

ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

Группа экспертов по статистике миграции

Женева, Швейцария, 26-28 октября 2022 года

Пункт С предварительной повестки дня

Использование новых источников данных для формирования статистики миграции

Использование новых источников данных для измерения международной миграции

Документ подготовлен Целевой группой по использованию новых источников данных для измерения международной миграции и трансграничной мобильности

В настоящем докладе представлены результаты работы Целевой группы по использованию новых источников данных для измерения международной миграции и трансграничной мобильности, созданной Бюро Конференции европейских статистиков (КЕС) в январе 2020 года.

В докладе описываются примеры национальной практики использования больших данных и новых источников данных, собранные в ходе двух обследований, охватывавших страны, участвующие в работе КЕС. В докладе делается вывод, что национальные статистические управления имеют ограниченный опыт использования новых источников данных в целях статистики миграции. Вместе с тем потенциал новых источников данных в плане решения возникающих проблем миграции и удовлетворения неотложных потребностей в данных весьма перспективен. Обмен соответствующей информацией и инструментами и их обновление облегчат использование больших данных и новых источников данных для измерения миграции и трансграничной мобильности.

Бюро КЕС рассмотрело доклад в феврале 2022 года и обратилось в Секретариат с просьбой направить его для электронных консультаций всем членам КЕС и другим заинтересованным участникам. Секретариат организовал электронные консультации в марте — апреле 2022 года. Результаты консультаций были представлены в ходе пленарного заседания КЕС в июне 2022 года. Тогда КЕС утвердила доклад. Данный документ представляет собой итоговую версию доклада, которая включает изменения и дополнения по итогам консультаций

Благодарности

Доклад подготовлен Целевой группой Конференции европейских статистиков по использованию новых источников данных для измерения международной миграции и трансграничной мобильности, в состав которой входят следующие члены:

Чинция Конти, Италия (сопредседатель Целевой группы)
Энрико Туччи, Италия (сопредседатель Целевой группы)
Тристан Кейн, Канада
Жюльен Берар-Шаньон, Канада
Шорена Циклаури, Грузия
Эдгар Вьельма, Мексика
Омар Муру, Мексика
Ким Данстан, Новая Зеландия
Пубуду Сенанаяке, Новая Зеландия
Нериман Джан Эрган, Турция
Шерифе Дилек Йылмаз, Турция
Альпер Акар, Турция
Ребекка Бриггс, Великобритания
Меган Бауэрс, Великобритания
Джейсон Шахтер, США
Анжелика Менчака, США
Джампаоло Ланциери, Евростат
Жан-Кристоф Дюмон, ОЭСР
Сесиль Торо, ОЭСР
Паоло Валенте, ЕЭК ООН
Андрес Викат, ЕЭК ООН
Мария Изабель Кобос Эрнандес, Статистический отдел ООН

Главы отчета были обсуждены и согласованы всей Целевой группой. Некоторые члены группы взяли на себя основную ответственность за составление некоторых разделов доклада, в частности:

Введение, Паоло Валенте
Национальный опыт работы с большими данными и новыми источниками данных, Чинция Конти и Энрико Туччи
В центре внимания опыт Соединенных Штатов, Джейсон Шахтер и Анжелика Менчака
Причины отказа от использования новых источников данных, Паоло Валенте
Важные инновации в измерении международной миграции, Пубуду Сенанаяке и Ким Данстан
Справочник по литературе: DIMiS, Джампаоло Ланциери
Выводы, Чинция Конти и Энрико Туччи

ЕЭК ООН благодарна Ким Данстан за редактирование отчета.

Оглавление

Благодарности	2
1. Введение	5
1.1. Справочная информация	5
1.2. Методология и содержание доклада	6
2. Национальный опыт использования больших данных и новых источников данных	8
2.1. Резюме опыта: первый опрос	8
2.2. Резюме опыта: второй опрос	9
2.3. Пример: опыт Соединенных Штатов	10
2.3.1 Традиционные источники обследований	11
2.3.2 Данные о воздушных пассажирских перевозках как альтернативный источник	11
2.3.3 Ураган «Мария»	12
2.3.4 Сочетание данных об авиаперевозках с обследованиями домашних хозяйств	12
2.3.5 Результаты	12
2.3.6 Корректировка с учетом COVID-19	15
2.3.7 Принятый в 2021 году метод на основе данных о воздушных пассажирских перевозках	16
3. Причины неиспользования новых источников данных	18
3.1. Использование новых источников данных для подготовки статистических данных в других областях статистики	18
3.2. Причины неиспользования новых источников данных	19
4. Масштабные нововведения в области измерения международной миграции	21
4.1. Нововведения в измерении международной миграции, связанные с данными	21
4.1.1 Краткая информация об использовании административных данных	21
4.1.2 Отдельные примеры нововведений с использованием административных данных	23
4.1.3 Краткая информация об использовании данных обследований	24
4.2. Нововведения в измерении международной миграции, связанные с методологией	24
4.2.1 Отдельные примеры методологических нововведений	24
5. Справочная система по специальной литературе: база данных по нововведениям в статистике миграции (DIMIS)	26
6. Выводы и последующие шаги	30
6.1. Пока ограниченные масштабы использования	30
6.2. Потенциальные преимущества	31
6.3. Следующие шаги — ведение Базы данных по нововведениям в статистике миграции	31

1. Введение

1.1. Справочная информация

1. Миграция и другие формы трансграничной мобильности являются вопросами большой политической важности. Спрос на статистические данные в этих областях дополнительно возрос в свете [Повестки дня в области устойчивого развития на период до 2030 года](#) и [Глобального договора о безопасной, упорядоченной и легальной миграции](#) (2018 год). Перед статистическим сообществом по-прежнему стоит задача отражения международной миграции и трансграничной мобильности методами, позволяющими удовлетворять растущие потребности пользователей.
2. Для измерения миграции и трансграничной мобильности используются различные источники, такие как переписи населения и жилищного фонда, обследования домашних хозяйств и данные административного учета, и каждому из них присущи свои достоинства и недостатки. Интеграция данных из различных источников — это способ повысить информативность данных и уменьшить проблемы охвата или точности. Однако даже это зачастую не позволяет охватить все аспекты миграции и трансграничной мобильности.
3. Новые нетрадиционные источники данных, например данные использования мобильных телефонов, кредитных карт и социальных сетей, — обычно называемые большими данными и данными социальных сетей, — могут быть полезны для ведения статистики миграции при применении в сочетании с традиционными источниками. В [Руководстве ЕЭК ООН по интеграции данных для измерения миграции](#) предлагается продолжить работу по использованию потенциала больших данных «для распространения новых методов на международном уровне, чтобы поддержать первые шаги стран по реализации потенциала таких данных для подготовки статистических материалов о миграции».
4. Несмотря на проблемы доступности, точности и конфиденциальности этих новых источников, в последние годы появились примеры, наглядно демонстрирующие их потенциал. В ходе [рабочей сессии ЕЭК ООН-Евростата по статистике миграции 2018 года](#) приводились примеры использования данных Facebook для получения возрастной структуры «экспатов» в разбивке по происхождению и данных твитов с географическими метками для оценки мобильности. На [рабочей сессии ЕЭК ООН-Евростата по статистике миграции в 2019 году](#) был приведен пример из сферы официальной статистики, демонстрирующий использование в Бюро переписей Соединенных Штатов данных о пассажирских авиаперевозках для более точных оценок миграции. По мере появления новых примеров их необходимо собирать и анализировать, чтобы помочь национальным статистическим управлениям (НСУ) в использовании новых источников данных, опираясь на результаты [проектов ЕЭК ООН по вопросам больших данных в официальной статистике](#) и других инициатив, в рамках которых изучаются взаимоотношения поставщиков официальных данных с владельцами больших данных.
5. В октябре 2019 года Бюро Конференции европейских статистиков (КЕС) на основе документа Мексики¹ и записки ЕЭК ООН провело углубленный обзор статистики международной миграции и трансграничной мобильности. Бюро подчеркнуло важность накопления примеров использования национальными статистическими управлениями новых источников данных для подготовки официальной статистики в этой области и упомянуло рабочую сессию ЕЭК ООН-Евростата по статистике миграции 2019 года в

¹ http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/bur/2019/October/02_In-depth_review_on_international_migration_Rev.1.pdf.

качестве оптимального форума для дальнейшего обсуждения этого вопроса. Оно просило секретариат и Руководящую группу по статистике миграции представить на следующем совещании Бюро предложение по дальнейшей работе в области использования новых источников данных для измерения миграции и трансграничной мобильности.

6. Участники рабочей сессии ЕЭК ООН-Евростата по статистике миграции 2019 года признали необходимость проработки вопросов использования новых типов источников данных, таких как мобильные устройства, социальные сети, спутниковые изображения и интернет-платформы, а также рассмотрения существующих примеров использования новых источников данных в интересах подготовки официальной статистики миграции. Для этого и с учетом высказанной Бюро просьбы Бюро КЕС в январе 2019 года создало Целевую группу по использованию новых источников данных для измерения международной миграции и трансграничной мобильности для анализа существующего опыта и планов НСУ в области использования новых источников данных для измерения миграции и трансграничной мобильности, выявления соответствующих примеров вне сферы официальной статистики, накопления примеров в справочной системе и разработки механизма ее обновления новыми примерами, а также анализа собранного материала для подготовки рекомендаций НСУ по использованию новых источников.

