории не достигает минимального порогового уровня, удаление итоговых данных (локальное удаление), за исключением, возможно, базовых итоговых показателей, или же слияние их с показателями достаточно крупной территории;

c) добавление «шума» к наборам микроданных перед составлением таблиц («предтабличный шум»), например замена некоторых характеристик единичных записей, подверженных наибольшему риску, путем нахождения совпадения в микроданных на основе набора заранее определенных совпадающих переменных и замены всех или некоторых других переменных между совпадающими записями («целенаправленная замена записей»);

d) добавление «шума» в созданные таблицы («посттабличный шум»), например, путем округления величин в ячейках вверх или вниз до ближайшего кратного predeterminedной базе округления (традиционное округление) или добавление шума ограниченной величины контролируемым и последовательным образом во все таблицы (контролируемое введение «шума», например, с помощью «метода ключа ячеек»).

20. В случае публикации переписных микроданных (таких, как микроданные из заказных файлов или файлов для широкого распространения) важно обеспечить, чтобы вся информация из баз данных, касающаяся имени, адреса и любой уникальной характеристики, позволяющая идентификацию лиц, удалялась из них. Наборы микроданных для научного использования позволяют проводить сложный анализ, выходящий за рамки публикуемых таблиц, визуализаций и показателей. Создание выборки микроданных, обеспечивающей доступ лишь к части всей совокупности, добавляет уровень защиты, сохраняя информацию о характеристиках на уровне совокупности. Дополнительно могут использоваться методы глобальной перекодировки и локального подавления микроданных для снижения риска раскрытия. Для защиты конфиденциальной информации могут также использоваться методы возмущения микроданных или целенаправленно применяться синтетические методы. Методы возмущения и синтетические подходы должны использоваться в пределах допустимого и тщательно проверяться, поскольку они потенциально могут снизить точность и полезность микроданных.

21. Систематическая оценка соотношения риска и полезности может показать, что для конкретной программы публикации данных наиболее подходящим является сочетание нескольких мер защиты. Кроме того, важно последовательно применять любую схему СКПИ во всей статистической продукции, поскольку расхождения обычно влекут за собой дополнительные риски раскрытия информации. Некоторые из основанных на шуме методов, упомянутых в пункте 19, были разработаны специально для преодоления известных недостатков более традиционных методов, например, чрезмерной потери информации и потери согласованности локального подавления в больших наборах таблиц.

22. Интерактивные средства публикации становятся все более популярными (см. п. 26). При их использовании важно, чтобы схема СКПИ была автоматизирована и интегрирована в эти инструменты. Очень гибкие инструменты, которые обслуживают запросы пользователей с широкими возможностями настройки путем прямых запросов к базовым микроданным, влекут за собой особые дополнительные риски раскрытия информации, например, начиная с массовых и систематических атак путем запросов с использованием сценариев. В таких случаях обычно требуются специальные схемы СКПИ.

23. Высокая географическая детализация является одной из уникальных особенностей выходных материалов переписи во многих странах, что находит отражение в растущей тенденции публикации большего объема данных по мельчайшим географическим единицам, которые в настоящее время часто дополняются сеточными продуктами (см. пп. [428–433]). Однако предоставление данных по очень малым и не пересекающимся географическим единицам (например, административным границам или границам сетки) влечет за собой особые дополнительные риски раскрытия информации («географическая дифференциация»),

которые должны быть учтены в схеме СКПИ. Например, в ходе подготовки к предыдущей переписи в ряде стран были признаны эффективными методы посттабличного добавления шума.

24. Независимо от конкретной принятой схемы СКПИ, важно объяснить общие свойства схемы, чтобы пользователи были в курсе. Это особенно актуально для методов СКПИ, при которых публикуемые данные выглядят как необработанные, или для методов, которые могут привести к ограниченной потере внутренней аддитивности таблиц.

