

Применение административных данных в традиционных переписях населения

Выбор тем при международных переписях населения¹

Опубликовано в мае 2022 года

ВВЕДЕНИЕ

Традиционно переписи населения и домашних хозяйств основывались на таких важных особенностях, как опросы физических лиц, одновременность, универсальность, данные по небольшой области и периодичность (рисунок 1). Однако за последние несколько десятилетий новые методологии расширили это определение. В таких новых методах используются административные данные, которые служат либо для замены, либо для поддержки данных, полученных при опросе населения во время переписи (UNECE, 2015). Поскольку национальные

¹ Настоящая техническая записка является частью серии «Избранные темы международных переписей населения» (STIC, в соответствии с английским акронимом), в которой рассматриваются вопросы, представляющие интерес для международного статистического сообщества. Бюро переписи населения США помогает странам совершенствовать национальные системы статистики, содействуя устойчивому расширению статистических компетенций. Любые выраженные мнения отражают точку зрения автора(-ов) и не обязательно совпадают с позицией Бюро переписи населения США.

статистические управления (NSO, в соответствии с английским акронимом) рассматривают возможность переопределения переписей населения и домашних хозяйств с учетом этих методологий, крайне важно, чтобы они как можно точнее придерживались перечисленных выше важных особенностей.

Источники административных данных (ADS, в соответствии с английским акронимом) — это наборы данных, которые содержат информацию, собираемую в основном для административных (а не исследовательских или статистических) целей (UNECE, 2018). Эти данные собираются государственными ведомствами, общественными органами и другими организациями в целях регистрации, обработки и учета данных о физических лицах, предприятиях и домашних хозяйствах. При переписях такие данные могут применяться по-разному. В некоторых странах они служат дополнением к традиционной переписи, в других — для проведения комбинированной переписи, в третьих

Рисунок 1.

Пять важных особенностей традиционной переписи населения



Источник: United Nations Economic Commission for Europe, 2015 and 2018.

– для проведения переписи, полностью основанной на реестрах (поле 1). Что касается переписи населения 2020 года, то из 113 ответивших стран, которые находятся не в Европе или Северной Америке, 12,4 процента сообщили, что при переписи использовались или будут использоваться ADS (UNSD, 2020).

В данной технической записке основное внимание уделяется различным типам ADS, которые могут использоваться в дополнение к традиционной переписи, типовому подходу к интеграции ADS, а также различным проблемам и возможностям, связанным с применением ADS. Таким образом, данная записка поможет NSO включать административные данные в перепись населения и домашних хозяйств для получения более качественных результатов.

ОСНОВНЫЕ ТИПЫ ADS

Многие государственные учреждения собирают и хранят административные данные различных типов, которые создаются в ходе их обычной деятельности. Такие данные можно потенциально использовать как на этапе составления списков для опроса, так и на этапе обработки результатов последнего. В таблице 1 представлены типовые формы ADS (разнообразные реестры). Реестры представляют собой структурированные списки единиц (например, физических лиц, предприятий) с соответствующими атрибутами (например, пол, тип отрасли), которые

Поле 1.

Три типа переписей

Традиционные

Прямой и исчерпывающий подсчет всех людей, домохозяйств и жилищных единиц с одновременным сбором данных об их характеристиках.

Комбинированные

Сочетание источников административных данных (ADS) с прямым подсчетом.

На основе реестра

Традиционный сбор данных заменяется на получение административных данных из различных регистров (то есть без опроса населения).

Источник: United Nations Economic Commission for Europe, 2018.

регулярно обновляются (например, ежегодно) (UNECE, 2011). Такие источники официально собираемых и хранимых административных данных помогают выполнять различные правительственные функции, но могут применяться и для повышения качества собранных в результате переписи данных (UNECE, 2018 and 2021).

Таблица 1.