1.2. Методология и содержание доклада

7. Целевая группа начала свою работу с выявления соответствующей научной литературы и информации, имеющейся у ее членов. Для сбора информации о соответствующем опыте (в НСУ и вне их рамок) по вопросам использования новых источников данных для измерения миграции и трансграничной мобильности Целевая группа провела в период с ноября 2020 года по февраль 2021 года онлайн-опрос среди НСУ в регионе ЕЭК ООН.
 - a. Большие данные, которые включают такие источники данных, как социальные сети (Facebook, Twitter, Tumblr и т. д.), блоги и комментарии, личные документы, изображения: Instagram, Flickr, Picasa и т. д., видео: Youtube и т. д., поиск в Интернете, данные мобильных устройств: текстовые сообщения, карты, созданные пользователем, электронная почта, медицинские записи, коммерческие транзакции, выписки по банковским счетам / книга регистрации владельцев акций, электронная коммерция, кредитные карты, домашняя автоматизация, датчики погоды/загрязнения, научные датчики, видео/изображения систем безопасности/наблюдения, местоположение мобильного телефона, автомобили, спутниковые изображения, журналы, веб-журналы)
 - b. Другие источники данных, такие как информация о пассажирах авиаперевозок, которые отличаются от традиционных источников (выборочные обследования, переписи, административные данные)
8. Из 34 стран ЕЭК ООН, принявших участие в опросе, только 5 (15 %) сообщили, что им известно о процессе или проекте (с участием или без участия НСУ), где для измерения миграции и трансграничной мобильности используются новые источники данных, такие как большие данные и данные социальных сетей.
9. Целевая группа проанализировала результаты опроса и решила провести повторный опрос, чтобы получить от НСУ дополнительную информацию:
 - a. о том, используют ли НСУ новые источники данных для подготовки статистических данных в других областях статистики;
 - b. об основных причинах, в силу которых НСУ не используют новые источники данных для статистики миграции;

- с. о том, прорабатываются ли в НСУ какие-либо значительные нововведения в области измерения международной миграции, включая новые подходы с использованием существующих источников, и изучаются ли новые источники данных.
10. Повторный опрос проводился в марте 2021 года, ответы были получены от 27 стран.
 11. Результаты этих двух опросов представлены соответственно в разделах «Национальный опыт использования больших данных и новых источников данных», «Причины неиспользования новых источников данных» и «Серьезные нововведения в области измерения международной миграции».
 12. Помимо информации, которую представили НСУ в ходе опросов, Целевая группа также собрала большой объем информации о различных видах исследований, дискуссионных докладах или научных журналах, прямо или косвенно связанных с использованием новых источников данных для измерения международной миграции и трансграничной мобильности. Для представления этого материала Целевая группа в соответствии с просьбой Бюро КЕС разработала концепцию онлайн-справочной системы с возможностью ее обновления новыми примерами, которые появятся в будущем. Подготовленный Целевой группой краткий обзор справочной системы представлен в разделе «Справочная система по специальной литературе: DIMiS».

2. Национальный опыт использования больших данных и новых источников данных

13. Судя по результатам двух опросов, проведенных Целевой группой, очевидно, что в настоящее время НСУ редко используют большие данные для измерения международной миграции. Вместе с тем некоторые учреждения используют новые источники данных, другие же изучают вопрос использования новых источников данных.

2.1. Резюме опыта: первый опрос

14. Единственный проект с использованием больших данных, о котором было сообщено, реализуется в Грузии. Целью проекта является измерение мобильности населения в рамках работы Комитета экспертов Организации Объединенных Наций по использованию больших данных и инструментария науки о данных для целей официальной статистики (КЭБА ООН)². Задача заключается в оценке тенденций мобильности групп населения, таких как внутренние мигранты, международные мигранты и туристы. НСУ Грузии (Геостат), грузинский регулятор мобильной связи (Национальная комиссия по коммуникациям Грузии) и другие члены целевой группы (Международный союз электросвязи (МСЭ), СОООН, Евростат, компания «Позитивум» и др.) совместными усилиями разрабатывают и тестируют методы и методологию оценки статистики миграции и туризма в Грузии с использованием данных мобильной телефонной связи. Однако официальная статистика в рамках проекта пока не выпускалась, и в настоящее время проект приостановлен по техническим причинам.
15. Один из примеров использования новых источников данных, а не больших данных, приводит Бюро переписей США. Этот проект основан на использовании данных о пассажирских авиаперевозках для корректировки чистой международной миграции между Пуэрто-Рико и Соединенными Штатами для учета последствий урагана «Мария» в 2017 году. В последние годы Бюро переписей США прилагает целенаправленные усилия по интеграции новых источников данных для совершенствования применяемого метода оценки чистой миграции между Соединенными Штатами и Пуэрто-Рико. Эти усилия вызывались необходимостью повышения своевременности оценок миграции, которые обычно производятся на основе крупных массивов данных обследования домашних хозяйств, собранных Бюро переписей. Методология, основанная на обследовании, работает успешно, когда динамика миграционных тенденций стабильна во времени, но, как правило, ее эффективность падает в случае их быстрых изменений, например из-за стихийных бедствий (скажем, ураганов, глобальных пандемий). Для подготовки более актуальных и точных оценок миграции Бюро переписей прорабатывает возможность интеграции административных данных других агентств с данными, собранными в ходе проводимых им обследований, с переходом при этом к методу измерения чистой миграции между Пуэрто-Рико и остальными странами мира исключительно на основе административных данных.
16. Стимулом к этим первым усилиям по объединению данных обследований и административных данных послужили последствия урагана «Мария» в 2017 году, который вызвал массовый отток мигрантов из Пуэрто-Рико. С помощью этих методов интеграции макроданных данные об авиаперевозках, публикуемые Бюро статистики транспорта (БСТ), объединялись с оценками на основе опросов, полученными в ходе Обследования американского общества (ОАО) и Обследования общества Пуэрто-Рико (ООПР), чтобы лучше оценить влияние этого стихийного бедствия на оценки 2018 года. Интеграция макроданных использовалась и для учета возвратной миграции в Пуэрто-Рико после урагана «Мария» (оценки на 2019 год), а также для корректировки

² <https://unstats.un.org/bigdata/blog/2019/mpd-task-team.cshtml>.

данных с учетом новых тенденций миграции, возникших в результате пандемии COVID-19 (оценки 2020 года). Начиная с 2021 года Бюро переписей переходит на метод, при котором общая чистая миграция в Пуэрто-Рико определяется на основе данных об авиарейсах, но при этом приток и отток населения в разбивке по демографическим характеристикам по-прежнему определяется с помощью данных ОАО и ОНР. Этот новый метод удобен, поскольку он позволит повысить своевременность наличия данных, снизить необходимость корректировок, улучшить согласованность с базисным периодом (год, закончившийся 30 июня) и расширить объем данных о миграции с включением международных перемещений в Пуэрто-Рико и из него (подробнее см. раздел «В фокусе внимания – опыт Соединенных Штатов Америки»).

17. В порядке примеров использования больших данных указывались другие проекты, но они связаны с внутренней миграцией (например, в Нидерландах), или их результаты еще не получены (например, в Латвии).

2.2. Резюме опыта: второй опрос

18. В ходе повторного опроса четыре страны сообщили, что они изучают возможность использования новых источников данных для измерения международной миграции.
19. На действительно «больших данных» основывается только один проект. В настоящее время в Управлении национальной статистики Соединенного Королевства изучаются агрегированные и анонимизированные данные мобильной телефонной связи (или данные о перемещении больших масс людей) от компании O2 Motion. В этих данных анализируется подключение мобильных устройств к вышкам «O2» для получения представления о тенденциях перемещения больших масс людей по всему Соединенному Королевству. Эти данные включают географическое местоположение мест (где мобильные телефоны подключаются к вышкам сотовой связи) на национальном уровне и уровне местных административных единиц и, в принципе, помогают понять закономерности международной миграции. Планируется проанализировать эти данные на предмет мобильности мигрантов, приезжающих в Соединенное Королевство и уезжающих из него, включая такие аспекты, как продолжительность пребывания при поиске работы, для получения более своевременных показателей миграции. Этот проект находится на очень ранней стадии, существует также более широкая программа работы по вопросам международной мобильности в научно-исследовательском центре интеллектуальной обработки данных, где изучаются и другие вопросы³. В качестве одного из нескольких источников при моделировании миграции используется также информация об пассажирских авиаперевозках.
20. В других странах изучаются новые источники данных, но не обязательно с охватом больших данных.
21. По сообщению Статистического управления Новой Зеландии, для расчета официальных показателей международной миграции используются административные данные (паспортные данные, полученные при пересечении границы). Для целей анализа географического распределения населения в Новой Зеландии изучаются и другие источники данных, например данные мобильной телефонной связи, включая людей, приехавших из-за границы, и население страны (включая недавних мигрантов).
22. Статистическое управление Словацкой Республики не использует большие данные для измерения статистики миграции, уделяя главное внимание инновационным формам использования административных данных. Главной задачей называется

³ Полученные от «O2 Motion» аналитические данные никоим образом не позволяют идентифицировать или локализовать отдельных людей и предоставляются в строгом соответствии с принципами конфиденциальности. [Using Mobile Phone Data for Enhancing International Passenger Survey Traveller Statistics.](#)

систематическое использование источников административных данных для целей официальной статистики миграции.