В. Распространение

25. Перепись не считается завершенной, пока собранная информация не будет представлена пользователям в той форме и в те сроки, которые отвечают их постоянно меняющимся потребностям. Таким образом, при распространении результатов переписи повышенное внимание следует уделять оперативному реагированию на потребности пользователей и высоким стандартам качества при подготовке статистических данных. Результаты переписи должны распространяться одновременно среди всех пользователей, и самое тщательное внимание следует уделять недопущению непреднамеренного разглашения информации, позволяющей идентифицировать конкретных лиц. Для защиты конфиденциальности должны применяться различные статистические приемы (см. раздел о конфиденциальности и безопасности, пункты 18–24).

26. Существует ряд традиционных способов распространения итогов переписи среди пользователей:

а) в виде публикуемых сводок (в печатном виде либо, что встречается чаще, на цифровых носителях), содержащих стандартные и заранее согласованные таблицы, обычно общенационального уровня, уровня регионов или уровня местных территориальных единиц, которые могут быть получены от правительственных агентств или напрямую из других источников;

б) в качестве специальных непубликуемых сводок/выдержек, содержащих стандартные данные переписи, но дезагрегированные по другим географическим границам и/или подгруппам, не публиковавшимся ранее (заказчику может быть предложено покрыть расходы на подготовку);

с) в виде продуктов данных, доступных в режиме онлайн через веб-сайты Национального статистического управления (НСУ) или другие электронные каналы. Эти продукты данных могут варьироваться от агрегированных до микробаз данных, доступных для онлайн-обработки или бесплатного скачивания, факультативно оснащенных динамическими или интерактивными средствами визуализации данных, включая картографические функции для повышения ценности статистики;

д) в виде заказных или специально подготовленных материалов, разрабатываемых на основе базы данных, автоматизированного строителя таблиц или сервиса статистических данных, содержащих специальные перекрестные таблицы переменных, отсутствующие в стандартных сводках или выдержках, которые должны соответствовать тем же требованиям статистического контроля для предотвращения идентификации, которые применимы к стандартным результатам; и

е) в виде микроданных (часто называемых общедоступными выборками данных), обычно доступных только в виде выборочной части всей совокупности (часто отбираемой на уровне домохозяйства, чтобы включить отдельных лиц из домохозяйства) с применением методов СКПИ (см. пункты 18–24). Чтобы сбалансировать повышенный риск предоставления микроданных, может применяться комбинация дополнительных мер контроля: потенциальные пользователи данных часто проходят проверку; данные предоставляются только в ограниченном формате; и предоставление данных или доступ к ним часто осуществляются в безопасных и строго контролируемых условиях, при которых реализуются тщательные меры по защите конфиденциальности данных.

27. Если настраиваемые средства распространения отсутствуют или их недостаточно для предоставления конкретных таблиц, необходимых лишь нескольким пользователям, таким как некоторые правительственные учреждения или специализированные исследовательские организации, они могут быть предоставлены по запросу. Однако если такие таблицы были сформированы, то не должно быть ограничений на их доступность для широкой общественности через группы поддержки клиентов и пользователей НСУ.

28. Затратоэффективность печатания становится проблемой, поскольку физические копии не могут охватить такое же количество пользователей, что и публикации в Интернете. Роль традиционных изданий, особенно в печатной форме, близка к полному устареванию. Хотя печатные публикации могут содержать согласованные и последовательные комментарии по отдельным темам и быть тем самым полезными для определенных пользователей или рынков, однако, как правило, сегодня пользователи, обычно, ожидают распространения в интерактивных и динамичных цифровых форматах.

29. Публикация данных в электронном формате в режиме онлайн должны предоставлять пользователям удобные средства поиска данных в стандартных форматах или через интерфейсы программирования приложений (API), в идеале дополненные удобными интерактивными функциями для настройки выходных продуктов. Пригодность для немедленного использования таблиц данных в режиме онлайн — важнейшая характеристика, обеспечивающая возможность поиска и релевантность переписной информации. Международные стандарты метаданных, такие как стандарт обмена статистическими данными и метаданными (SDMX), должны считаться приоритетными для форматирования базы выходных данных. Стратегии распространения должны быть также приведены в соответствие с государственной политикой в области открытых данных.

