Предотвращение раскрытия данных и перепись населения

Выбор тем в международных переписях населения[1]

По Sam Dupre

Выпущено в октябре 2020 года

ВВЕДЕНИЕ

Одна из важнейших функций национальных статистических служб (NSO, в соответствии с английским акронимом) заключается в проведении национальных переписей населения и жилищного фонда. При этом у служб NSO имеются два предписания по контролю данных, которые могут быть взаимоисключающими. Надлежащий контроль данных предполагает как охрану приватности респондентов, доверивших свою информацию службе NSO, так и обнародование точных и актуальных данных переписи. Поскольку службе NSO нужно, чтобы жители участвовали в переписи и представляли точные ответы, их необходимо убедить в том, что ответы будут надежно защищены.

В настоящей технической записке рассматриваются три вида случаев разглашения данных, возможные формы атак, а также четыре последовательных этапа предотвращения разглашения данных, которые служба NSO может выполнить в ходе переписи для обеспечения надлежащего контроля данных.

КЛЮЧЕВЫЕ ПОНЯТИЯ

Типы данных

Существует два типа данных, которые NSO может обнародовать: микроданные и обобщенные данные. Микроданные — это набор ответов для единицы наблюдения (в случае переписи населения и жилищного фонда — домохозяйство). Микроданные обычно обнародуются в форме небольшой выборки всех данных респондентов. Обобщенные данные — это сводная информация по полным группам лиц в форме показателей частотности или величин (например, средние значения, диапазоны и другие сводные статистические данные).

Для поддержания доверия общественности службам NSO не следует публиковать полный неотредактированный набор данных переписи до тех пор, пока не пройдет достаточно времени, чтобы приватность перестала быть актуальной (UNECE-CES, 2015). Например, Соединенные Штаты выпускают файлы данных переписи только через 72 года после ее проведения (USCB, дата не указана).

Формы раскрытия

Существуют три основные формы раскрытия. Каждая из них создает различный уровень риска для респондентов (Buron and Fontaine, 2018; UNSD, 2015). *Identity disclosure* происходит в тех случаях, когда личность респондента непосредственно связана с обнародованной записью данных (например, через имя, адрес, идентификационный номер, отпечаток пальца, адрес электронной почты или номер телефона). *Attribute disclosure* происходит, когда значения обнародованных данных разглашают другие атрибуты лица. *Inferential disclosure* (с последующей реконструкцией и идентификацией) происходит, когда обнародованные данные используются для вывода значений по конкретным респондентам на основе статистических свойств публикуемых данных. Частота и вероятность разглашения данных различаются в зависимости от обнародованного результата обработки (McKenna and Haubach, 2019). Например, отчет, в котором говорится, что все члены той или иной этнической группы проживают в одном географическом районе, сопряжен с высоким риском раскрытия атрибутов, но низким риском раскрытия личности. Это обусловлено тем, что злоумышленник может определить район, в котором проживает тот или иной человек, если ему известно, что этот человек является членом соответствующей этнической группы. Методов осуществления атак много; наиболее распространенные из них приведены в поле 1.

[1] Настоящая техническая записка является одной из 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.

U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov

United States Agency for International Development
www.usaid.gov

United Nations Population Fund
www.unfpa.org

Поле 1.

Формы атак

Этапы атаки

Повторная идентификация: запись соотносят с конкретным человеком.

Реконструкция: выполняется деанонимизация анонимизированных значений в каждой записи.

Эти этапы не обязательно идут в указанной последовательности, так как реконструированные значения могут использоваться для повторной идентификации и наоборот.

Распространенные виды атак

Атака реконструкции базы данных: сопоставление поля (ключа) в анонимизированном наборе данных со стандартным полем в общедоступном наборе данных.

Атака отслеживания: попытка найти целевые данные в наборе данных. В зависимости от полноты набора данных или уровня конфиденциальности темы даже простое определение факта присутствия в наборе сведений о конкретном лице может быть серьезным нарушением приватности или создать предпосылки для дальнейшей реконструкции.

Разностная атака: один из самых крупных рисков для пространственных данных. Осуществляется с использованием отличий повторяющихся запросов для выяснения сведений о записях путем сравнения подмножеств.

Примечание: по материалам Dwork et al., 2017; McKenna and Haubach, 2019; и UNSD, 2015.