23. В Республике Молдова начиная с 2018 года Национальное бюро статистики обрабатывает большие объемы данных о пересечении государственной границы физическими лицами. Это данные из административных источников, находящихся в ведении Генеральной инспекции пограничной полиции, которые используются для оценки международной миграции на основе международного определения обычного места жительства. При расчете используются административные данные за 3 года подряд, то есть для определения статуса эмигранта или иммигранта за учетный год (например, для 2020 года используются данные за предыдущий 2019 год и последующий 2021 год). В настоящее время доступны окончательные оценки международной миграции за 2014–2019 годы. Эти результаты используются для оценки численности населения Республики Молдова, обычно проживающего в стране.
24. Статистическое управление Литвы сообщило, что применение математических методов к новым источникам данных (и большим данным) может служить подспорьем при решении проблем, связанных со статистикой международной миграции. Параллельно в Литве существует государственная информационная система управления данными, обеспечивающая общую платформу управления данными для всех учреждений, доступ к данным для государственных нужд, эффективный обмен между учреждениями и обмен данными с деловыми и научными кругами. Эта система расширит возможности использования государственных информационных систем, реестров и других источников для статистики международной миграции.
25. Национальный институт статистики Италии также прорабатывает вопросы интеграции новых источников данных для статистики миграции. В Италии статистика международной миграции ведется на основе уведомлений о прибытии или отъезде, направляемых мигрантами в регистр населения. Осуществляется специальный исследовательский проект по теме использования данных регистрации и снятия с учета *ex-officio* для оценки недостающей части этих уведомлений, особенно для целей статистики эмиграции, поскольку хорошо известно, что эмигранты могут не сообщать о своем отъезде. В некоторых случаях после снятия с учета *ex-officio* через несколько месяцев производится регистрация ввиду повторного появления того же лица. Итальянская система продольных данных (ANVIS), которая основана на микродемографическом учете, позволяет выявлять эти взаимокомпенсирующие административные перемещения и исключать их из демографического баланса. Напротив, те, кто не появляются вновь, вероятно, покинули страну. В некоторых случаях (продолжительность периода с момента снятия с учета *ex-officio* до регистрации ввиду повторного появления составляет более одного года) не ясно, покинул ли человек страну, а потом вернулся или же он постоянно проживал в стране в течение всего этого периода. Для этого факт пребывания таких лиц проверяется по Интегрированному архиву постоянного населения, который включает и объединяет множество административных источников (регистры рабочей силы и образования, регистр налоговых деклараций, регистры трудовых доходов, пенсий и непенсионных пособий и архив видов на жительство). Этот архив, который уже используется для оценки охвата переписи постоянного населения, можно использовать и для подтверждения пребывания отдельных лиц в Италии по методологии «признаков жизни». Имеется в виду проверка предположения о том, что человек покинул страну, путем анализа его «признаков жизни» в данный период.

2.3. Пример: опыт Соединенных Штатов

26. Этот раздел посвящен работе Бюро переписей США по включению новых источников данных для измерения миграции, в частности для оценки миграции в Пуэрто-Рико. Эти новые источники данных и методологии помогают преодолеть ограничения

предыдущих методов, зависящих от более традиционных источников данных. Интеграция новых и прежних источников данных позволяет получить более своевременные и точные оценки миграции.

2.3.1 Традиционные источники обследований

27. Бюро переписей США составляет ежегодные оценки численности населения по состоянию на 1 июля каждого года. Обследование американского общества и обследование общества Пуэрто-Рико (ОАО/ООПР) — это ежегодные обследования домашних хозяйств населения Соединенных Штатов, в ходе которых запрашивается подробная информация. В настоящее время в рамках ОАО обследуется около 3,5 млн домашних хозяйств в год, а в ООПР выборка включает 36 000 адресов в Пуэрто-Рико.
28. За основу в оценках миграционных потоков из Пуэрто-Рико в Соединенные Штаты брались ответы на задававшийся при проведении ОАО вопрос о месте жительства год назад (МЖГН), когда уточнялось, где респонденты жили за год до обследования. В противоположность этому миграционные потоки из Соединенных Штатов в Пуэрто-Рико рассчитывались на основе ответов на вопрос о МЖГН, полученных в ходе ООПР. Данные ОАО/ООПР собираются на непрерывной основе в течение календарного года, хотя переезд мог произойти в любое время в течение двухлетнего периода, в зависимости от того, когда респондент был включен в выборку и когда он фактически переехал.

2.3.2 Данные о воздушных пассажирских перевозках как альтернативный источник

29. В Соединенных Штатах данные об авиаперевозках собираются из ежемесячных отчетов, которые свыше 200 коммерческих американских и иностранных авиаперевозчиков представляют в БСТ как по внутренним, так и по международным рейсам. Данные представляются по всем авиаперевозкам (таким образом, выборка не производится) в соответствии с федеральными правилами отчетности, которые вступили в силу в октябре 2002 года.
30. В случае Пуэрто-Рико внутренние данные по воздушным пассажирским перевозкам (ВПП) позволяют получить ежемесячную информацию о количестве авиапассажиров на рейсах между Пуэрто-Рико и континентальной частью Соединенных Штатов. Следует отметить, что данные о ВПП включают информацию обо всех пассажирах без дифференциации по их типу и, соответственно, включают туристов и временно приезжающих лиц, составляющих большинство пассажиров. Немигранты учитываются на авиарейсах в обоих направлениях, мигранты же учитываются только в одном направлении — если только они не уезжают временно, в этом случае они будут учтены снова по возвращении.
31. Одно из ограничений этого метода заключается в том, что он позволяет получить цифру только «чистой» миграции, без информации о совокупном притоке или оттоке, поскольку мигрантов нельзя отличить от общего числа пассажиров, въезжающих в Пуэрто-Рико или покидающих его. Еще одним ограничением является то, что данные не включают демографические характеристики мигрантов. Кроме того, этот метод применим только к стране или территории, не имеющей сухопутных границ, например к острову Пуэрто-Рико, поскольку основным способом прибытия или отъезда является авиатранспорт.
32. Ежемесячные данные о чистой динамике пассажиропотока авиакомпаний отражают сезонные колебания, связанные с туризмом, с более интенсивным движением в Пуэрто-Рико и из него в летние и зимние месяцы отпусков. В зависимости от периода измерения (например, календарный год), это может вызывать колебания от года к году, связанные с годовой динамикой туризма. Например, большое количество туристов может быть

зарегистрировано в декабре, в то время как возвращаться они же могут в январе следующего года. Считается, что со временем эти колебания взаимоуравняются.

2.3.3 Ураган «Мария»

33. В сентябре 2017 года на Пуэрто-Рико обрушился ураган пятой категории «Мария», повлекший масштабные разрушения, человеческие жертвы и отток населения на континентальную часть Соединенных Штатов. Стихийные бедствия могут влечь демографические последствия — через перемещение людей из пострадавших районов, а также смертность в результате катаклизмов. По мере увеличения частоты и масштабов стихийных бедствий возрастают и потребности в оценке численности населения для точного измерения последствий таких бедствий, что нередко требует использования различных источников данных или применения новых методов.

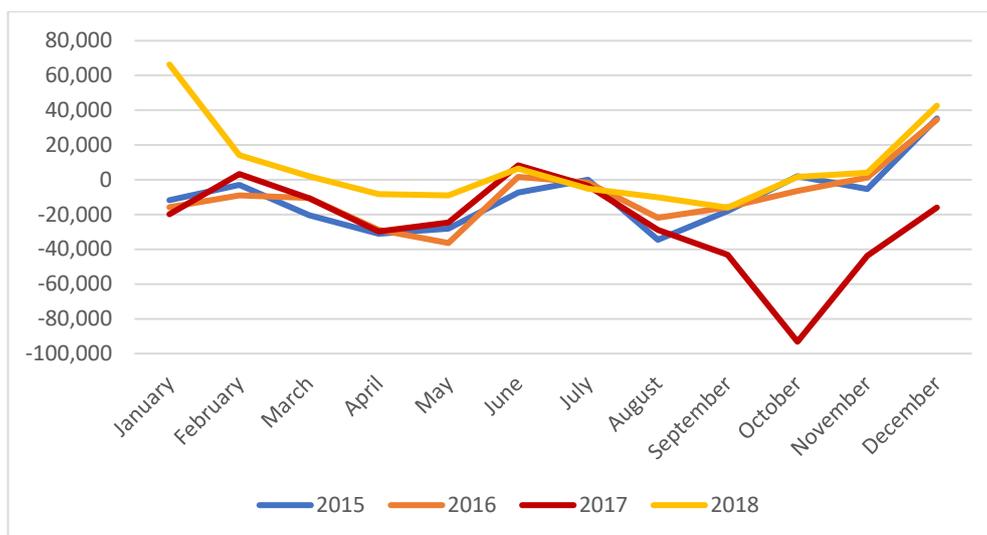
2.3.4 Сочетание данных об авиаперевозках с обследованиями домашних хозяйств

34. Первоначально для оценки влияния урагана «Мария» на миграцию в Пуэрто-Рико и из него Бюро переписей США использовало данные ВПП в сочетании с данными ОАО и ООПР. Выборочные обследования домашних хозяйств, такие как ОАО, не предназначены для отражения внезапных массовых перемещений людей, поскольку ретроспективные данные о миграции, полученные по результатам обследований, как правило, «запаздывают» по сравнению с фактическими миграционными событиями. В рамках обследований измеряется не событие миграции в реальном времени, а событие, когда мигрант включен в выборку. Этот метод нормально работает при стабильной динамике миграции, но в случае значительных колебаний масштабов перемещений они будут полностью отражены только позже (обычно в следующем году обследования). Учет, в принципе, усложнялся тем, что ураган «Мария» пришелся на поздний этап (конец сентября) цикла сбора данных ОАО в 2017 году, и соответственно в выборку включался короткий период времени. В результате нужно было искать альтернативный источник данных для измерения влияния урагана «Мария» на миграцию из Пуэрто-Рико.

2.3.5 Результаты

35. Данные о ВПП показали большой чистый отток пассажиров из Пуэрто-Рико в последние месяцы 2017 года (с сентября по декабрь), за которым последовал обратный поток в первые месяцы 2018 года (диаграмма 1). Возврат в Пуэрто-Рико из Соединенных Штатов уменьшился в течение первого квартала 2018 года и опять сменился чистым оттоком к апрелю 2018 года. До урагана «Мария» (20 сентября 2017 года) динамика чистого пассажирооборота между Соединенными Штатами и Пуэрто-Рико следовала относительно стабильным ежемесячным тенденциям, когда больше пассажиров покидали Пуэрто-Рико, чем въезжали в него, за исключением некоторых летних или зимних месяцев (в частности, июня и декабря). Это соответствует сезонной структуре авиарейсов, когда больше туристов прибывает в летние и зимние месяцы и пуэрториканцы, живущие в Соединенных Штатах, приезжают на родину в периоды отпусков.