30. По мере возможности, должны разрабатываться онлайн-инструменты для заказа, спецификации, настройки и получения переписных таблиц и общедоступных выборок (микроданных), обеспечивающие соответствующие меры по защите статистической конфиденциальности и безопасности передачи данных. При разработке форматов представления результатов переписи следует учитывать все новые технологии, популярные среди пользователей, такие как смартфоны и другие портативные устройства. Также важно учитывать гибкость контента, который будет нужен пользователям. Тенденции быстро меняются, и для обеспечения релевантности переписной информации крайне важно иметь полный набор таблиц или продуктов, позволяющих быстро обеспечить доступ к данным. Онлайн-генераторы таблиц позволяют пользователям быстро настраивать свои запросы и напрямую запрашивать базу данных для получения немедленных результатов. Это уменьшит необходимость размещения в Интернете массивной таблицы данных, в которой пользователи будут пытаться найти свои данные, если они были опубликованы.

31. Социальные сети становятся все более популярным и эффективным средством распространения небольших объемов выходных данных, особенно среди пользователей-неспециалистов, или для своевременного распространения результатов, значимых для каких-то событий или важных дат. Учитывая разнообразие доступных в настоящее время платформ, текстовые изображения и инфографика могут быть использованы в кампаниях в социальных сетях, чтобы подчеркнуть удобство использования выходных материалов переписи. Использование социальных сетей будет во многих случаях служить демонстрацией готовности НСУ к выстраиванию и укреплению диалога с пользователями, с тем чтобы с большей готовностью реагировать на их вопросы и озабоченности. см. также пункты [297].

32. Хотя онлайн-доступ к таким базам микро- и макроданных и их распространение в сети с помощью компьютерных носителей могут в значительной мере способствовать расширению круга пользователей и, таким образом, повышению спроса на переписные данные, необходимо сделать два предостережения:

а) определенные перекрестные таблицы могут страдать значительными проблемами качества вследствие ошибок, обусловленных непредоставлением ответов,

формированием выборки или обработкой данных или недостатками процедур обработки или импутации. Переписным органам следует обеспечить процедуры предупреждения потенциальных пользователей о таких проблемах, чтобы способствовать сохранению доверия к переписи в целом. НСУ запрещают публикацию определенных перекрестных таблиц по причинам, связанным с качеством их содержания, хотя такая политика зачастую может оттолкнуть пользователей. Другие НСУ публикуют такие перекрестные таблицы только при наличии четко сформулированной политики, основанной как на технических, так и на содержательных соображениях;

b) некоторые подробные перекрестные таблицы и все файлы с данными об отдельных лицах создают риски разглашения информации о поддающихся идентификации лица, что является нарушением правил о конфиденциальности переписных данных. Этот вопрос более подробно обсуждается в пунктах 15–24.

Необходимо учитывать обе эти проблемы — качество содержания и конфиденциальность — и применять соответствующие меры. С другой стороны, ни одна из этих проблем не создает существенных препятствий распространению широкого ряда материалов переписи.

33. Для удовлетворения эволюционирующих потребностей пользователей следует использовать разнообразные продукты. Скорее всего, понадобятся:

a) сводки общенационального, регионального и местного уровней;

b) отчеты о ключевых результатах по конкретным признакам, дополняемые подробными результатами и анализом либо в стандартной форме вплоть до местного географического уровня, либо в виде более подробных дезагрегированных статистических данных по конкретным темам и представляющим интерес совокупностям;

c) анализ демографических тенденций или ключевые обзорные данные по малым районам и малым группам населения;

d) пространственный и графический анализ, включая атлас переписи, возможно, дополненный интерактивным графическим анализом или картографическими инструментами, позволяющими проводить анализ по спецификациям пользователя;

e) продукты дополнительной полезности, такие как классификации по районам и/или домохозяйствам; и

f) дополнительные метаданные, охватывающие определения, классификации, оценки охвата и качества данных. Метаданные могут быть различными в зависимости от вида переписи — прямой, комбинированной и регистровой, но они должны быть одинаково подробными для всех способов проведения.

34. Распространение результатов переписи должно осуществляться в соответствии с заранее разработанным структурированным планом, включающим заранее объявленный график. Это подразумевает оперативное опубликование общенациональных данных в соответствии с графиком, а затем последующее опубликование дезагрегированных данных после всесторонней обработки, все в соответствии с установленным планом. Тем не менее своевременность является важной проблемой на этапе публикации, и необходимо прилагать усилия для сокращения периода публикации, обеспечивая при этом целостность данных.

