

**Европейская экономическая комиссия****Конференция европейских статистиков****Шестьдесят девятая пленарная сессия**

Женева, 23–25 июня 2021 года

Пункт 7 с) предварительной повестки дня

**Доклады, руководящие принципы и рекомендации,
подготовленные под эгидой Конференции:****Оценка качества административных источников
для использования в целях переписи****Оценка качества административных источников
для использования в целях переписи****Документ подготовлен Целевой группой по оценке качества
административных источников для использования в целях
переписи***Резюме*

Настоящий документ представляет собой сокращенную версию «Руководства по оценке качества административных источников для использования в целях переписи». Цель Руководства заключается в том, чтобы предоставить организаторам переписей практические рекомендации по оценке качества административных данных с использованием последовательных этапов оценки: источник, данные, процесс и результат, причем первые два из этих этапов являются основным объектом Руководства.

Настоящее Руководство было подготовлено Целевой группой по оценке качества административных источников для использования в целях переписи в составе Соединенного Королевства (председатель), Австрии, Германии, Израиля, Ирландии, Испании, Италии, Канады, Нидерландов, Новой Зеландии, Польши, Португалии, Республики Молдова, Соединенных Штатов Америки, Турции, Франции, Черногории, Фонда Организации Объединенных Наций в области народонаселения, Евростата и независимого эксперта.

Настоящий сокращенный вариант Руководства подготовлен для целей перевода. Он включает в себя большую часть введения, главы, посвященные системе качества и этапу «Источник», а также полную главу, содержащую выводы и рекомендации. Просьба учесть, что нумерация разделов в этой сокращенной версии не соответствует нумерации глав и разделов в полном тексте Руководства.



Остальные разделы можно найти в полном варианте Руководства на английском языке, который был разослан всем членам Конференции европейских статистиков (КЕС) для консультаций в марте–апреле 2021 года, а также размещен на веб-странице Конференции по адресу: <https://unece.org/statistics/events/CES2021>. Резюме отзывов, полученных в рамках консультаций, будет представлено в документе ECE/CES/2021/3/Add.1.

При условии получения положительного отклика в рамках консультаций пленарной сессии КЕС будет предложено одобрить данное Руководство.

I. Введение

A. Справочная информация

1. В 2017 году Целевая группа ЕЭК ООН по регистровым и комбинированным переписям подготовила Руководство по использованию регистров и административных данных в целях переписей населения и жилищного фонда¹. Руководство включало раздел, посвященный «источникам данных и их качеству», с общим обсуждением этой темы. Эксперты, принявшие участие в Совещании экспертов ЕЭК ООН–Евростата по переписям населения и жилищного фонда (Женева, 4–6 октября 2017 года), определили качество административных источников в качестве темы, имеющей первостепенное значение для многих стран. Исходя из этого Совещание экспертов призвало к созданию новой Целевой группы ЕЭК ООН по измерению качества административных источников для использования в целях переписи, которая будет опираться на результаты работы предыдущей Целевой группы.
2. Данная Целевая группа была учреждена в 2018 году, и ее круг ведения² был утвержден на совещании Бюро Конференции европейских статистиков (КЕС), состоявшемся в феврале 2018 года в Хельсинки (14–15 февраля 2018 года). Целевая группа представила доклад Руководящей группе ЕЭК ООН по переписям населения и жилищного фонда, которая, в свою очередь, представила доклад КЕС и ее Бюро.
3. Задача Целевой группы заключалась в разработке руководства по измерению качества административных источников для использования в целях переписи³. Круг ведения предусматривал разработку руководства, которое будет актуальным для всех стран ЕЭК ООН, с опорой на работу по использованию административных источников в целях формирования официальной статистики, проводимую в рамках проекта Евростата ESS.VIP ADMIN⁴.
4. Настоящее Руководство служит практическим пособием по оценке и измерению качества административных источников для использования в целях переписей населения и жилищного фонда.

B. Использование административных данных в целях переписи

5. Источниками административных данных являются массивы данных, содержащие собранную главным образом для административных целей информацию. Они содержат данные, которые собираются правительственными ведомствами, государственными органами и другими организациями в целях регистрации, проведения операций и учета, обычно в связи с предоставлением каких-то услуг. К ним относятся административные регистры (с уникальным идентификатором), такие как регистры населения, предприятий, адресов, образования, здравоохранения, занятости и налогообложения, а также другие административные источники (без уникального идентификатора). Административные регистры и/или другие административные источники используются для создания статистических регистров, которые конкретно используются в статистических целях, в том числе для проведения переписи.
6. Использование в переписях административных данных в разных странах характеризуется различиями. Такие источники могут использоваться для расширения

¹ URL: <http://www.unece.org/index.php?id=50794>.