Различные общие типы источников административных данных (ADS)

Тип ADS	Описание	Применение
Реестры населения	Формируется для получения данных о людях, живущих в стране (например, о дате и месте рождения, поле, дате прибытия/отъезда, гражданстве, семейном положении).	Население и его перемещения на национальном и субнациональном уровнях.
Реестры социального страхования	Формируются для национальной программы социального страхования в целях администрирования и распределения ресурсов.	Пол, возраст, семейное положение, доход, состояние инвалидности/здоровья.
Налоговые реестры	Формируются для администрирования и сбора подоходного и других налогов.	Состояние занятости, род занятий, место работы, место жительства.
Реестры юридических лиц	Формируются для хранения данных о юридических лицах и контроля за ними.	Экономические вопросы, особенно связанные с промышленностью.
Реестры трудоустройства	Формируются для получения сведений о занятости и безработице.	Экономическая деятельность, занятость, род занятий, стаж и место работы.
Образовательные реестры	Формируются для получения сведений об успеваемости студентов и занятости преподавательского состава.	Посещаемость занятий, грамотность, наивысший уровень образования.
Реестры здравоохранения	Формируются для предоставления услуг здравоохранения (например, частная система здравоохранения на основе страхования или национальная система здравоохранения).	Состояние здоровья, уровень инвалидности, деторождение.
Реестр зданий и проживания	Формируется местными службами по оценке земли/собственности, а также местными властями, ответственными за городское планирование и политики в области постройки зданий.	Собственность, размер, физическая конструкция домов.
Реестр адресов	Формируется для составления списков мест проживания людей, определения местоположения, а также для связи с реестром населения. (Примечание: в некоторых странах этот реестр связан с реестром зданий и проживания).	Домовладения и жилищные единицы.

Источник: United Nations Economic Commission for Europe, 2018 and 2021; United Nations Statistics Division, 2022 draft.

ПРЕИМУЩЕСТВА ПРИМЕНЕНИЯ ADS ПРИ ТРАДИЦИОННЫХ ПЕРЕПИСЯХ

Многие NSO все больше заинтересованы в расширении возможностей применения ADS (поле 2). Это решение мотивировано несколькими ключевыми преимуществами, которые дает применение ADS при традиционной переписи.

Снижение затрат на душу населения, повышение эффективности

Самая дорогая перепись — это традиционная перепись, при которой все данные о людях, домашних хозяйствах и жилищных единицах собираются путем непосредственного опроса специально обученными переписчиками. Во многих странах традиционные переписи с полным поквартирным опросом могут обходиться примерно в два раза дороже годового бюджета NSO. NSO могут сократить время и затраты на проведение переписей, используя данные, уже имеющиеся в ADS. Практика показывает, что, сокращая объем опросов людей за счет интеграции ADS, NSO могут сократить расходы на 22 процента по сравнению с традиционной переписью. Наибольшие возможности для снижения затрат имеются при проведении переписей, полностью основанных на реестрах. При переписи 2010 года медианные затраты на проведение традиционной переписи составили 5,57 доллара США (7,04 доллара США в долларах с поправкой на инфляцию 2022 г.) на душу населения, тогда как медианные затраты на проведение переписи, полностью основанной на реестрах составили 0,24 доллара США (0,30 доллара США) (UNECE, 2014 and 2018).

Меньше вопросов, оставшихся без ответа, меньше нагрузка на респондента

Когда ADS обеспечивает полный охват всего населения, эти данные можно использовать для создания и обновления основной базы ответов, предварительного заполнения анкет, восполнения данных, которые не были получены из-за отсутствия ответа на тот или иной вопрос (Mule, 2021; U.S. Census Bureau, 2021). ADS (например, налоговые декларации за предыдущие годы) содержат более точную информацию, чем та, которую NSO может получить путем условной подстановки — статистического метода, который позволяет восполнить недостающие данные на основе других доступных сведений. Благодаря интеграции ADS респондентам потребуется сообщать меньше сведений о себе, своих домохозяйствах или жилье, что снижает нагрузку на респондентов (UNECE, 2018).

Повышение качества данных и расширение покрытия

NSO могут использовать ADS для подтверждения предоставленных респондентами данных. Такие проверки качества позволяют повысить качество результатов переписи до максимума. В то же время в странах с актуальными почтовыми и строительными реестрами ADS позволяют расширить покрытие за счет обновления базы адресов переписи (U.S. Census Bureau, 2020).

НЕОБХОДИМЫЕ УСЛОВИЯ ДЛЯ ИНТЕГРАЦИИ ADS

Прежде чем интегрировать ADS, при планировании переписи необходимо выполнить несколько важных условий (более подробные сведения приведены в документе UN Guidelines on Registers-Based Censuses, UNSD, 2022 draft).