35. Первоначальную публикацию данных о численности населения обычно с нетерпением ожидают все пользователи — от широкой общественности до администраторов программ и стратегий. Поэтому в некоторых странах промежуточные результаты публикуются вскоре после завершения переписи. Хотя после завершения обработки и проверки всех данных эти цифры уточняются, предварительные данные дают общую картину демографических тенденций. Тем не менее пользователей данных необходимо предупреждать о возможных последствиях

использования промежуточных данных о численности населения, которые могут значительно отличаться от окончательных данных, прошедших валидацию.

36. Чтобы поддерживать интерес общественности в отношении переписи, необходимо в самом начале процесса публиковать график и описание предстоящих выпусков окончательных результатов и материалов переписи. Выпуски могут быть организованы по нарастающей: от простых кратких описательных сводок, охватывающих основные административно-территориальные единицы страны, вначале и до более подробных перекрестных таблиц и описательных тематических и аналитических докладов на более поздней стадии.

37. В целом данные должны предоставляться бесплатно в точке доступа или доставки, но размер платы в тех случаях, когда она необходима (например, в случае подготовки материалов по специальным заказам/спецификациям), должен устанавливаться таким образом, чтобы результаты оставались доступными для всех типов пользователей. Гибкие интерактивные инструменты публикации, которые автоматически обрабатывают индивидуальные запросы о предоставлении данных (см. пункт 26), могут помочь эффективно удовлетворять индивидуальные потребности пользователей и тем самым повысить ценность бесплатных выпусков данных. Возможности распространения информации в режиме онлайн позволили сократить потребность в обширных бумажных материалах. Тем не менее НСУ могут предоставлять платные услуги по печати по запросу, чтобы обеспечить материалами переписи пользователей, которые не могут иметь доступ к цифровым копиям или получить их. Такие пользователи не должны оказываться в ущемленном положении из-за отсутствия результатов итоговых материалов на бумажных носителях.

38. Переписные материалы должны разрабатываться таким образом, чтобы позволять распространение статистической и географической информации вместе с ГИС и/или использование других интерфейсов программирования приложений, с тем чтобы соответствовать широким запросам пользователей и обеспечивать максимальную гибкость и взаимную совместимость, по возможности, соразмерно с необходимыми гарантиями конфиденциальности. Ниже перечислены некоторые желательные свойства продуктов:

а) пользователи должны иметь возможность быстро и беспрепятственно находить информацию и в открытом или множественных форматах;

б) НСУ значительно повысят полезность своих данных переписи, если встроит в продукты соответствующие графические и картографические возможности. В идеале пользователи должны иметь возможность самостоятельно составлять графики и/или карты без каких-либо затруднений, затем распечатывать или трассировать их либо использовать эти изображения для других целей. Ряд стран в настоящее время производят такие переписные материалы, иногда в сотрудничестве с коммерческими организациями;

в) приоритет следует отдавать открытым форматам данных, в частности тем, которые облегчают обмен, позволяя «смешивать» переписную информацию с другими базами данных, что даст еще больше возможностей для более широкого использования данных.

39. Тематическое картирование и визуализация данных стали важным элементом распространения итоговых материалов НСУ по причине их способности обеспечить взаимодействие с пользователями и расширить охват распространения переписных данных. Вместе с тем визуализация данных является достаточно широкой областью, контент и структура которой могут варьироваться от простой инфографики до сложных инструментов многомерного анализа данных. Кроме того, визуализация данных может создавать трудности для некоторых переписных органов из-за высокой стоимости, требований к инфраструктуре и соответствия техническим требованиям отдельных стран (например, доступность, единое оформление). Примерами могут служить нехватка экспертных навыков для эффективной визуализации, а также проблемы выделения достаточных ресурсов для ее разработки, особенно с учетом бюджетных ограничений, с которыми сталкиваются многие переписные органы в десятилетний период. Тем не менее сегодня пользователи все чаще ожидают, что

веб-контент будет визуальным, привлекательным и индивидуальным, исходя из чего создание потенциала визуализации данных должно стать одним из ключевых приоритетов в списке ожиданий многих НСУ.