ПРОЦЕСС ПРЕДОТВРАЩЕНИЯ РАСКРЫТИЯ ДАННЫХ

Процесс предотвращения раскрытия данных обычно состоит из четырех этапов: 1) оценка риска, 2) консультации с общественностью, 3) меры контроля раскрытия данных и 4) архивирование и доступ/выпуск. Эти этапы начинаются до переписи и продолжаются после завершения основного графика переписи. В таблице 1 приведены сведения о четырех этапах и их стадиях.

На каждом этапе NSO применяет три стратегии обеспечения приватности респондентов. NSO *ограничивает сбор данных*, по возможности сводя к минимуму получение сведений по конфиденциальным темам (NASEM, 2017; UNSD, 2015). NSO *ограничивает данные,* контролируя типы и формы выпускаемых компонентов данных (например, обобщенные данные или микроданные), а также следя за тем, какие статистические меры контроля применяются к данным. Кроме того, NSO *ограничивает доступ*, контролируя круг пользователей, имеющих доступ к данным, а также объем такого доступа.

В таблице 1 представлен обзор трех перечисленных стратегий. Данные сведения приводятся не для изложения технических деталей, а в качестве общего руководства по техническим особенностям процесса. Учитывая сложность и огромную важность таких мероприятий, в обзор включена дополнительная информация об оценке рисков после регистрации, этапах статистического контроля и доступе к архивам и выпуске.

Таблица 1.

Этапы предотвращения раскрытия информации

Этап	Стадия этапа	Описание
Оценка рисков	Внутренняя оценка	На начальном этапе планирования переписи следует проанализировать риски, связанные с типами и конфиденциальностью собираемых и публикуемых данных.
	Внешняя оценка	После внутренней оценки, еще до проведения переписи, нанимаются внешние консультанты для независимой оценки рисков.
	Повторная внутренняя оценка	После проведения переписи и принятия мер контроля раскрытия данных еще раз проводится внутренняя оценка рисков, включая количественную проверку собранных данных.
	Совет по рассмотрению вопросов раскрытия данных	До, во время и после переписи населения совет по рассмотрению вопросов раскрытия данных проводит оценку рисков для новых планов обнародования данных. Данный орган также пересматривает обнародованные продукты по мере появления новых технологий.
Консультация с общественностью	N	До проведения переписи необходимо проконсультироваться с заинтересованными лицами по вопросам, касающимся их приватности и потребностей в данных. то есть выяснить, какие данные нужно для них обнародовать и в каком формате. Эти сведения следует использовать в качестве руководства при оценке рисков NSO с самого начала процесса планирования, ориентируясь на группы населения, исторически с трудом поддающиеся учету (см. руководство Бюро переписи населения США по регистрации групп населения, с трудом поддающихся учету [2019 год, A]). В поле 2 описана конкретная реализация: действия Бюро национальной статистики Великобритании на этапе консультаций с общественностью.
Меры по контролю раскрытия информации	Правовые меры	До проведения переписи следует принять законодательные акты, возлагающие на NSO ответственность за защиту данных респондентов и в частности оговаривающие порядок публикации данных. Таким образом будет обеспечена законодательная поддержка для обоснования прав NSO на принятие решений.
	Физические меры	До проведения переписи следует разработать политику в отношении удаления материалов, доступа на объекты и порядка обработки сохраненной репрезентативной выборки. Удаление включает уничтожение бумажных форм и стирание данных на устройствах для личного опроса с применением компьютера.
	Технические меры	Перед проведением переписи следует ввести политику, препятствующую перехвату ответов в ходе переписи по Интернету; обеспечить защиту данных на устройствах для личного опроса с применением компьютера; усилить сетевую безопасность NSO и обеспечить разграничение доступа сотрудников к данным респондентов.
	Статистические меры	После проведения переписи следует применить статистические меры контроля к микроданным респондентов (до составления таблицы) или к обобщенным данным (после составления таблицы). Конкретные меры зависят от формы запланированного обнародования.
Доступ к архивам/ выпуск	N	После переписи следует архивировать микроданные (необработанные файлы и отредактированные файлы после статистического контроля), метаданные и параданные и предоставить их заинтересованным лицам. Подробные сведения об архивировании защищенных данных см. в Руководстве по архивированию и сохранению данных Бюро переписи населения США (2019 год, B).

N — Не применимо.
Примечание: по материалам Lauger et al., 2014; McKenna and Haubach, 2019; NASEM, 2017; UNECE-CES, 2015; и UNSD, 2015.