Правовая и политическая база

NSO должны сотрудничать с правительствами, чтобы подготовить правовые положения, необходимые для доступа к административным данным и обеспечения их защиты. В идеале правительству следует принять дополнительные положения, предоставляющие NSO полномочия по созданию, изменению (или удалению) и применению данных (и метаданных), которые будут использоваться в статистических реестрах. (UNECE, 2011)

Организационная структура и сотрудничество

NSO также следует сотрудничать с органами, которые отвечают за хранение данных в частном и государственном секторах, в целях создания структуры, подробно описывающей следующее:

- Ключевые участники процесса передачи данных.
- Стандарты для метаданных, частота обмена данными, требования к качеству данных и метаданных, правила обеспечения конфиденциальности.
- Оплата (желательно бесплатно).
- Срок действия соглашения.
- Непредвиденные обстоятельства для изменений.
- Порядок разрешения споров.

Такая структура не заменяет прозрачных и открытых рабочих отношений с другими органами, которые отвечают за хранение данных. Открытые рабочие отношения должны поддерживаться для своевременного предоставления держателями качественных данных (UNECE, 2011 and 2018).

Заинтересованные стороны и общественное одобрение

Отношение к применению административных данных в разных странах разное (UNECE, 2011). Например, некоторые люди не верят в то, что их конфиденциальность будет сохранена, поскольку данные передаются между различными организациями. Поэтому такие люди категорически отказываются от применения ADS. Чтобы избежать путаницы, недоверия и получить общественное одобрение, NSO должны обеспечить прозрачность, то есть информировать общественность о том, как будут использоваться данные (например, публиковать четкие ограничения и правила использования). При этом пользователи данных информируются о любых изменениях в правилах доступа к данным конкретных типов (UNECE, 2018).

Единая система идентификации и надежная система регистрации статистических данных

Интеграция ADS (рисунок 2) требует времени для установления связи между данными о физических лицах, которые хранятся в различных реестрах, и данными, полученными при переписи. В идеальном случае NSO могут связывать такие данные при помощи уникальных идентификаторов, которые определяют отдельных людей и домохозяйства в разрозненных источниках, что значительно облегчает проведение переписей на основе реестров (UNECE, 2018). Эти реестры, безусловно, должны быть всеобъемлющими и надежными (то есть содержать точные, релевантные и актуальные данные): только при этом условии NSO смогут интегрировать данные в структуру переписи. Однако, как правило, перед интеграцией данные должны пройти тщательный процесс преобразования.

МЕТОДЫ И ОБРАБОТКА ДЛЯ ПРОВЕДЕНИЯ КОМБИНИРОВАННОЙ ТРАДИЦИОННОЙ ПЕРЕПИСИ

Поскольку многие страны переходят от традиционных переписей к комбинированным, а также к переписям на основе реестров, для оптимизации процесса интеграции может оказаться полезной общая схема (рисунок 3) (UNECE, 2018). В этом разделе кратко описаны пять ключевых процессов, которые показаны на рисунке 3.

Идентификация и сбор источников данных

Потенциальные, перспективные и актуальные данные должны быть определены в ходе встреч между NSO и поставщиками административных данных. В этом процессе NSO должны быть движущей силой, центром хранения и передачи данных в этой развивающейся сети

соединений. Такие соединения являются ключевыми, поскольку механизмы первоначальной передачи (то есть сбора) и обновления данных должны быть разработаны до начала интеграции.

Связывание записей

Связывание записей является одним из наиболее важных процессов при применении ADS во время переписи. При отсутствии в ADS уникальных (общих) идентификаторов NSO могут использовать для связи между источниками общие идентифицирующие переменные (совпадающие ключи), такие как имя, пол и дата рождения (UNECE, 2011). Ключи следует выбирать на основе их уникальности или «различительной способности». Например, «Роберт Смит, 1962 г.р.» имеет гораздо меньшую различительную способность, чем «Роберт Герберт Смит, род. 13.06.1962, Бомонт, Техас». Способы сопоставления, при которых совпадения записей возможно только с некоторой вероятностью, позволяют получить меньшее количество ложных совпадений и ложных несовпадений, снижают потребность в дальнейшем исследовании посредством автоматизированных процессов сопоставления с участием человека (рисунок 2).