40. НСУ должны попытаться найти баланс между подготовкой более традиционных табличных данных и более высокими затратами на создание визуализаций данных. Это требует глубокого понимания сегментов пользователей и их конкретных потребностей. Необходимо провести исследование, чтобы лучше понять потребности конечных пользователей в данных (подробнее о сегментации пользователей см. главу [x] о коммуникации и информационно-просветительской деятельности). Универсального решения не существует, и многие НСУ учитывают это, создавая специальные подразделения или группы, отвечающие за технологию визуализации данных.

С. Документация и метаданные

41. Важной составной частью программы распространения результатов переписи каждой страны является всеобъемлющий набор справочной документации и метаданных, помогающих разъяснить, уточнить и дополнить статистические данные, особенно в том, что касается сопоставления с данными предыдущих переписей и другими источниками данных.

42. Система метаданных служит источником дополнительной информации о характеристиках регистрируемых и публикуемых данных. Каждое НСУ будет использовать свою собственную систему метаданных, опирающуюся на международные стандарты и одновременно соответствующую конкретным национальным требованиям. Поскольку перепись населения и ее результаты связаны с другими областями статистической деятельности, рекомендуется, чтобы в каждой стране система метаданных переписи населения использовала те же элементы, что и общая система метаданных конкретного НСУ. Однако также необходимо, чтобы система метаданных переписи населения содержала некоторые элементы, которые используются только в целях данной переписи. Кроме того, система метаданных переписи населения должна обеспечивать по возможности максимальную международную сопоставимость данных.

43. Раунд переписей населения 2030 года должен обеспечить сопоставимость с данными предыдущих переписей населения, но в то же время содержать новые элементы, учитывающие изменения, произошедшие с момента проведения предыдущей переписи. Таким образом, система метаданных должна быть сопоставима с системой предыдущей переписи, но должна быть обновлена в соответствии с вновь возникшими потребностями. Системы метаданных индивидуальных НСУ также должны отражать степень использования данных прямой переписи и/или административных источников данных.

44. Система метаданных должна как минимум включать в себя:

- a) определения используемых терминов и концепций;
- b) словарь данных или глоссарий терминов;
- c) пояснительные примечания к таблицам;
- d) классификации и номенклатуры;
- e) вопросы переписи (в том случае, когда информация собирается традиционным опросным методом); и
- f) цели сбора информации, особенно в случае использования административных данных; и
- g) используемые источники данных, особенно когда данные получают из административных регистров.

45. В случае показателей, по которым были созданы международные стандартные классификации, должны использоваться именно эти классификации. В отношении показателей, которые не могут быть классифицированы с использованием международных стандартов, может потребоваться разработка новых номенклатур. Вспомогательная документация может охватывать широкий ряд конкретных вопросов, таких как базовая методология, охват, предоставление ответов, источники данных, пробные переписи и тесты, производные переменные, предоставление ответов через Интернет, импутация, слепопереписные обследования, а также доклады, содержащие более общее описание всего процесса переписи в целом и качества данных. Набору документации и конкретным вопросам охвата и качества данных посвящена глава [IV], однако рекомендуется, чтобы метаданные переписи включали в себя как минимум конкретные показатели качества и охвата, такие как коэффициенты предоставления ответов (на национальном и местном уровнях) и масштаб импутации данных (по источнику данных в целом и по отдельным признакам).

46. Методологические доклады играют особенно важную роль в том случае, когда по сравнению с предыдущей переписью была изменена базовая методика (например, переход от традиционного опросного метода к полному или частичному использованию регистрового метода). Такие изменения, как представляется, окажут влияние на используемые определения и понятия и, следовательно, на сопоставимость результатов разных раундов переписи.

47. Чтобы получить полные и всеобъемлющие метаданные, переписные органы должны провести консультации с заинтересованными сторонами из разных сфер, включая государственные учреждения, исследователей, политиков и общественность, используя, например, структурированные опросы, фокус-группы или другие механизмы консультаций (см. пункты [255–265]). Отзывы могут собираться для уточнения описаний метаданных, что сделает их более интуитивными и всеобъемлющими.