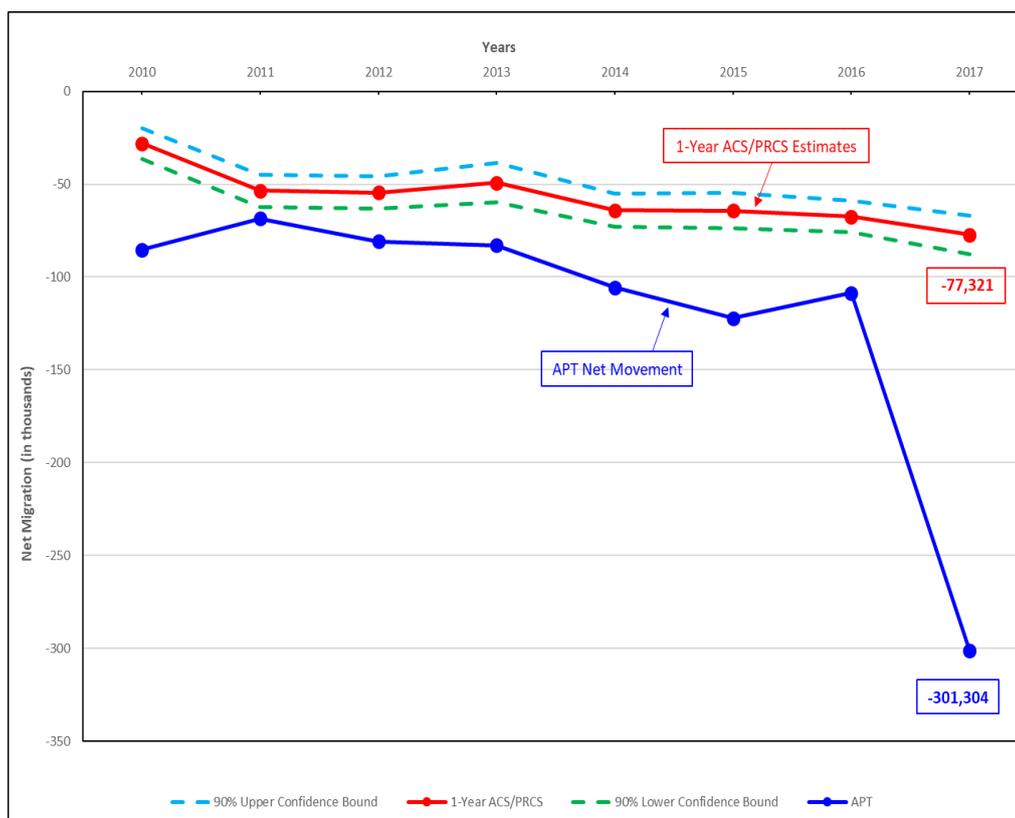
Рисунок 1. Ежемесячный чистый пассажирооборот на авиарейсах между Пуэрто-Рико и Соединенными Штатами, 2015–2018 годы



Источник: бланк 41 Бюро статистики транспорта США, сегмент T100 (международный), все перевозчики.

36. Согласно «смешанной» методологии, использованной для оценок 2018 года, делалось допущение, что данные о ВПП лучше отражают истинное влияние урагана «Мария» на миграционные процессы, чем данные ОАО/ООПР, и что историческая динамика данных ОАО и ВПП примерно совпадает; таким образом, эту взаимосвязь можно было использовать для получения более точной оценки за этот период. До 2017 года тенденции в динамике данных ОАО и ВПП, как правило, были схожими, при этом по данным о ВПП величина чистой эмиграции из Пуэрто-Рико была устойчиво выше, чем по данным ОАО (диаграмма 2).

Рисунок 2. Динамика чистой миграции между Пуэрто-Рико и Соединенными Штатами, 2010–2017 годы



Источники: Бюро переписей США, Обследование американского сообщества и Обзор общества Пуэрто-Рико; бланк 41 Бюро статистики транспорта США, сегмент T100 (международный), все перевозчики.

37. Поскольку предыдущие оценки основывались на ОАО/ООПР, были предприняты усилия для обеспечения максимальной сопоставимости этих двух источников данных и брались ежемесячные данные об авиарейсах за 2017 календарный год, чтобы добиться совпадения с периодом оценки ОАО/ООПР. Информация об авиарейсах ограничивалась внутренними рейсами между Соединенными Штатами и Пуэрто-Рико, международные авиарейсы исключались. В данном методе применялся простой коэффициент с использованием соотношения чистой миграции по данным ОАО/ООПР и о ВПП за два года — 2015 и 2016. Для сохранения методологического соответствия с предыдущими оценками на основе ОАО/ООПР рассчитанный коэффициент применялся к показателю чистой миграции между Пуэрто-Рико и США, измеренному по данным о ВПП за 2017 календарный год.
38. Поскольку оценки численности населения Бюро переписей за 2018 год отражают численность населения на 1 июля 2018 года, учитывалась также возвратная миграция в Пуэрто-Рико в начале 2018 года. Для учета январской возвратной миграции брался 12-месячный период ВПП с февраля 2017 года по январь 2018 года, затем применялся коэффициент корректировки, чтобы добиться максимального совпадения периода с ОАО, но при этом учесть возвратную миграцию в Пуэрто-Рико после урагана «Мария». Смещение периода на один месяц помогло учесть возвратную миграцию, при этом большинство месяцев (11 из 12) остались в пределах календарного года, эквивалентного 2017 году в случае ОАО/ООПР. В результате этой модификации чистая миграция из Пуэрто-Рико в Соединенные Штаты по данным о ВПП составила 215 166 человек, и с поправкой на коэффициент ВПП-ОАО была получена окончательная цифра — 123 399 человек (Рисунок 3).

Рисунок 3. Скорректированная динамика чистой миграции между Пуэрто-Рико и Соединенными Штатами, 2010–2017 годы



Источники: Бюро переписей США, Обследование американского сообщества и Обзор общества Пуэрто-Рико; бланк 41 Бюро статистики транспорта США, сегмент T100 (международный), все перевозчики.

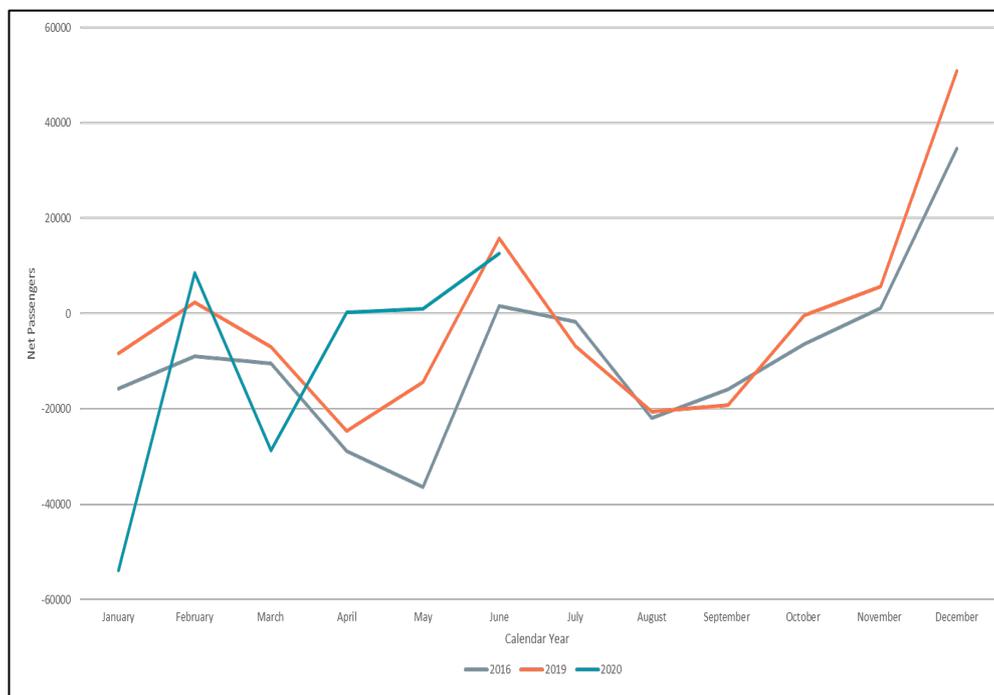
2.3.6 Корректировка с учетом COVID-19

39. Пандемия COVID-19 значительно повлияла на динамику числа лиц, прибывающих в Соединенные Штаты и уезжающих из страны начиная с марта 2020 года. Серьезно повлияв на международные миграционные потоки, это сказалось и на динамике числа прибывающих в Пуэрто-Рико и уезжающих из него, особенно в начале пандемии. Предполагалось, что данные ОАО/ООПР за 2019 год (которые не охватывают период пандемии) не будут адекватно отражать динамику чистой миграции в Пуэрто-Рико за 2020 год, поэтому и в этом случае для корректировки использовались данные о ВПП. Однако, поскольку начало пандемии пришлось только на последнюю треть года, оценки (март — июнь 2020 года), применить тот же смешанный метод было невозможно, поскольку он основан на данных ВПП и ОАО/ООПР за полный календарный год. Поэтому для корректировки чистой миграции Пуэрто-Рико за март — июнь 2020 года использовались данные о ежемесячных авиарейсах в Пуэрто-Рико и США и из них.
40. В динамике авиарейсов в Пуэрто-Рико и из Пуэрто-Рико обычно прослеживаются устойчивые тенденции (Рисунок 44). Как правило, в марте — мае наблюдается чистый отток пассажиров, в то время как июнь — начало туристического сезона — является месяцем притока пассажиров. Данные ВПП за март 2020 года продемонстрировали значительное увеличение чистого числа пассажиров, убывающих на континентальную часть Соединенных Штатов (из Пуэрто-Рико); в апреле и мае наблюдался небольшой

⁴ 2017 и 2018 годы исключены из-за месячных аномалий, вызванных ураганом «Мария».