Существует пять основных возможных результатов вероятностного сопоставления пар записей: (1) совпадение: пара, в реальности представляющая один и тот же объект, (2) несовпадение: пара, в реальности представляющая два разных объекта, (3) возможное совпадение: пара без достаточного количества информации, сделать вывод о совпадении невозможно; (4) ложное совпадение: пара, ошибочно обозначенная как совпадение (то есть ложно положительный результат), и (5) ложное несовпадение: пара, ошибочно обозначенная как несовпадение (то есть ложно отрицательный результат) (UNECE, 2011).

Рисунок 2.

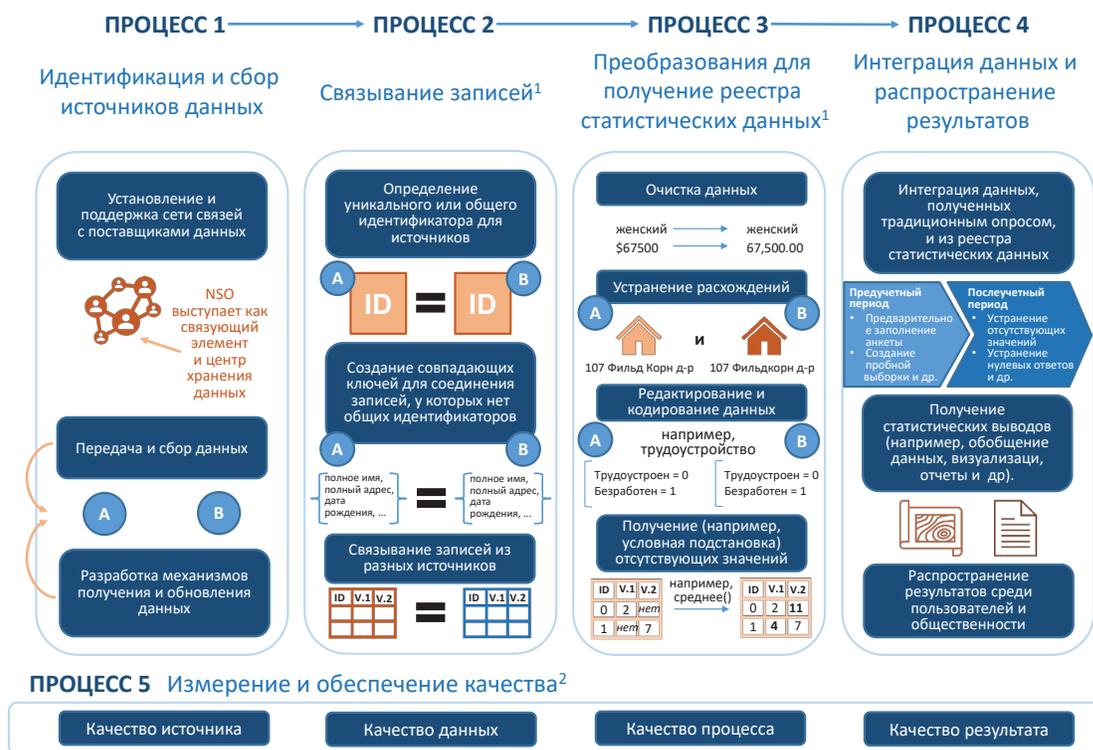
Основные потенциальные выводы из совпадающих пар записей



Источник: United Nations Economic Commission for Europe, 2011.

Рисунок 3.

Общая схема интеграции источников административных данных в результатах переписи населения



¹ Более подробно процессы 2 и 3 описаны в документе United Nations Statistics Division (UNSD), 2022 draft.

² Более подробно процесс 5 описан в документе United Nations Economic Commission for Europe, 2021; UNSD, 2022 draft.

Источник: United Nations Economic Commission for Europe, 2018 and 2021.

Преобразования для формирования реестров статистических данных

Преобразование данных происходит после того, как установлена связь. Целью преобразования может быть учет недостаточного или избыточного охвата, устранение расхождений между источниками или условной подстановки отсутствующих данных. Во время этих процессов административные данные преобразуются при помощи статистических алгоритмов, а полученные таким образом сведения заносятся в реестры статистических данных. К процессам преобразования относятся:

- Очистка данных.
- Устранение расхождения (например, обеспечение совпадения адресов в разных источниках).
- Редактирование и кодирование данных (например, согласование кодов в разных источниках).
- Преобразование административных единиц (например, налоговых и юридических) в статистические единицы (например, люди и домохозяйства).
- Получение (например, условная подстановка) отсутствующих значений.