D. Архивирование и доступ к закрытым переписным данным

48. Перепись является особым источником статистических данных, который может обеспечивать наличие непрерывной и сопоставимой информации, покрывающей период длительностью до 100–150 лет. Это дает возможность создать уникальный источник данных, но возлагает на НСУ (или соответствующее национальное архивное учреждение) большую ответственность за сохранение этой особой исторической картины общества для будущего. Таким образом, данные переписи представляют ценность не только для лиц, принимающих решения, и пользователей сегодня, но также и для будущих поколений. НСУ отвечают за обработку, архивирование и хранение этого «сокровища».

49. Многие страны хранят переписную информацию об отдельных лицах и домохозяйствах только до тех пор, пока она необходима для обработки данных и подготовки статистических результатов, или до окончания переписи. Вместе с тем, когда речь идет об оценке общих затрат и выгод переписи, нельзя недооценивать научную, социально-историческую и генеалогическую ценность индивидуальных записей. Поэтому НСУ могут захотеть разрешить (полный или ограниченный) доступ к полному набору переписных записей по истечении периода, в течение которого они должны оставаться закрытыми. Страны, которые намереваются хранить эти данные для таких исследований, должны обеспечить наличие надежной правовой основы и физической инфраструктуры для защиты безопасности и конфиденциальности таких данных до момента их открытия для общественности. Национальные правительства должны отдавать себе отчет в том, что способность НСУ собирать информацию среди широкой общественности может быть серьезно подорвана, если не будет обеспечено соблюдение гарантий, обещанных в отношении конфиденциальности собираемой информации. Доверие общественности к безопасности и конфиденциальности информации, обрабатываемой в целях переписи, должно рассматриваться в качестве имеющего первостепенное значение.

50. Период неразглашения переписных данных должен быть достаточно длительным либо для того, чтобы обеспечить защиту конфиденциальности информации, особенно чувствительной информации, о еще живущих людях, либо для того, чтобы, во всяком случае, минимизировать риск нарушения такой конфиденциальности. Продолжительность периода, в течение которого данные остаются закрытыми, во многих странах конкретно определяется законом, но может быть различной в зависимости от страны. В других странах могут существовать более общие положения в рамках законодательства о защите данных и/или о свободе информации, предусматривающие сохранение конфиденциальных данных в качестве закрытых до тех пор, пока не исчезнет риск раскрытия персональной информации о живущих лицах. Исходя из этого рекомендуется использовать 100-летний период, хотя, учитывая постоянный рост продолжительности жизни, в зависимости от национальных особенностей страны, возможно, пожелают изучить возможность увеличения этого порога.

51. Помимо запросов о предоставлении открытого доступа, НСУ могут получать запросы от других государственных учреждений о предоставлении доступа к переписным записям с целью проверки или подтверждения существующей информации, если исторические данные скудны или отсутствуют. Вопрос о предоставлении доступа к этим записям должен рассматриваться, если это требуется по закону или если это явно служит общественному благу.

52. В дополнение к архивированию переписных данных (для тех стран, которые это делают) для всех стран не менее — а то и более — важно также обеспечивать сохранение всех метаданных и материалов о процедурах/операциях, включая всю документацию по управлению проектами, созданную в ходе процесса проведения переписей, и удобный доступ к ним. Это не только обеспечивает наличие материалов для аудита в целях оценки успешности и эффективности переписи, но также позволяет при планировании будущих переписей извлекать уроки из успехов и неудач предшественников.

53. В связи с вопросами архивирования переписным органам следует также учитывать быстрое развитие технологий и регулярно пересматривать носители и системы, которые используются для хранения этой ценной информации, чтобы обеспечить ее быстрое извлечение в случае необходимости в течение следующих даже 20–50 лет.

V. Заключение

54. Проект рекомендаций по распространению и смежным вопросам: безопасность, документация, метаданные и архивирование для раунда переписей населения и жилищного фонда 2030 года представляется для комментариев и обсуждения.
