

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

Двадцать пятое совещание

Женева, 20–22 сентября 2023 года

Пункт 4 l) предварительной повестки дня

**Пересмотр рекомендаций Конференции европейских
статистиков по проведению переписей населения
и жилищного фонда для раунда 2030 года:****распространение дезагрегированных данных переписи****Предварительный доклад о ходе работы Целевой группы
Конференции европейских статистиков
по распространению дезагрегированных данных
переписи****Записка Целевой группы Конференции европейских статистиков
по распространению дезагрегированных данных переписи*¹***Резюме*

После начала процесса пересмотра Рекомендаций КЕС данная Целевая группа провела три совещания по телефону и обмены различными письменными материалами. Работа организована по трем направлениям (общее распространение, статистический контроль для предотвращения идентификации, материалы с географической привязкой). Целевая группа рассмотрела вопросник обследования 2013 года и разработала предложения по обновлению вопросов и включению новых вопросов в рамках своей тематики.

* Автор: Фабиан Бах, Европейская комиссия — Евростат. Высказанные мнения принадлежат исключительно автору и ни при каких обстоятельствах не могут рассматриваться как изложение официальной позиции Европейской Комиссии.

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.

¹ Настоящий документ был представлен с опозданием в связи с задержкой представления документа Целевой группой.



Среди тем, которые лучше освещаются в обновленных вопросах, можно отметить следующие: современные технологии и инструменты распространения информации, включая инструменты построения 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 информации, а также последние разработки в области географии функциональных выходных материалов, включая сетки.

I. Введение: сфера деятельности и план работы Целевой группы

1. Основной частью Рекомендаций КЕС 2020 (РКЕС)², пересмотром которой будет заниматься данная Целевая группа (ЦГ), является раздел «Распространение», пункты 292–307 раздела «Распространение, документация, метаданные и архивирование» Главы III «Полевая и другая оперативная деятельность». Это также имеет потенциальные последствия для раздела «Конфиденциальность и безопасность», пункты 140–148 главы I «Методология».

2. Круг ведения ЦГ включает в себя:

- a) форматы и продукты распространения с указанием и изложением соответствующих обоснований, целевых аудиторий и конкретных соображений;
- b) обеспечение качества при распространении информации на высоком уровне детализации;
- c) последствия различных подходов к распространению данных переписи населения для контроля конфиденциальности и предотвращения идентификации;
- d) планирование сроков распространения;
- e) распространение материалов с географической привязкой (Целевой группе необходимо будет координировать эту часть своего пересмотра с Целевой группой, занимающейся геопространственной информацией);
- f) документация и метаданные;
- g) архивирование.

3. В состав ЦГ входят девять членов от Канады, Евростата (председатель), проекта «Международные ряды интегрированных микроданных общего пользования» (МРИМОП), Ирландии, Израиля, ЭКЛАК ООН, ЮНФПА Кыргызстана, СОООН и Регионального представительства структуры «ООН-Женщины» для Европы и Центральной Азии.

4. ЦГ согласовала план работы, разделив свою деятельность на два больших этапа: 1) пересмотр соответствующих вопросов вопросника обследования 2013 года (к апрелю 2023 года); 2) пересмотр упомянутых частей РКЕС на основе анализа результатов обследования 2023 года (т.е. начиная с осени 2023 года после возвращения вопросников). На обоих этапах члены ЦГ объединяются в три тематические подгруппы для формулирования первоначальных идей, которые будут обсуждаться всей группой:

- a) общие аспекты распространения, метаданных и архивирования (под руководством Ирландии);
- b) статистическое предотвращение идентификации (под руководством Евростата);

² https://unece.org/DAM/stats/publications/2015/ECECES41_RU.pdf.

с) материалы с географической привязкой (под руководством Канады).

5. В данном документе сообщается об успешно завершеном первом этапе работы, а именно о пересмотре вопросника. ЦГ предложила Руководящей группе обновить вопросы и включить новые вопросы в следующие разделы вопросника 2013 года: методология; технология; коммуникационная и информационно-разъяснительная деятельность; документация, метаданные и архивирование; конфиденциальность и статистический контроль для предотвращения идентификации; распространение; инновации; качество и охват; географические характеристики. Далее ЦГ представила комментарии и соображения (не имея возможности предложить обновления к вопросам) по некоторым элементам разделов: проблемы, успехи и извлеченные уроки; затраты и выгоды.