Дополнительные сведения об оценке рисков и мерах статистического контроля после переписи

Оценка рисков и меры статистического контроля, как правило, — самые сложные в техническом отношении элементы контроля раскрытия данных (таблица 1). Конкретные меры, которые необходимо применять, зависят от следующего:

- Форма выпуска (например, выборки микроданных общего пользования [PUMS, в соответствии с английским акронимом] или обобщенные данные) (McKenna and Haubach, 2019).
- Планируемый уровень детализации (для файла PUMS [в соответствии с английским акронимом] с грубым обобщением данных может потребоваться минимальная численность населения 100 тыс. человек, а для файла с высокой степенью детализации может потребоваться минимальная численность населения 400 тыс. человек) (Buron and Fontaine, 2018).

Поле 2.

Пример: Бюро национальной статистики (ONS, в соответствии с английским акронимом) Великобритании

В ходе подготовки к Переписи населения 2021 года в период с 2015-го по 2018 год ONS провело несколько раундов консультаций с общественностью по темам и требованиям переписи 2021 года. После каждого раунда ONS публиковало подробную информацию о 1) первоначальных планах, 2) ответах общественности, 3) планах ONS, подготовленных с учетом ответов, и 4) влиянии изменения планов на обеспечение равного представительства в переписи 2021 года.

Источник: ONS, 2018.

После проведения переписи NSO может выполнить повторную оценку рисков на основе собранных данных с учетом обновленных планов выпуска и принять меры статистического контроля. Это комплексный процесс, состоящий из пяти этапов:

Этап 1. Устранение сведений, идентифицирующих личность (PII, в соответствии с английским акронимом)

Удаление прямых идентификаторов, включая имя, адрес и любые государственные идентификационные номера, из записей для предотвращения прямого раскрытия личности (UNSD, 2015).

Этап 2. Идентификация конфиденциальных записей, ячеек и категорий

Для оценки риска раскрытия статистической информации требуются высококвалифицированные эксперты в предметной области для выявления конфиденциальных/деликатных тем и уязвимых групп на местном уровне (UNSD, 2015), но существуют и количественные методы, позволяющие оценить риск раскрытия данных. Использование количественных методов позволяет четко сопоставить различные варианты обнародования и обеспечивает состоятельное юридическое обоснование процесса принятия решений NSO (NASEM, 2017).

Таблица 2.
Типичные признаки конфиденциальных записей, ячеек и категорий

Проблема	Причина появления проблемы	Количественная оценка риска
Наличие ячеек, соответствующих небольшому подмножеству.	Если для какой-либо группы доступно очень мало записей, риск разглашения личных данных возрастает.	Следует отметить все единицы, находящиеся ниже стандартного порога. Для Анкетирования населения США по месту жительства (ACS, в соответствии с английским акронимом), которое проводится Бюро переписи населения США, и файла PUMS переписи 2010 года: • Каждая категория конкретной категориальной переменной должна включать не менее 10 000 человек или домохозяйств в невзвешенном выражении. • Все географические районы (включая городские и сельские) должны содержать не менее 50 человек или домохозяйств в невзвешенном выражении для одной переменной. • Среднее значение размера ячейки таблицы должно составлять не менее трех невзвешенных случаев.
Наличие ненулевых значений подсчета уязвимых групп.	Даже само знание факта существования лиц с определенными характеристиками может привести к нарушению приватности.	Следует отметить все ячейки с заранее идентифицированными характеристиками конфиденциальности или их сочетаниями.
Различные подмножества результатов включают одну и ту же совокупность (совокупности) населения.	Такие подмножества можно сравнить в ходе разностной атаки, чтобы вывести данные респондента.	Следует отметить невложенные географические группы и группы респондентов для дальнейшей проверки, уделяя особое внимание случаям, когда существуют лишь незначительные различия между повторяющимися подмножествами населения.
Лица, находящиеся в домохозяйстве, отмечаются в качестве находящихся в группе риска.	Если какое-либо лицо подвергается риску раскрытия данных, то под угрозой может быть все домохозяйство.	Оценить риск на индивидуальном уровне для каждой переменной и географического уровня. Обобщить данные по лицам, составляющим домохозяйство, и отметить все домохозяйства с членом в группе риска.
В ответах для любой переменной имеются выпадающие значения.	Тех, кто находится в верхней или нижней части распределения ответов, идентифицировать легче, чем респондентов, у которых ответы ближе к среднему значению.	Для непрерывных переменных следует отметить записи со значениями, близкими к максимуму и минимуму распределения. Обычно это нижние и верхние 0,5 процента всех значений (или 3 процента всех ненулевых значений, если соответствующее число записей больше).
	Респондент(ы) с наибольшими значениями в группе будут с максимальной вероятностью подвергаться риску со стороны других респондентов в предельных значениях этой группы.	С помощью правил (n, k), p% и pq можно отметить случаи, когда респонденты с выпадающими показателями могли бы идентифицировать других таких респондентов на основе своих собственных ответов. Правило (n, k) позволяет отметить переменную, если значения для максимум n респондентов составляют не менее k процентов от суммарных значений. Правило p% и правило pq позволяют отметить случаи, когда другие респонденты могут приблизительно указать значения для респондента с наибольшим значением с точностью p процентов от истинного значения. Конкретные значения n, k, p и q, которыми пользуется NSO, как правило, конфиденциальны, поскольку даже эта информация может сделать данные уязвимыми для атаки косвенного раскрытия.