- Выбор записей, соответствующих целевой группе населения. Полученные реестры преобразованных и связанных статистических данных представляют собой базы данных с постобработкой, которые можно использовать для получения результатов, аналогичных переписи. Общие связанные реестры статистических данных включают в себя реестры населения, реестры адресов/жилых помещений, реестры юридических лиц и реестры деятельности. Записям в этих реестрах соответствуют физические и юридические лица.

Интеграция данных и распространение результатов

Затем сформированные реестры статистических данных можно использовать в сочетании с данными, собранными посредством традиционной переписи, для получения типовых статистических результатов требуемого качества (рисунок 4). Успешная интеграция реестров позволит NSO получать наборы статистических данных, подобные результатам переписи. Такие наборы можно получать более часто, чем результаты переписи, а затем передавать пользователям и общественности.

Рисунок 4.

Измерения качества выходных данных

- 1 РЕЛЕВАНТНОСТЬ**
Степень, в которой результаты переписи соответствуют требованиям пользователя с точки зрения охвата и содержания.
- 2 ТОЧНОСТЬ И НАДЕЖНОСТЬ**
Степень, в которой полученная информация точно описывает измеренное явление. Степень близости оценки к последующим оценкам.
- 3 СВОЕВРЕМЕННОСТЬ**
Время, прошедшее между датой сбора данных и датой публикации статистики.
- 4 СОГЛАСОВАННОСТЬ И СРАВНИМОСТЬ**
Степень, в которой данные, полученные из разных источников и разными методами, одинаковы. Степень, в которой данные можно сравнить с точки зрения географии, времени и групп населения.
- 5 ДОСТУПНОСТЬ И ПРОЗРАЧНОСТЬ**
Простота доступа пользователей к результатам переписи, наличие понятных статистических выводов, а также метаданных.

Источник: United Nations Economic Commission for Europe, 2021.

Измерение и обеспечение качества

На протяжении каждого из процессов, показанных на общей схеме, NSO должны проводить проверки качества и осуществлять его менеджмент. В широком смысле NSO могут рассматривать качество входных данных (включая качество источников), качество процесса (изменение качества данных по ходу процесса) и качество выходных данных (качество данных, переданных общественности или пользователям данных) (рисунок 4). Более подробные сведения приведены в документе UNECE, 2021 and UNSD, 2022 draft.

ТРУДНОСТИ ИНТЕГРАЦИИ ADS

Несмотря на множество преимуществ применения ADS, существует также и много проблем. Для NSO важно проанализировать проблемы, связанные с интеграцией ADS, до начала их использования.

Ограниченный контроль над сбором административных данных

Некоторые из самых сложных проблем с интеграцией ADS в традиционную структуру переписи связаны с ограниченным контролем над сбором административных данных со стороны NSO. NSO полагаются на государственные органы, собирающие данные для конкретных административных целей, которые могут отличаться от потребностей NSO. Это может привести к различиям в концепциях и определениях, единицах анализа и учетных периодах (UNECE, 2011 and 2018). NSO должны учитывать эти различия путем, например, создания систем классификации для сопоставления определений между ADS и данных переписей или преобразования административных единиц (например, налоговых и юридических) в статистические единицы (например, люди и домохозяйства). Создание таких систем может быть довольно дорогостоящим. Процессы преобразования и связывания, которые позволяют NSO использовать ADS для целей, отличных от тех, для которых изначально создавались, могут вызывать озабоченность с точки зрения конфиденциальности, безопасности и юридических вопросов (UNECE, 2018).

Поле 2.

Использование административных данных в Совместной группе

В 2020 году [United Nations Statistics Division](#) и [Global Partnership for Sustainable Development Data](#) организовали совместную работу заинтересованных сторон с конкретной целью: укрепить потенциал NSO по применению ADS в статистических целях. Эта группа, называемая Совместной группой по использованию административных данных для статистики (далее — *Совместная группа*), выступает в качестве платформы для обмена ресурсами, инструментами, передовой практикой и опытом в различных страновых, региональных и международных контекстах.