II. Обмен опытом распространения информации между членами ЦГ

6. На втором совещании все члены ЦГ имели возможность представить конкретные продукты или аспекты распространения из собственного опыта. Выступили семь членов ЦГ со следующими сообщениями:

а) Eurostat рассказал об основных каналах распространения результатов переписи населения ЕС 2021 года, а именно: Census Hub³ для всеобъемлющего доступа ко всем результатам через интерфейс построения таблиц, Eurobase⁴ для предоставления выбора гиперкубов через браузер данных и статьи Statistics Explained⁵. Наконец, материалы также включают ключевые показатели на европейской сетке с квадратом 1 км (общая численность населения уже опубликована⁶);

б) Ирландия представила обзор продуктов распространения данных национальной переписи населения, включающих таблицы PX Stat (более 1000 таблиц до 4 измерений)⁷, электронные публикации⁸, статистику населения малых территорий, включая картографирование переписи и интерактивные карты⁹, исследовательские микрофайлы¹⁰ и сетки;

в) ЭКЛАК ООН продемонстрировала свою региональную платформу распространения данных REDATAM¹¹, сделав акцент на двух продуктах в контексте переписи населения: гармонизированных микроданных переписи и матрицах миграции для латиноамериканского региона;

г) Канада рассказала о своих продуктах для распространения переписных данных¹², включая специальные переписные профили¹³ и карты¹⁴, а также о некоторых планах на будущее, в том числе о новом гибридном онлайн-генераторе таблиц;

³ <https://ec.europa.eu/CensusHub2> (пока только итоги переписи населения ЕС 2011 года; итоги переписи 2021 года будут добавляться по мере поступления).

⁴ <https://ec.europa.eu/eurostat/data/database> (табулятор переписи 2021 года содержит некоторые наборы данных, добровольно предоставленные членами ЕС, а именно: население по широким возрастным группам и регионам уровня КТЕС 3, а также различные специальные наборы данных о населении с украинским гражданством — в контексте неспровоцированного вторжения России на территорию Украины).

⁵ <https://ec.europa.eu/eurostat/statistics-explained/index.php?search=census>.

⁶ https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_and_housing_census_2021_-_population_grids.

⁷ <https://data.cso.ie>.

⁸ <https://www.cso.ie/en/census>.

⁹ <https://visual.cso.ie/?body=entity/ima/cop/2016&boundary=C03736V04484>.

¹⁰ <https://www.cso.ie/en/census/census2016reports/powscar>.

¹¹ <https://www.cepal.org/en/subtopics/redatam>.

¹² <https://www12.statcan.gc.ca/census-recensement/index-eng.cfm>.

¹³ <https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E>.

¹⁴ <https://www12.statcan.gc.ca/census-recensement/2021/geo/maps-cartes/index-eng.cfm>.

е) МРИМОП¹⁵ рассказал о своей деятельности и продуктах, сосредоточив внимание на сопоставимых в международном масштабе выпусках микроданных переписей и обследований из всех регионов мира;

ф) Израиль рассказал о своей национальной ситуации и текущих задачах, в основном связанных с глубокой модернизацией всей системы распространения информации переписи населения. Одна из самых серьезных проблем заключается в том, как обеспечить безопасную публикацию данных;

г) страновой офис ЮНФПА в Кыргызстане рассказал о своей прошлой и текущей региональной деятельности, включая поддержку публикаций по переписи населения в Кыргызстане (например, начиная с 2010 года) и текущие усилия по более полному охвату вопросов Вашингтонской группы¹⁶ по статистике инвалидности и миграции.

III. Пересмотр вопросника

7. Основные соображения и вытекающие из них обновления вопросника, предложенные ЦГ, представлены ниже в соответствии с тремя тематическими направлениями, описанными в пункте 4.

A. Общие аспекты распространения, метаданных и архивирования

8. Со времени последнего обновления РКЕС ЦГ отметила общую тенденцию к использованию более интерактивных веб-инструментов распространения информации. Например, в ряде стран в качестве инструмента распространения информации были внедрены интерактивные генераторы таблиц, позволяющие пользователям создавать по своим настройкам перекрестные таблицы. На этапе окончательной компиляции эти инструменты могут быть связаны либо с микроданными (например, Australian Table Builder¹⁷ и Bangladesh Bureau of Statistics¹⁸), либо с заранее определенным набором очень подробных перекрестных таблиц (например, EU Census Hub, см. сноску 3). Несмотря на то, что выбор постпроцессорной части и некоторых других вариантов дизайна может иметь последствия для сценариев статистического контроля для предотвращения идентификации (см. пункт 16), ЦГ в целом согласилась с тем, что интерактивный и удобный для пользователя внешний интерфейс таких инструментов являются важным улучшением в распространении данных переписи. Часто такие инструменты комбинируются с интерактивными средствами визуализации данных.