Примечание: информация получена из Antal et al., 2017; Buron and Fontaine, 2018; Lauger et al., 2014; McKenna and Haubach, 2019; и OECD, 2005.

В таблице 2 приведены проблемы, с которыми NSO может столкнуться при идентификации конфиденциальных записей, ячеек и категорий, а также рекомендации в отношении количественных методов, позволяющих оценить, следует ли отметить конкретный конфиденциальный элемент в качестве подлежащего статистическим мерам контроля.

Этап 3. Устранение риска

Статистический контроль может быть *пертурбативным* и *непертурбативным* (Antal et al., 2017). Пертурбативные меры незначительно изменяют данные контролируемым образом, минимально влияя на их структуру. Непертурбативные меры основаны на удалении (или обобщении) табличных ячеек, географических районов или записей данных, отвечающих определенным уровням риска. Пертурбативные методы, как правило, более надежно сохраняют структуру данных и влекут меньшую потерю информации, чем непертурбативные методы (Antal et al., 2017).

Непертурбативные

Первичное и вторичное/дополнительное подавление. Первичное подавление защищает от идентификации и раскрытия атрибутов: ячейки или записи заменяются меткой, указывающей на их подавление или вызывающей отображение сообщения «Нет данных» (Antal et al., 2017). Вторичное подавление предполагает подавление дополнительных неотмеченных ячеек, так чтобы подавленные значения не могли быть получены путем косвенного раскрытия. В качестве альтернативы можно исключить из обнародования все проблемные переменные либо отмеченные группы или географические районы целиком (UNECE-CES, 2015).

Перекодирование. Когда в наличии слишком мало записей для какого-либо значения или диапазона значений, его можно объединить с другими группами, записями, столбцами или строками, чтобы достигнуть необходимого порога. В тех случаях, когда общедоступные данные можно увязать с данными переписи, может потребоваться перекодирование для предотвращения раскрытия атрибутов или косвенного раскрытия — даже в ситуации, когда стандартный порог уже достигнут. Варианты перекодирования количественных данных включают округление, интерполяцию в пределах предопределенного диапазона/распределения и сокращение до квантилей для снижения специфичности данных (Dajani et al., 2017). Для маскировки выпадающих показателей непрерывных переменных применяют перекодирование по методу верхнего и нижнего кода. Выпадающие показатели выше или ниже порогового перцентиля заменяются граничным значением либо средним или медианным значением для всех значений с верхним/нижним кодом.

Пертурбативные

Добавление шума. В сопряженные с риском ячейки вносится случайный шум путем незначительного изменения исходных значений. Добавляемый шум полностью сохраняет структуру данных переменной благодаря контролю систематической ошибки, дисперсии, частотности, а также корректировке нулевых ячеек (Antal et al., 2017).

Перестановка записей, перестановка рангов и перетасовка. Перестановка записей подразумевает сопоставление пар записей по некоторым критериям и обмен неравными значениями в таких парах (Antal et al., 2017). Географический район обмениваемых пар должен быть по возможности одинаковым, с тем чтобы свести к минимуму порчу данных и географические сдвиги (например, перестановку следует выполнять внутри регионов, но не между регионами) (Buron and Fontaine, 2018), однако это не обязательно, если имеется существенный риск раскрытия данных (Lauger et al., 2014). Перестановка рангов и перетасовка предусматривают обмен значениями некоторой переменной между записями с близкими значениями этой переменной.