Страны, участвующие в Совместной группе, имеют широкий опыт интеграции ADS. Некоторые страны, такие как Гана, находятся на начальных этапах интеграции, что предполагает получение политической поддержки на применение ADS. В Гане официальные лица проводили мероприятия и конференции, на которых небольшие пилотные проекты, иллюстрирующие процесс передачи данных от министерств в Статистическую службу Ганы. Была показана возможность применения ADS и в более крупных масштабах. Другие страны, такие как Эквадор, работают с ADS уже несколько лет, разработали протоколы и методологии для их интеграции. В случае с Эквадором официальные лица смогли улучшить качество идентификационных данных ENEMDU (в соответствии с испанским акронимом) на участников опроса COVID-19, связав эти данные с базой данных «Cedulados» из Реестра актов гражданского состояния Эквадора.

Благодаря Совместной группе развитые и развивающиеся страны объединяются в ответ на острую потребность в более своевременных и дезагрегированных статистических данных для достижения SDG на период до 2030 года.

Источник: United Nations Statistics Division and the Global Partnership for Sustainable Development Data, 2021.

Несоответствия между источниками и отсутствующие данные

Проблемы также возникают, когда кажется, что несколько источников административных данных противоречат друг другу, или когда соответствующие данные полностью отсутствуют. В первом случае NSO должны принимать решения относительно того, какие ADS использовать (то есть выбирать наиболее актуальные и надежные), а какие – игнорировать. Как вариант, возможна разработка процессов согласования административных данных из различных источников. В последнем случае NSO должны внедрить процессы либо для сбора дополнительных данных, либо для условной подстановки.

Стойкость к изменениям и сохранение знаний

Помимо одобрения общественности и заинтересованных сторон, штатные статистики могут не доверять качеству данных, которые были получены другими лицами. Также возможно сосредоточение на отрицательных аспектах интеграции ADS, а не на ее преимуществах (UNECE, 2011).

Такое сопротивление изменениям внутри NSO может привести к ограниченному применению ADS, что снизит эффективность их применения. Кроме того, текучесть кадров в NSO может привести к потере ценных знаний в промежутках между годами сбора статистики (UNECE, 2018).

ЗАКЛЮЧЕНИЕ

Источники административных данных имеют важное и возрастающее значение для повышения эффективности и качества данных, собираемых на традиционных переписях. Рассматривая возможность интеграции ADS, NSO должны сотрудничать с правительством и потенциальными партнерами в области данных, чтобы обеспечить законное и своевременное предоставление качественных данных. В данной технической записке перечислены мероприятия, контроль и оценка которых требуют особого внимания при интеграции ADS. Конечной целью интеграции является повышение качества данных, своевременности и эффективности традиционных переписей.

ЛИТЕРАТУРА

Mule, Thomas, "Administrative Records and the 2020 Census," U.S. Census Bureau, Decennial Statistical Studies Division, April 2021.

United Nations Economic Commission for Europe, "Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices," New York and Geneva, 2011.

_____, "Measuring population and housing: Practices in UNECE countries in the 2010 round of censuses," New York and Geneva, 2014.

_____, "Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses," New York and Geneva, 2018.

_____, "Guidelines for Assessing the Quality of Administrative Sources for US in Censuses," Geneva, 2021.

_____, "Recommendations for the 2020 Censuses of Population and Housing," New York and Geneva, 2015.

United Nations Statistics Division, "Handbook on Registers-Based Population and Housing Censuses," second draft, March 2022.

_____, "Report on the results of the UNSD survey on 2020 round population and housing censuses," March 2020.

United Nations Statistics Division and the Global Partnership for Sustainable Development Data, "Collaborative on the Use of Administrative Data for Statistics," 2021.

U.S. Census Bureau, "Intended Administrative Data Use in the 2020 Census," 2020.

U.S. Census Bureau, "Administrative Data Used in the 2020 Census," 2021.



USAID
FROM THE AMERICAN PEOPLE



Серия «Избранные темы международных переписей населения» публикуется в рамках Международных программ Отдела народонаселения Бюро переписи населения США. Агентство США по международному развитию финансирует подготовку серии STIC и двустороннюю поддержку статистических организаций, которые предоставляют информацию авторам. Фонд ООН в области народонаселения участвует в подготовке содержания и обнародовании документов STIC, способствуя их распространению среди более широкой аудитории.