9. ЦГ также отметила, что способы сбора и хранения переписных данных продолжают быстро развиваться по мере цифровизации общества во многих странах. С одной стороны, многие страны сокращают прямые контакты с населением, все больше полагаясь на административные и другие источники данных. С другой стороны, переписи, полностью или частично основанные на прямом опросе, все чаще проводятся с использованием интеллектуальных и цифровых инструментов.

10. Наконец, ЦГ отметила, что общение с сообществом пользователей становится все более важным аспектом успешного проведения переписи. Это относится как к периоду до проведения переписи (например, консультации с заинтересованными сторонами для планирования переписи), так и к периоду после публикации результатов (например, типы документации, метаданные и качественная информация).

11. В связи с этим ЦГ предложила внести в вопросник различные изменения, чтобы улучшить получение информации по этим направлениям. В частности, несколько вопросов были расширены, а также добавлены новые, чтобы лучше отразить,

¹⁵ www.ipums.org.

¹⁶ <https://www.washingtongroup-disability.com>.

¹⁷ <https://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder>.

¹⁸ <http://redatam.bbs.gov.bd/redbin/RpWebEngine.exe/Portal>.

предоставляются ли генераторы таблиц в рамках распространения; был добавлен вопрос о том, выделяет ли переписной орган кадровые ресурсы специально для технологий визуализации данных; вопросы были расширены, а также добавлены новые с целью адаптации к эволюционирующим способам проведения переписи (использование административных данных, интеллектуальных средств предоставления ответов на вопросы обследований); и были предложены вопросы, касающиеся консультаций с заинтересованными сторонами, типов документации и показателей качества, предоставляемых вместе с данными.

В. Аспекты статистического контроля для предотвращения идентификации

12. ЦГ признала значительные изменения, произошедшие в этой области со времени последнего обновления РКЕС. Следует отметить, что в мире ведется активная работа по изучению и внедрению новых методов статистического контроля для предотвращения идентификации (СКПИ), основанных на использовании шума, т. е. внесении небольших возмущений в микроданные или таблицы сопряженности с целью защиты конфиденциальной информации.

13. Например, после переписи населения ЕС 2011 года, в ходе которой в более сложных перекрестных таблицах было обнаружено большое количество ячеек с удаленными данными, а также после первых работ, проведенных, например, в Австралии¹⁹ и других странах, Евростат организовал проект с участием экспертов из разных стран по разработке более эффективных методов СКПИ для переписи населения ЕС 2021 года. В результате были разработаны рекомендации экспертов²⁰ и общедоступные программные средства²¹, в основе которых лежат два метода создания шумов: метод целенаправленной замены записей до таблицы и метод ключа ячеек после таблицы. Они были выбраны не только для того, чтобы уменьшить объем удаляемых данных, но и для того, чтобы устранить специфические дополнительные риски, связанные с публикацией данных по невложенным географическим признакам (например, сетка с квадратом 1 км и муниципальные границы).

14. Ожидается, что при проведении переписи населения в ЕС в 2021 году значительное число стран ЕС будет применять СКПИ, основанные на этих рекомендациях. В связи с продолжающейся модернизацией статистики населения в ЕС эти рекомендации будут оценены и, вероятно, получат дальнейшее развитие, а, возможно, будут дополнены другими идеями — например, Бельгия и Франция разработали собственные методы, основанные на шумах, специально для защиты своих выходных основанных на сетках материалов.

15. Бюро переписи населения США впервые опубликовало результаты переписи населения 2020 года, защищенные методом дифференцированного по доступу шума²². Это вызвало неоднозначную реакцию экспертов и пользователей данных, что

¹⁹ Thompson G., Broadfoot S., Elazar D. (2013), "Methodology for the Automatic Confidentialisation of Statistical Outputs from Remote Servers at the Australian Bureau of Statistics," [Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality](#).

²⁰ https://cros-legacy.ec.europa.eu/content/harmonised-protection-census-data_en.

²¹ <https://github.com/sdcTools/CensusProtection>.

²² Dwork C., McSherry F., et al. (2006), «Calibrating Noise to Sensitivity in Private Data Analysis», in Theory of Cryptography, eds. Halevi S., Rabin T., pp. 265-284. Berlin, Heidelberg: Springer Berlin Heidelberg.

²³ Abowd J. M. (2018), «The U.S. Census Bureau Adopts Differential Privacy,» in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18. London, UK: Association for Computing Machinery, p. 2867. DOI: 10.1145/3219819.3226070.

связано с серьезными опасениями²⁴ по поводу полезности, и последующие дискуссии²⁵.