положительный чистый прирост для Пуэрто-Рико, поскольку количество рейсов было резко сокращено; и в июне был отмечен крупный положительный чистый прирост для Пуэрто-Рико с увеличением количества авиарейсов на остров и с острова.

Рисунок 4. Ежемесячный чистый пассажирооборот на авиарейсах между Пуэрто-Рико и Соединенными Штатами, 2016, 2019 и 2020 годы



Источник: бланк 41 Бюро статистики транспорта США, сегмент T100 (международный), все перевозчики.

- Для расчета поправочного коэффициента с учетом пандемии COVID на 2020 год с использованием данных о ВПП была рассчитана совокупная чистая численность пассажиров за месяцы с марта по июнь 2020 года. Затем рассчитывался необходимый коэффициент с использованием данных прошлых периодов о среднем чистом пассажиропотоке за март — июнь 2010–2019 годов, исключая 2018 год (ураган «Мария»), и чистом пассажиропотоке за март — июнь 2020 года. Затем совокупный сезонный показатель по данным ОАО/ООПР был уменьшен путем умножения одной трети годового показателя по данным ОАО/ООПР 2019 года на коэффициент чистого пассажиропотока. Для получения окончательной оценки две трети оценочных показателей по данным ОАО/ООПР 2019 года были применены к сокращенному совокупному сезонному показателю по данным ОАО/ООПР. В результате корректировки с учетом общего влияния пандемии COVID-19 были получены данные о сокращении чистой эмиграции из Пуэрто-Рико.

2.3.7 Принятый в 2021 году метод на основе данных о воздушных пассажирских перевозках

- Начиная с 2021 года Бюро переписей США использует данные о ВПП для прямого измерения чистой миграции в Пуэрто-Рико и остальные страны мира и из них, хотя ОАО/ООПР будут по-прежнему необходимы для оценки притока, оттока и демографических характеристик мигрантов. Переход к методу расчета чистой миграции на основе данных об авиарейсах позволит повысить точность и регулярность оценок чистой миграции в Пуэрто-Рико и сократить объем будущих корректировок, необходимых для учета основных событий, влияющих на миграцию.

-
43. Методология на базе данных об авиарейсах позволяет устранить ряд ограничений, присущих прежней методологии:
- a. Поскольку данные обследования публикуются не сразу, период сбора, обработки и публикации данных не совпадает с оцениваемым периодом. Ввиду этого изменения в динамике численности населения становятся заметны только в следующем году оценки;
 - b. Методология учитывала только миграцию из Пуэрто-Рико в и из США, при этом делалось неверное допущение о нулевой чистой миграции в островные территории Соединенных Штатов и зарубежные страны и из них. ООПР позволяет измерить количество иммигрантов, прибывших в Пуэрто-Рико не из Соединенных Штатов, но не эмигрантов, выехавших за границу.
 - c. Период, охватываемый ОАО/ООПР (январь — декабрь календарного года), не совпадает с годом оценок (июль — июнь).
 - d. С 2018 года для сохранения точности и надежности данных приходится производить коррекцию данных и использовать прогнозы, что в будущем становится менее вероятным.
44. Ежегодные общие показатели чистой миграции будут рассчитываться на основе данных о входящем/исходящем пассажиропотоке за год оценки (июнь — июль) с использованием данных о ВПП: для получения оценочного показателя чистой миграции общее количество прилетающих в Пуэрто-Рико пассажиров вычитается из общего количества пассажиров, вылетающих из него. В отличие от «смешанного метода», описанного выше, сюда включаются пассажиры международных рейсов.
45. Поскольку данные об авиаперевозках позволяют оценить только чистый пассажиропоток, не охватывая никаких конкретных потоков прибывающих или убывающих мигрантов или демографических характеристик, эта информация будет по-прежнему браться из ОАО/ООПР. Таблицы данных в разбивке по полу будут составляться с использованием данных ООПР (МЖГН) о притоке мигрантов за один год с последующим распределением по годам возраста с помощью данных о половозрастном распределении из пятилетних оценок ООПР (МЖГН). Эта процедура будет повторно применяться для оттока мигрантов с использованием оценок ОАО (МЖГН) за один год в разбивке по полу и оценок ОАО (МЖГН) за пятилетний период для получения данных о половозрастном распределении.
46. Как уже отмечалось, сезонные колебания, связанные с туризмом, особенно в летние и зимние месяцы, могут вызывать тенденции, требующие корректировки с учетом годовых колебаний. Корректировки колебаний будут зависеть от периода измерения, чтобы сбалансировать сезонные тенденции в динамике притока и оттока мигрантов.

3. Причины неиспользования новых источников данных

47. Судя по результатам опроса, в настоящее время лишь небольшое число стран региона ЕЭК ООН используют для измерения миграции и трансграничной мобильности новые источники данных, таких как большие данные и данные социальных сетей, или рассматривают возможность их использования. Из 36 стран, представивших ответы в ходе опроса, только 5 стран сообщили об осуществляемых на данном этапе мероприятиях/проектах, в которых используются новые источники данных, и еще 4 страны изучают такую возможность. Таким образом, в целом новые источники данных для измерения миграции и трансграничной мобильности используют — или рассматривают возможность такого использования — около четверти стран.
48. Почему большинство стран не используют новые источники данных для измерения миграции и мобильности? Используют ли они новые источники данных в других областях статистики? Какие факторы могут повлиять — положительно или отрицательно — на использование новых источников данных для измерения миграции и мобильности?

3.1. Использование новых источников данных для подготовки статистических данных в других областях статистики

49. Странам, которые в настоящее время не используют новые источники данных для измерения миграции и трансграничной мобильности, в ходе опроса задавался вопрос, используют ли НСУ новые источники данных для формирования данных в других областях статистики. Из 24 стран, ответивших на этот вопрос, 8 стран (одна треть) сообщили об использовании — или апробировании использования — новых источников данных в других областях статистики, остальные ответили «нет» (9 стран) или «не знаю» (7 стран). Германия сообщила, что новые источники данных используются для получения экспериментальной статистики в других областях статистики (https://www.destatis.de/EN/Service/EXDAT/_node.html).
50. Франция и Новая Зеландия сообщили об использовании новых источников данных для ведения статистики внутренней миграции. В Новой Зеландии данные мобильной телефонной связи объединяются с административными данными о пересечении границ для оценки географического распределения лиц, прибывших из-за границы, в стране. Кроме того, для оценки внутренней миграции используются интегрированные административные данные различных государственных учреждений. Интегрированные административные данные используются также для дополнения и проверки информации переписи, проверки официальных оценок численности постоянного населения и оценки международной миграции в разбивке по этнической принадлежности.
51. Об использовании новых источников данных для получения статистических данных о внутренней миграции и ежедневных поездках на работу и обратно сообщили четыре страны. В Португалии для расчета показателей мобильности населения на региональном уровне используется программа «Фейсбук» «Data for Good». В Испании с использованием данных мобильной телефонной связи получены экспериментальные статистические данные о ежедневных поездках на работу и обратно. Швейцария сообщила о пилотном проекте по сбору данных о повседневной мобильности с помощью смартфонов: в рамках опроса выборочная совокупность участников устанавливает на свои смартфоны приложение, собирающее данные об их повседневной мобильности —

вместо сообщения о своих перемещениях с заполнением опросной анкеты. В Соединенном Королевстве научно-исследовательский центр интеллектуальной обработки данных⁵ использует различные новые источники данных, в том числе изучает вопрос использования данных «Фейсбук» о перемещениях по стране.

52. Шесть стран сообщили об опыте использования новых источников данных в различных областях экономической статистики. В области статистики цен Португалия использует веб-данные, Литва — данные супермаркетов, а Словакия — как веб-скрейпинг, так и данные сканеров розничных сетей. В сфере статистики туризма Словакия использует данные с веб-сайтов систем бронирования жилья, а в Испании для оценки расходов на туризм проведено исследование по использованию кредитных карт. Относительно международной торговли Португалия сообщила об исследованиях по вопросам использования интернет-поиска на конкретных сайтах для подтверждения статистических данных о международной торговле, в том числе об объемах и ценах международной торговли электроэнергией. Португалия также использует систему электронных счетов-фактур (E-factura) Налогового и таможенного управления для получения статистических данных о региональной экономической деятельности, а также возможности поиска вакансий и сайтов предприятий в Интернете для ведения статистики рынка труда. Наконец, во Франции новые источники данных используются для целей статистики потребления, а в Соединенном Королевстве — для получения более оперативных показателей состояния экономики.

3.2. Причины неиспользования новых источников данных

53. 24 странам, где для измерения миграции и трансграничной мобильности новые источники данных не используются, в ходе опроса задавался вопрос об основных причинах этого. В вопроснике предлагалось на выбор несколько возможных причин, также можно было вписать свой ответ, при этом респонденты могли указывать до трех причин (таблица 1).

Таблица 1. Причины неиспользования новых источников данных для статистики международной миграции

Причины	Страны (из числа 24 стран)
Соображения конфиденциальности индивидуальных данных, юридические ограничения и правила, ограничивающие использование больших данных и новых источников в статистических целях	12
Высококачественная статистика миграции уже выпускается с помощью традиционных источников данных	11
Отсутствие специалистов по работе с большими данными	8
Большие данные не подходят для использования в официальной статистике	6
Отсутствие договоренности между ведомствами (например, НСУ и отраслевым министерством) и потенциальными поставщиками услуг	3
Нехватка средств для получения доступа к большим данным	1
Замечания и/или другие причины	3

54. Большинство стран упомянули в качестве причины **соображения конфиденциальности индивидуальных данных, юридические ограничения и правила, ограничивающие использование больших данных и новых источников в статистических целях** (12 стран). Любопытно было бы получить более подробную информацию об этих опасениях и проблемах, а также о том, относятся ли они к конкретным источникам данных. Вопросы

⁵ <https://datasciencecampus.ons.gov.uk/>.