Синтетические данные. Создайте статистическую модель, которая описывает набор данных, и замените уникальные записи на смоделированное значение (Dajani et al., 2017). Такие наборы данных также требуют оценки рисков ввиду сохранения вероятности инцидентов, связанных с раскрытием информации; вместе с тем они позволяют исследователям получать доступ к данным, которые в ином случае могли бы создать непозволительный риск.

Этап 4. Проверка результатов

Проверьте остаточный риск с помощью мер, описанных для этапа 2, и проверьте уровень потери данных (например, приращение дисперсии оценки параметров или появление систематического отклонения). Для оценки потери данных проверьте следующее:

- Не оказывает ли пертурбация существенного влияния на минимумы/максимумы, среднее/медианное значение/моду или перцентили (абсолютную и относительную разницу) (Antal et al., 2017).
- Долю ячеек, в которых пертурбация превышает заданный порог изменения, — поскольку незначительные изменения в районах с низкой плотностью могут иметь более существенные последствия, чем относительно крупные изменения в районах с высокой плотностью (Buron and Fontaine, 2018).
- Не добавляет ли пертурбация существенный процент «ложноположительных» значений (нулевые значения, искаженные до ненулевых) и «ложноотрицательных» значений (ненулевые значения, искаженные до нулевых) (Buron and Fontaine, 2018).

- Если отношения данных после пертурбации сохранились (например, между двумя переменными сохранилось ожидаемое равенство, неравенство или определенное статистическое соотношение) (Antal et al., 2017).

Этап 5. Внутреннее исследование атак

Служба NSO может провести внутреннее исследование атак, направленных на раскрытие данных, с целью выявления уязвимостей на фоне возникновения новых угроз (McKenna and Haubach, 2019). В рамках таких исследований необходимо применять те же методы и технологии, которыми могут пользоваться злоумышленники, включая государственные и частные наборы данных и новые технические разработки. Такие тесты можно применять как к новым, так и к старым выпускам данных, чтобы анонимизированные ранее записи не оказались уязвимыми для раскрытия.

Дополнительные сведения об архивировании и доступе/выпуске

Микроданные (как необработанные, так и отредактированные файлы после предотвращения раскрытия статистических данных), метаданные и параданные могут создавать риски раскрытия в отдельности или при использовании с другими источниками (UNECE-CES 2015). Службе NSO следует сохранить оригинальные неотредактированные версии данных и создать журнал всех изменений, хранящийся отдельно от анонимизированных файлов данных (Van den Eynden et al., 2011). В некоторых случаях NSO может сохранять репрезентативную выборку заполненных форм переписи. При этом перед любым выпуском могут применяться все принципы обеспечения приватности данных.

В числе механизмов, которые служба NSO может использовать для сохранения безопасных форм доступа, — анклавы данных, средства удаленного доступа, онлайн-базы данных для запроса или анализа наборов данных, соглашения лицензирования для уполномоченных пользователей и опубликование файла PUMS (FCoSM, 2005; Hundepool et al., 2012). Общие принципы применяются всегда, независимо от действующего механизма ограничения доступа. По соображениям безопасности NSO не должна предоставлять внешним пользователям доступ к внутренним сетям. Данные, не одобренные для выпуска в открытый доступ, могут быть зашифрованы перед любой операцией передачи из внутренних защищенных сетей. Анклавы данных не должны иметь доступа к Интернету, внешним сетям и портам USB (UNSD, 2015). Необходимо автоматически применять дополнительные меры по недопущению раскрытия данных в случае, если одни и те же таблицы запрашиваются несколько раз, чтобы защититься от разностных атак с использованием повторяющихся подмножеств. Такие меры должны предусматривать проведение необъявленных проверок хранилищ данных, проверку статистических результатов и контроль удаления данных и производных файлов. Для обеспечения эффективности все меры могут иметь обязательную юридическую силу и предусматривать наказания за нарушения (FCoSM, 2005; UNSD, 2015).