16. Наконец, ЦГ приняла к сведению специфические проблемы СКПИ, возникающие при выпуске подробных данных с использованием интерактивных инструментов, позволяющих пользователям настраивать запросы о данных²⁶. Если такие средства предполагается предоставлять в рамках программы распространения, то стратегия СКПИ должна учитывать соответствующие сценарии атак.

17. В связи с этим ЦГ предложила внести в вопросник различные изменения, чтобы улучшить получение информации по этим направлениям. В частности, в качестве вариантов ответа в отношении используемых методов СКПИ были добавлены (целенаправленная) замена записей, метод ключа ячейки и дифференцированные по доступу методы; были добавлены вопросы об использовании и свойствах средств интерактивного вывода и последствиях СКПИ; и добавлен вопрос о конкретных мерах по защите от географического дифференцирования не вложенных географических признаков материалов.

С. Географические аспекты распространения

18. В качестве наиболее заметной тенденции, наметившейся со времени последнего обновления РКЕС, ЦГ назвала сильно возросший интерес к данным и анализу с географической привязкой. В первую очередь это касается широко распространенной тенденции предлагать данные переписи населения с привязкой к географическим сеткам, а также более мощные и интерактивные средства географической визуализации, такие как генераторы карт. Также были рассмотрены новые возникающие задачи для стратегий СКПИ (см. пункт 13).

19. Расширение доступности данных с географической привязкой и по наименьшим единицам и соответствующих средств анализа, как правило, требует и адаптации предоставляемых метаданных. Например, любые методы оценки, используемые для получения результатов переписи на самом малом географическом уровне (например, на уровне сетки), скорее всего, будут документироваться отдельно. Кроме того, страны могут публиковать вместе с результатами переписи соответствующую вспомогательную информацию, позволяющую проводить более мощный геопространственный анализ (например, данные о дорожных сетях или границах фиксированных или заданных пользователем географических зон).

20. В связи с этим ЦГ предложила внести в вопросник различные изменения, чтобы улучшить получение информации по этим направлениям. В частности, были обновлены вопросы, связанные с географией малых территорий, с тем чтобы в них явно охватывали сетки, включая разрешение сетки и то, какие статистические единицы (люди, жилища, домохозяйства и т. д.) будут доступны на сетке; было добавлено несколько вопросов для отражения географических методов оценки и вспомогательной информации, публикуемой вместе с данными по малым территориям или сетке.

21. ЦГ также обсудила вопрос о том, следует ли адаптировать или дополнить вопросы, касающиеся определения и использования «населенных пунктов» для разграничения городских и сельских территорий, чтобы охватить также более современные методологии разграничения городов, городских и сельских территорий.

²⁴ Например, Ruggles S., Fitch C., Magnuson D., Schroeder J. (2019), «Differential Privacy and Census Data: Implications for Social and Economic Research», AEA Papers and Proceedings, 109, 403–408.08.

²⁵ Например, Muralidhar, Domingo-Ferrer (2023), «Legacy Statistical Disclosure Limitation Techniques for Protecting 2020 Decennial US Census: Still a Viable Option», Journal of Official Statistics, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=4472525> or <http://dx.doi.org/10.2139/ssrn.4472525>.

²⁶ Например, Asghar H. J., Kaafar D. (2020), «Averaging Attacks on Bounded Noise-Based Disclosure Control Algorithms», Proceedings on Privacy Enhancing Technologies, 2020, 358–378.

В частности, Статистическая комиссия ООН на своей 51-й сессии в 2020 году одобрила новую методологию и руководство по делимитации городов, поселков и сельских районов²⁷, которые, по сути, определяют кластеры населения на основе комбинированного анализа малых административных районов и демографических сеток. Однако ЦГ решила препроводить эти соображения Целевой группе 11 по геопространственной информации и статистике малых территорий для переписей населения, а не предлагать собственные обновления к вопроснику в этом отношении.

IV. Выводы и последующие шаги

22. Целевая группа рассмотрела вопросник опроса 2013 года в контексте его тематической направленности и предложила Руководящей группе ряд обновленных и новых вопросов по трем направлениям работы: общие аспекты распространения, метаданных и архивирования; аспекты статистического контроля для предотвращения идентификации и географические аспекты распространения. Насколько известно автору на момент завершения работы над настоящим докладом, значительное большинство предложений было учтено в окончательном варианте вопросника для предстоящего обследования 2023 года.

23. Как только будут получены результаты обследования 2023 года, Целевая группа проанализирует их и приступит к работе по пересмотру действующих Рекомендаций КЕС.

²⁷ <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-02-20-499>.