конфиденциальности и юридические ограничения могут охватывать разные вещи (этические и юридические причины). Однако, к сожалению, страны представили очень ограниченную информацию по этому вопросу. Польша уточнила, что основным препятствием являются юридические проблемы, поскольку использование данных операторов мобильной связи станет возможным только после внесения поправок в соответствующие нормативные документы.

55. 11 стран ответили, что они **уже выпускают высококачественную статистику миграции с помощью традиционных источников данных**. Некоторые из них уточнили, что они используют административные данные и считают, что новые источники данных не будут гораздо полезнее. Новая Зеландия указала, что с 2019 года (с оценками вплоть до 2001 года) для оценки международной миграции с высокой точностью и своевременностью используются административные данные о туристах, пересекающих границу. Кроме того, Новая Зеландия отметила, что надежное и непрерывное использование коммерческих/неправительственных источников данных, как правило, сопряжено с немалыми трудностями и ненамного эффективнее по сравнению с получением данных через интегрированную систему административных данных.
56. Несколько стран сослались на **отсутствие специалистов по работе с большими данными** (8 стран) и на то, что **большие данные не подходят для использования в официальной статистике** (6 стран). Некоторые страны представили дополнительную информацию о причинах, в силу которых, по их мнению, большие данные не подходят для использования в официальной статистике. В Словакии демографическая статистика и статистика миграции формируется путем проведения сплошных статистических обследований демографических изменений (естественное движение населения и миграция), а источники административных данных используются в качестве дополнительных; отделы Министерства внутренних дел Словацкой Республики являются подотчетными подразделениями по статистике миграции. Эти источники данных являются гарантией требуемого качества статистических данных. Аналогичным образом в Швейцарии данные о международных миграционных потоках или численности мигрантов черпаются в основном из административных данных или данных обследований, которые уже доступны при низких затратах и хорошем качестве. Эти данные используются для целей государственного управления и выработки политики в области миграции, и в более обширных данных о движении населения нет необходимости. Кроме того, в официальной статистике Швейцарии главное внимание уделяется населению, постоянно проживающему в стране на законных основаниях. Поэтому в ней не рассматриваются группы населения, не представленные в официальных реестрах, например мигранты без документов или краткосрочные мигранты, в случае которых использование новых данных, возможно, более актуально. В Испании Национальный институт статистики наладил контакты с операторами мобильной связи для получения данных о ежедневных поездках на работу и обратно (а также, в меньшей степени, о туризме), но ценной информации для статистики миграции не получено, других же подходящих источников больших данных выявлено не было. Соединенное Королевство отметило, что пока нет полной уверенности в полезности новых источников данных для статистики миграции. Прежде чем использовать эти данные при ведении статистики, необходимо проверить, могут ли они — и насколько точно — отображать миграционные потоки отдельно от других форм перемещения людей. В целом, при использовании больших данных может оказаться нелегко выделять мигрантов или миграцию в огромном массиве данных. Например, с помощью данных мобильной телефонной связи трудно выделить перемещения через границу, которые не являются просто посещениями или короткими поездками.
57. **Отсутствие договоренности между ведомствами (например, НСУ и отраслевым министерством) и потенциальными поставщиками** было указано в качестве причины тремя странами. В Чехии данные о миграции собираются из различных

административных источников, включая Министерство внутренних дел (Служба иностранной полиции), которое является одним из основных поставщиков данных. Было высказано мнение, что любая инициатива по использованию больших данных должна исходить от этого министерства, поскольку это потребует внесения изменений в Систему информации об иностранных гражданах.

58. **Нехватка средств для получения доступа к большим данным** не представляется серьезным препятствием, поскольку эту причину назвала только одна страна.

4. Масштабные нововведения в области измерения международной миграции

59. Как подробно показано в предыдущих разделах настоящего доклада, большинство НСУ, принявших участие в опросе (об использовании альтернативных источников данных в статистике миграции), указали, что большие данные в настоящее время не используются.
60. Однако 12 НСУ указали, что они «работают над масштабными нововведениями в области измерения международной миграции», включая «новые подходы с использованием существующих источников и/или с изучением новых источников данных». В данном разделе обобщены выводы из этих ответов в разбивке по двум темам:
- изменения в источниках данных по сравнению с ранее использовавшимися в НСУ;
 - новая или инновационная методология, которую НСУ используют в отношении этих или существующих данных.

4.1. Нововведения в измерении международной миграции, связанные с данными

61. Большинство стран, в которых изучаются потенциальные нововведения в области данных, начали использовать текущие оценки, расширять масштабы их использования или улучшать их с помощью административных данных (также называемых админданными). Из упомянутых выше респондентов восемь указали, что они в той или иной мере работают с административными данными, достигнув различных уровней зрелости и прогресса на этом пути.
62. Три страны сообщили, что для оценки миграции, определения ее демографической структуры или численности населения они используют метод обследований. В некоторых случаях дополнительно применяются показатели, рассчитанные на базе административных данных.

4.1.1 Краткая информация об использовании административных данных

63. Страны достигли разных уровней зрелости при использовании административных данных (таблица 2). В данном разделе для того, чтобы наглядно показать различные стадии развития и типы подходов, которые изучаются в сфере административных данных, использованы примеры стран, основанные на подробных ответах, полученных в ходе опроса.

Таблица 2. Резюме использования и изучения административных данных для оценки международной миграции в странах-респондентах

Степень зрелости использования административных данных	Страна	Подробная информация об использовании
Изучение/ранняя стадия разработки	Франция	В настоящее время осуществляется проект по использованию и объединению административных данных для оценки миграции
	Португалия	Изучение административных данных для целей ежегодных оценок численности постоянного населения и достижение согласованности миграционных потоков с численностью мигрантов
Зрелая степень разработки/отработка процесса	Армения	При проведении переписи 2022 года планируется использовать регистр населения и данные пограничного контроля. После переписи населения эти источники могут быть использованы для измерения миграции
	Венгрия	Улучшение статистики миграции на основе многочисленных источников административных данных (в основном регистр населения и адресов, регистр иностранных граждан, регистр социального обеспечения).
	Словакия	Перепись 2021 года проводилась на основе интеграции административных данных с дополнительными данными о миграции из статистических обследований
	Соединенное Королевство	Переход от статистики миграции на основе обследований к новым показателям на базе административных данных
	Соединенные Штаты	Использование полных данных об авиарейсах для оценки чистой миграции в Содружестве Пуэрто-Рико, дополняемых демографическими характеристиками из данных обследований
В работе	Новая Зеландия	Административные данные используются в качестве основного источника для оценки миграции, статистические модели применяются для заблаговременного получения предварительных оценок
	Швейцария	Очень развитое и широко распространенное использование административных данных при ведении статистики. В настоящее время прорабатываются возможности выпуска продольных данных о международных миграционных потоках на основе уже подготовленных данных регистров

4.1.2 Отдельные примеры нововведений с использованием административных данных

Португалия

64. Статистическое управление Португалии изучает и исследует возможность использования административных данных при составлении официальной статистики населения на основе различных источников, включая:
- национальный регистр актов гражданского состояния населения;
 - регистр иностранного населения;
 - регистр достигнутого уровня образования;
 - налоговый регистр;
 - регистр социального обеспечения;
 - регистры занятых и безработных.
65. Эти данные учета используются для создания базы данных о постоянном населении, которая будет ежегодно обновляться. При переходе к использованию административных данных ключевой проблемой является согласованность данных о миграционных потоках и численности мигрантов. Для измерения численности мигрантов можно использовать метод сравнения численности постоянного населения между контрольными датами, но при этом невозможно получить валовой показатель миграционного потока между этими контрольными датами. В настоящее время Статистическое управление Португалии изучает методы, с помощью которых можно было бы справиться с проблемами и ограничениями административных источников данных для измерения миграционных потоков.

Армения

66. Для целей переписи населения в 2022 году Армения планирует использовать регистр населения в сочетании с данными из информационной системы пограничного контроля. Впервые эти административные данные будут использоваться в сочетании с большим выборочным обследованием (с охватом 25 % адресов из регистра населения). Такой подход к переписи населения потенциально позволяет получить новые показатели миграции и мобильности. Разработка систем объединения административных данных и упомянутого обследования находится на завершающей стадии.

Соединенное Королевство

67. Управление национальной статистики (УНС) переходит к измерению миграции на основе административных данных вместо показателей, опирающихся на обследование международных пассажиропотоков (ОМП) с присущими им давно признанными недостатками. Этот процесс дополнительно ускорился в связи с пандемией COVID-19, из-за которой в 2020 году УНС приостановила проведение ОМП, хотя сейчас эта работа возобновилась с уделением повышенного внимания международным пассажирам. УНС использует различные административные источники, в основном интегрированные в рамках базы данных регистрации и учета доходов и пособий населения (RAPID)⁶, разработанной Министерством труда и пенсий (МТП). Это позволяет получить единое согласованное представление о взаимодействии систем МТП, Королевской налогово-таможенной службы (КНТС) и местных органов через систему жилищных льгот. В рамках такого взаимодействия охватываются пособия для малоимущих, доходы занятого и самозанятого населения, пенсии и пособие для работающих лиц с низким доходом.

⁶ В базе RAPID содержатся записи о всех лицах, имеющих номер национального страхования. Подробнее см. [International migration: developing our approach for producing admin-based migration estimates: Measuring international migration with administrative data.](#)

68. Коррективы вносятся для учета различных типов мигрантов, которые по-разному представлены в административной системе. Для решения проблемы лага, связанного с классификацией миграции на основе этих данных, УНС также прибегает к моделированию для оценки миграционных потоков, о чем говорится в разделе [Национальный опыт использования больших данных и новых источников данных](#).

4.1.3 Краткая информация об использовании данных обследований

69. В таблице 3 приводится краткая информация о том, как данные обследований используются или изучаются в странах, представивших ответы.