ЗАКЛЮЧЕНИЕ

В отсутствие обнародования актуальных, пригодных для использования данных польза переписи населения резко снижается. Вместе с тем повышение степени детализации публикуемых данных увеличивает риск нарушения приватности респондентов. Этот риск возрастает в эпоху Больших данных, поскольку прогресс в сфере инструментов добычи, географической привязки и статистической обработки данных повышает вероятность инцидентов, связанных с раскрытием. Службы NSO могут выполнять свои обязанности перед обществом путем управления рисками с использованием политики и процедур, представленных в настоящей записке.

ЛИТЕРАТУРА

Antal, L., T. Enderle, and S. Giessing, *Harmonised Protection of Census Data in the ESS: Statistical disclosure control methods for harmonised protection of census data*, Eurostat Centre of Excellence on Statistical Disclosure Control, The Hague, 2017.

Buron, M. L., and M. Fontaine, Confidentiality of Spatial Data, in Loonis, V. and Marie-Pierre de Bellefon, *Handbook of Spatial Analysis: Theory and application with R*, chapter 14, Insee Méthodes No. 131, Eurostat, The Hague, 2018.

Council of Europe (CoE), *Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data*, Directorate General of Human Rights and Rule of Law, Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Strasbourg, France, 2017.

Dajani, A.N., A.D. Lauger, P.E. Singer, D. Kifer, J.P. Reiter, A. Machanavajjhala, S.L. Garfinkel, S.A. Dahl, M. Graham, V. Karwa, H. Kim, P. Leclerc, I.M. Schmutte, W.N. Sexton, L. Vilhuber, and J.M. Abowd, The Modernization of Statistical Disclosure Limitation at the U.S. Census Bureau, in *September 2017 Meeting of the Census Scientific Advisory Committee*, Suitland, MD, 2017.

Dwork, C., A. Smith, T. Steinke, and J. Ullman, *Exposed! A Survey of Attacks on Private Data*, Annual Review of Statistics and Its Applications, 4(12): 1–24, 2017.

Federal Committee on Statistical Methodology (FCoSM), *Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology Version 2*, U.S. Office of Management and Budget, Washington, DC, 2005.

Hundepool, A., J. Domingo-Ferrer, L. Franconi, S. Giessing, E. Shulte Nordholt, K. Spicer, and P.P. de Wolf, Statistical Disclosure Control, In: *Wiley Series in Survey Methodology, Wiley*, Chichester, United Kingdom, 2012.

Lauger, A., B. Wisniewski, and L. McKenna, Disclosure Avoidance Techniques at the *U.S. Census Bureau: Current practices and research, research report series (Disclosure Avoidance #2014-02)*, Center for Disclosure Avoidance Research, U.S. Census Bureau, Washington, DC, 2014.

McKenna, L. and M. Haubach, *Legacy Techniques and Current Research in Disclosure Avoidance at the U.S. Census Bureau*, Research and Methodology Directorate, U.S. Census Bureau, Washington, DC, 2019.

National Academies of Sciences, Engineering, and Medicine (NASEM), *Innovations in Federal Statistics: Combining data sources while protecting privacy*, The National Academies Press, Washington, DC, 2017, <https://doi.org/10.17226/24652>.

OECD, Glossary of Statistical Terms, <https://stats.oecd.org/glossary/detail.asp?ID=6943>, 2005, accessed on July 15, 2020.

Office for National Statistics (ONS), Initial View on 2021 Census Output Content Design: Response to consultation, Office for National Statistics, United Kingdom, 2018.

United Nations Economic Commission for Europe—Conference of European Statisticians (UNECE-CES), *Recommendations for the 2020 Censuses of Population and Housing*, United Nations Publications, New York, NY, 2015.

United Nations Statistics Division (UNSD), *Principles and Recommendations for Population and Housing Censuses, Revision 3*, United Nations Publications, New York, NY, 2015.

United States Census Bureau, *Counting the Hard to Count in a Census*, Select Topics in International Censuses, <www.census.gov/content/dam/Census/library/working-papers/2019/demo/Hard-to-Count-Populations-Brief.pdf>, 2019a.

_____ , *Census Data Archiving and Preservation*, Select Topics in International Censuses, <www.census.gov/content/dam/Census/library/working-papers/2019/demo/Archiving-Brief.pdf>, 2019b.

_____, The "72-Year Rule," <www.census.gov/history/www/genealogy/decennial_census_records/the_72_year_rule_1.html>, n.d., accessed on July 15, 2020.

Van den Eynden, V., L. Corti, M. Woollard, L. Bishop, and L. Horton, *Managing and Sharing Data*, UK Data Archive, UK, 2011.

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