Таблица 3. Подготовленное на основе полученных ответов резюме применения и изучения данных обследований в странах-респондентах, с разбивкой по предполагаемому этапу зрелости проработки и использования административных данных

Степень зрелости использования данных опроса	Страна	Подробная информация об использовании
Зрелая степень разработки/отработка процесса	Армения	Для дополнения административных данных на основе переписи населения 2022 года будет использовано 25-процентное выборочное обследование (см. подраздел «Армения» в предыдущем разделе)
	Чехия	Данные о миграции из Обследования рабочей силы, включая дату миграции и место рождения родителей
	Соединенные Штаты	Обследование позволяет получать половозрастные характеристики потоков мигрантов для Содружества Пуэрто-Рико

4.2. Нововведения в измерении международной миграции, связанные с методологией

70. Некоторые страны также подробно рассказали о методологических разработках, проводимых независимо от нововведений в сфере данных. Подробная информация была представлена Новой Зеландией и Соединенным Королевством, и она будет использована в качестве кратких примеров различных тестируемых методологических нововведений.

4.2.1 Отдельные примеры методологических нововведений

Новая Зеландия

71. Новая Зеландия собирает информацию о прибывающих и отъезжающих на основе паспортного контроля и информации об авиаперевозках по всем пассажирам. С использованием этих данных разработан показатель миграции по факту, с помощью которого иммигранты и эмигранты классифицируются по продолжительности времени,

проведенного путешественником в стране или за ее пределами⁷. Как правило, это требует наблюдения за пересечением границ мигрантами в течение 16 месяцев, что неизбежно влечет за собой задержку в получении своевременных окончательных оценок миграции. Поэтому Статистическое управление Новой Зеландии использует модель прогнозного машинного обучения для получения предварительных оценок миграции на основе прогнозной классификации данных о пересечении границ.

72. В результате изучения миллионов записей за прошлые периоды модель позволяет выявить характерные особенности пересечения границы, в силу которых можно с большей или меньшей вероятности считать, что речь идет о пересечении границы мигрантами. В данной модели анализируются «особенности» в данных о пересечении границы за прошлые периоды, такие как направление и дата пересечения, количество времени, проведенное в стране и за ее пределами, количество времени, прошедшего с момента пересечения границы. Она также позволяет узнать из паспортных данных такие параметры, как возраст, пол и гражданство лиц, совершающих поездку, и тип обычно имеющихся у них виз. Кроме того, в каждой из записей также содержится информация о том, совершалось ли конкретное пересечение границы мигрантом (результат). Благодаря этому модель позволяет установить связь между различными сочетаниями вышеуказанных характеристик и вероятностью того, что конкретном случае границу пересекает мигрант. Для обучения в модели используется метод градиентного бустинга. Для этого используется хорошо отработанный алгоритм «XGBoost».
73. В результате этой процедуры создается комплекс предикторов, которые могут быть применены к случаям пересечения границы, где результат неизвестен, на основе отмеченных при этом характерных особенностей. Используя установленные уровни вероятности, оценивается количество пересечений границ мигрантами за определенный базисный период, агрегированное по различным признакам, включая возраст, пол и страну предыдущего проживания.

Соединенное Королевство

74. Как кратко обсуждалось выше, Управление национальной статистики (УНС) переходит на административные данные для измерения миграции. Миграция оценивается на основе показателей занятости в рамках имеющихся данных. В целом лица, прибывшие, имея 12 месяцев занятости, классифицируются как мигранты, хотя при этом требуются некоторые исключения и корректировки (подробности приведены в сноске 6 выше). Помимо поправок на охват и другие факторы, УНС также разрабатывает методологию моделирования для повышения своевременности оценок миграции, учитывая, что, согласно текущему стандартному определению, классификация доступна не ранее, чем через 12 месяцев вследствие того, что при сборе административных данных период наблюдения за сроками занятости составляет 12 месяцев⁸. В УНС, правда, изучается возможность использования более широкого набора определений с учетом стандарта ООН. Они призваны обеспечить показатели измерения численности населения в режиме реального времени с учетом их возросшей востребованности среди пользователей во время пандемии.
75. В Соединенном Королевстве изучается возможность использования в дополнение к подходу, основанному на административных данных, моделей пространства состояний, которые, возможно, позволят давать предварительный прогноз миграции. Эти модели до сих пор использовались для оценки миграции до декабря 2020 года⁹. Как указано в

⁷ Подробнее о конкретных правилах классификации см. [Migration Data Transformation: The 12/16-month rule](#).

⁸ Подробнее о своевременности методов и последних разработках см. [International migration statistical design progress report](#).

⁹ Основные результаты последних модельных оценок представлены в документе [Long-term international migration, provisional: year ending December 2020](#).

публикациях УНС, имеется в виду изучить использование таких моделей для повышения своевременности оценок миграции на основе административных данных, о которых говорилось ранее¹⁰.

5. Справочная система по специальной литературе: база данных по нововведениям в статистике миграции (DIMiS)

76. В научных журналах или в виде рабочих документов опубликован ряд работ по миграции с использованием инновационных источников данных; эти публикации представляют потенциальный интерес для специалистов по официальной статистике. Для преодоления разрыва между исследованиями и практикой Целевая группа отобрала некоторые из этих работ и при поддержке секретариата ЕЭК ООН создала базу данных¹¹, в которой даются ссылки на опубликованные работы по вопросам миграции с использованием инновационных методов и/или новых источников данных. **База данных по нововведениям в статистике миграции¹² (DIMiS)** призвана повысить осведомленность о научном прогрессе в этой области и облегчить доступ к соответствующей информации для специалистов по официальной статистике и практических работников, интересующихся данной темой. DIMiS должна охватывать исследования, в которых используются «нетрадиционные» источники данных, а также инновационные методы, применяемые к «традиционным» источникам данных, которые актуальны при ведении статистики миграции (в широком смысле).
77. Для этой первоначальной версии базы данных было определено около ста работ по таким темам, как международная миграция, внутренняя миграция, мобильность населения, перемещение населения и распределение населения. Из этих документов в первый список были включены только те, которые касаются миграционных потоков (как международных, так и внутренних). В зависимости от интересов пользователей в будущих версиях DIMiS спектр тем может быть расширен или дополнительно детализирован (например, с разграничением данных о потоках и численности мигрантов, конкретно по лицам, ищущим убежища, и беженцам, с учетом интеграции мигрантов и т. д.).
78. Структурно DIMiS организована в соответствии с набором ключевых переменных, которые могут быть использованы в качестве фильтра (фильтров):
1. **Автор (авторы):** имя автора (имена авторов). В имени первого автора сначала указывается фамилия, затем начальная буква (инициалы) второго и последующих имен. Для всех остальных авторов сначала указывается начальная буква имени, а затем полная фамилия. Перечень авторов приводится полностью и соответственно не сокращается с помощью «et al.». Поэтому можно либо искать конкретную фамилию, либо сортировать статьи по фамилии первого автора.
 2. **Год:** год публикации. Если год публикации в Интернете отличается от года выхода в свет печатной публикации, предпочтительно использовать самый последний год, поскольку он может соответствовать версии, включающей исправления. Для не ре-

¹⁰ Обсуждение последствий и ограничений моделирования, а также вопросов будущего использования см. [Using statistical modelling to estimate UK international migration: Discussion](#).

¹¹ Обзор отдельных научных работ см. доклад Объединенного исследовательского центра Европейской комиссии (ОИЦ) о науке для политики «*Data Innovation in Demography, Migration and Human Mobility — Data gaps, opportunities, and challenges of non-traditional data*», Iacus, S.M., C. Bosco, S. Grubarov-Boskovic, U. Minora, F. Sermi and S. Spyrtatos (готовится к выпуску).

¹² <https://statswiki.unece.org/display/DIMiS>

цензурируемых коллегами статей, например для статей в SocArXiv (открытый архив общественных наук), берется год первой загрузки/ рассылки, так как с течением времени статья может несколько раз подвергаться изменениям, в том числе и после включения в DIMiS.

3. **Название:** название статьи. Единого стиля в отношении использования заглавных букв, копирующего принятый в статьях, нет.
4. **Журнал:** научный журнал/книга/исследовательский проект/хранилище веб-сайта, в котором содержится документ. Единый стиль не применяется.
5. **Аннотация:** опубликованное краткое изложение цели, метода и результатов статьи — при наличии.
6. **Тема (темы):** основной аспект, связанный с миграцией, обсуждаемый в статье/докладе. Для этой первоначальной версии базы данных были рассмотрены следующие категории:
 - a) международная миграция;
 - b) внутренняя миграция¹³;
 - c) мобильность населения;
 - d) перемещение населения;
 - e) картирование/распределение населения;
 - f) прочие.

В DIMiS пока включены только те документы, тема которых охватывает первые две категории, хотя в них могут затрагиваться и другие темы. Несколько тем, рассматриваемых в одной работе, перечисляются последовательно в нужной графе в соответствии с приведенным выше списком. Поэтому невозможно найти работы, относящиеся исключительно к какой-либо категории (категориям), кроме a) или b), но можно найти работы, охватывающие также категории c) и/или последующие.

7. **География:** страна или страны, охваченные исследованием, с указанием в скобках соответствующей географической области (областей) на субнациональном уровне. В случае чисто методологических документов или обзорных статей, в которых анализируется состояние научной литературы, этот ключ не применяется, и в данной графе указывается «не применимо».
8. **Источники данных:** источник данных, использованных в статье. Предусмотрены следующие категории:
 - a) оператор(ы) мобильной связи (ОМС);
 - b) социальные сети (например, Twitter, Facebook, WhatsApp, LinkedIn);
 - c) поисковые системы (например, Google);
 - d) веб-демография (базы персональных микроданных, доступные во Всемирной паутине);
 - e) регистры (административные данные);
 - f) обследования (например, перепись населения, выборка домашних хозяйств);
 - g) прочие (например, электронная почта, интернет-провайдеры, информация о трафике, спутниковые снимки, конкретные базы данных).

Как правило, источник данных имеет свою собственную категорию, если он используется как минимум в двух статьях. В статье может использоваться один или несколько источников данных, все они указываются в этой графе. По возможности конкретный источник данных указывается в скобках (например, Facebook в категории «Социальные сети»). Затем пользователи могут начать поиск по всем статьям с

¹³ Для целей настоящего доклада миграция между США и Пуэрто-Рико классифицируется как внутренняя миграция.

использованием категории источника данных (например, ОМС) или конкретного источника (например, Twitter).

79. Из 52 статей, включенных в эту первоначальную версию DIMiS, 36 статей посвящены международной миграции (отдельно или охватывая также другие темы), 23 — внутренней миграции (отдельно или охватывая также другие темы) и 7 статей охватывают как международную, так и внутреннюю миграцию (см. таблицу 4).

Таблица 4. Статьи в DIMiS в разбивке по темам

Тема	Количество статей
Внутренняя миграция	9
Внутренняя миграция/мобильность населения	4
Внутренняя миграция/перемещение населения	3
Международная миграция	25
Международная миграция/мобильность населения	1
Международная миграция/внутренняя миграция	3
Международная миграция/внутренняя миграция/мобильность населения	4
Международная миграция/перемещение населения	3
Итого	52

80. As for the geography, the United States of America¹⁴ is the prominent country in terms of applications (14 papers), but European countries, and particularly the EU Member States, are well represented. There are also 5 articles with African countries (Kenya, Namibia, Rwanda, and Senegal) and 6 papers whose application is on all countries of the world (even if not all results are reported).

Таблица 5. Статьи в DIMiS в разбивке по географии

География	Количество статей
Бельгия, Франция, США	1
Канада	1
Китай	2
Эстония	1
Государства-члены ЕС	2
Государства-члены ЕС, Исландия, Норвегия, Швейцария, Великобритания	1
Государства-члены ЕС, Ирак, Сирия	1
Государства-члены ЕС, Норвегия, Швейцария	1
Государства-члены ЕС, Норвегия, Швейцария, Великобритания	1
Франция	1
Франция, Германия, Ирландия, Италия, Нидерланды, Испания, Великобритания	1
Кения	1
Намибия	1
Государства-члены ОЭСР	2

¹⁴ Для целей настоящего доклада Пуэрто-Рико считается частью США. Также см. сноску 12.

Португалия	1
Румыния, Великобритания	1
Руанда	2
Сенегал, США	1
Швейцария	1
Турция	1
Великобритания	1
США	4
США (Калифорния, Техас)	1
США (Пуэрто-Рико)	3
США (Техас)	2
Венесуэла	2
Все страны	6
Не применимо	9
Итого	52

81. Среди статей, перечисленных в DIMiS, наиболее часто используемыми источниками данных являются социальные сети (21 статья), в частности Facebook и Twitter с 6 и 4 научными статьями, соответственно (см. Таблицу 6). Данные от операторов мобильной связи также широко используются (14 статей), но, в отличие от социальных сетей, в основном для внутренней миграции. Интересно отметить, что есть также работы, в которых совместно используются «нетрадиционные» и «традиционные» (например, регистры и обследования) источники данных или применяются инновационные методы для первых. Диапазон «нетрадиционных» источников данных выходит далеко за рамки обычных операторов мобильной связи и социальных сетей, что также показано в категории «Другие».
82. DIMiS публикуется в предварительном формате, открытом для дальнейших улучшений, и доступен по адресу <https://statswiki.unecse.org/display/DIMiS>. В частности, по мере появления в будущем нового опыта потребуется механизм для добавления новых статей. Отзывы пользователей DIMiS будут иметь решающее значение для превращения этой первоначальной версии в справочный инструмент, который будет полезен для официальных статистиков и потенциальных специалистов по большим данным в целом.

Table 6. Статьи в DIMiS по источнику данных

Источник данных	Количество статей
Оператор мобильной связи	9
Оператор мобильной связи / Социальные сети	4
Оператор мобильной связи / Обследование (перепись)	1
Поисковая(ые) система(ы)	3
Поисковая(ые) система(ы) (Google)	2
Поисковая(ые) система(ы) (Google) / Регистры (незаконное пересечение границы, прошения о предоставлении убежища) / Другое (Глобальная база данных событий, языка и настроения)	1
Социальные сети	7
Социальные сети (Facebook)	4
Социальные сети (Facebook, LinkedIn)	1
Социальные сети (Facebook, Twitter)	1

Социальные сети (LinkedIn)	1
Социальные сети (Twitter)	3
Веб-демография	2
Регистры (незаконное пересечение границы, прошения о предоставлении убежища) / Другое (Глобальная база данных событий, языка и настроения)	1
Регистры/Обследование (Обследование рабочей силы)	1
Обследование (набор данных о потребительских кредитах)	1
Другое (большие данные в целом)	5
Прочее (данные об авиапассажирах)	1
Другие (Microsoft Academic Knowledge Graph)	1
Прочее (услуга мобильных денежных переводов)	1
Другие (электронные письма Yahoo!)	1
Другие (пользователи Yahoo!)	1
Итого	52

6. Выводы и последующие шаги

83. Опыт НСУ в области **использования** больших данных **ограничен**, и собранного для настоящего доклада материала недостаточно для предоставления конкретных рекомендаций. Вместе с тем **потенциал использования новых источников данных** в плане решения и существующих, и возникающих проблем миграции и удовлетворения неотложных потребностей в данных весьма перспективен. Обмен соответствующей **информацией и инструментами** и их обновление **облегчают использование** больших данных и новых источников данных для измерения миграции и трансграничной мобильности.

6.1. Пока ограниченные масштабы использования

84. Многие страны сталкиваются с ограничениями в использовании больших данных; в частности, в качестве причины неиспользования новых источников данных многие страны упомянули соображения конфиденциальности индивидуальных данных, юридические ограничения и существующие правила. Для решения этой проблемы необходимо рассматривать в международных дискуссиях этические вопросы и вопросы конфиденциальности. В то же время важно применять методологии и методы, обеспечивающие сохранение конфиденциальности информации, позволяя специалистам в области статистики и исследовательским работникам использовать потенциал больших данных. Однако такое сочетание как использование инновационных источников данных, например данных социальных сетей, и сохранение тайны личной жизни и общественного доверия может вызвать сложности.
85. В настоящее время многие страны предпочитают направлять свои усилия на использование административных данных и интеграцию данных из различных источников, что, возможно, связано с трудностями, возникающими при доступе к большим данным. Источники административных данных могут использоваться более полно. Например, лица, ищущие убежища, и нелегальные иммигранты могут не учитываться в официальной миграционной статистике, иногда из-за отсутствия идентификационных номеров. Устранение этих ограничений в регистрационных данных повысит качество миграционной статистики. В случае же административных данных то, что является новым источником данных для одной страны, порой является традиционным источником для другой. Возможности использования зависят от

множества различных аспектов, таких как особенности исторического развития стран, правовые и институциональные рамки и имеющиеся ресурсы.

86. Даже в странах, которые сообщили о ряде случаев использования больших данных в других областях статистики (туризм, ежедневные поездки на работу и обратно, экономика и т. д.), в области измерения миграции большие данные пока используются в ограниченных масштабах. Это объясняется многими причинами. Многие страны имеют долгую общую историю ведения статистики миграции и часто делятся опытом использования международных стандартов и определений с международными организациями. По данным проведенного опроса, эти страны, как правило, предпочитают повышать качество традиционных данных, а не изучать новые источники данных, которые требуют преодоления множества административных и юридических препятствий.

6.2. Потенциальные преимущества

87. Анализ, проведенный Целевой группой, подтверждает потенциал новых источников данных для измерения миграции, включая возникающие тенденции в динамике миграции и трансграничной мобильности, а также для удовлетворения срочных потребностей в данных, возникающих в результате неожиданных природных (пандемии, ураганы и т. д.) и геополитических (войны, преследования, политические кризисы и т. д.) событий.
88. Большие данные и новые источники данных могут помочь преодолеть разрыв между потребностью в своевременной информации и временем, которого требует процесс сбора официальной статистики. Поэтому они могут сыграть важную роль в повышении качества выпускаемых в настоящее время статистических данных. Многие страны переходят к интегрированным статистическим системам, и новые источники данных могут стать важными ресурсами для включения в эти системы. Поэтому можно надеяться, что в новых интегрированных статистических системах измерения миграции большие данные и новые источники данных найдут достойное место и смогут повысить качество статистики миграции.

6.3. Следующие шаги — ведение Базы данных по нововведениям в статистике миграции

89. В настоящем докладе показана текущая работа по использованию больших данных и новых источников данных в НСУ, которые, как ожидается, будут полезны для стран, проявляющих интерес к деятельности в этой области.
90. Целевая группа создала Базу данных по нововведениям в статистике миграции (DIMiS) для расширения горизонтов работы путем включения исследований и изысканий в области миграции и больших данных, проводимых вне рамок НСУ, таких как статьи и документы, посвященные большим данным для статистики миграции (включая вопросы международной миграции, мобильности населения, внутренней миграции и т. д.). База данных облегчает обмен информацией о проектах, интеграции данных, подходах к моделированию и передовой практике, что позволит НСУ учитывать общемировой опыт исследований в различных областях изучения миграции.
91. НСУ и исследовательским работникам предлагается сообщать ЕЭК ООН о любом новом опыте и публикациях, чтобы можно было регулярно обновлять созданную базу данных. Не являясь исчерпывающей, база данных DIMiS задумана как первый шаг к созданию динамично развивающегося хранилища данных, которое должно постоянно пополняться новым опытом, чтобы облегчить обмен информацией о проектах, основанных на новых источниках данных.

92. Эта база данных должна быть гибкой. Включенная в нее информация может меняться и увеличиваться в объеме по мере расширения и усложнения накопленного опыта, что позволит DIMiS стать ключевым справочным инструментом, помогающим НСУ использовать новые источники данных и связанный с этим метод.
-