² URL: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/bur/2018/February/06Add.1-TF_on_quality_of_admin_data_for_censuses_ToR_apr.pdf.

³ Впоследствии Целевая группа приняла решение скорректировать свое название и соответствующую задачу на «оценку», а не «измерение» качества административных источников для использования в целях переписи.

⁴ Более подробную информацию об этом проекте можно найти на сайте https://ec.europa.eu/Eurostat/cros/content/ess-vision-2020-admin-administrative-data-sources_en.

или дополнения традиционной переписи, для проведения комбинированной переписи или для разработки полностью регистровой переписи. Наблюдается явная тенденция к расширению использования административных данных в целях переписи. Это было мотивировано теми выгодами, которые могут принести административные данные, включая сокращение расходов и нагрузки на респондентов, повышение своевременности и периодичности получения результатов, улучшение качества и повышение гибкости для удовлетворения потребностей пользователей (см., например, раздел 4.1 публикации ЕЭК ООН 2018 года). Кроме того, во многих странах изменились условия в плане поддержки и облегчения использования административных данных во всех национальных статистических системах (см. раздел 4.2 публикации ЕЭК ООН 2018 года). Это включает в себя внесение изменений в законодательство, приемлемость для общественности и заинтересованных сторон, а также разработки в области технологий и статистических методологий.

7. Важность административных данных была еще раз подчеркнута проблемами, с которыми в настоящее время сталкиваются национальные статистические управления (НСУ), когда речь идет о сборе данных непосредственно от населения, будь то из-за нежелания общественности участвовать в переписи или из-за ее способности делать это. Эта проблема ярко проявилась в начале кризиса Covid-19 в 2020 году, который существенно затронул как способность общественности взаимодействовать с НСУ, так и способность НСУ взаимодействовать с общественностью.

С. Ключевые риски для качества

8. При всех преимуществах, которые могут обеспечить административные данные, существует ряд ключевых соображений качества, которые необходимо оценить и изучить, прежде чем включать административный источник в перепись. Во-первых, НСУ будет иметь лишь ограниченный контроль над тем, как собираются и обрабатываются данные. Поэтому существует значительная зависимость от органов, являющихся держателями административных данных. Например, если административный орган не в состоянии выполнить требования НСУ в отношении предоставления требуемых данных в требуемое время, это повлияет на своевременность результатов переписи. Аналогичным образом, если административный орган не будет адекватно взаимодействовать с НСУ в отношении любых потенциальных изменений в источнике, это может повлиять на согласованность и сопоставимость.

9. Во-вторых, использование НСУ административных данных в целях, отличных от тех, для которых эти данные были изначально собраны, вызывает озабоченность по поводу конфиденциальности, безопасности и правовых аспектов. Поэтому НСУ должно оценивать приемлемость для общественности исходя из того, что требуемые гарантии имеются в наличии и что они доводятся до сведения общественности (и административного органа). Использование также должно быть законным. Без одобрения или согласия как со стороны общественности, так и со стороны поставщика данных, или без внушающей доверие правовой основы для использования административного источника, возникнет значительный риск для репутации НСУ и его способности проводить высококачественную перепись. Это может произойти, если общественность изменит свое отношение к взаимодействию с административным органом или НСУ, по причине обеспокоенности в отношении того, как НСУ использует их данные.

10. В-третьих, административные данные (как правило) не собираются для статистических целей. Как следствие, в источниках данных, возможно, приняты иные концепции и определения, чем те, которые требуются в целях переписи; они могут относиться к различным отчетным периодам; и они могут иметь ограниченный охват переписной совокупности. Кроме того, точность и полнота данных будут в значительной степени зависеть от важности этих данных для функции административного органа. Административные источники также могут быть

подвержены изменениям во времени и расхождениям в методах сбора данных о единицах совокупности.

11. И наконец, сложность административных данных, а также наличие и полнота связанных с ними метаданных будут влиять на способность НСУ понимать, использовать административный источник и получать доступ к нему. Например, административные данные могут храниться в больших, сложных структурах данных, что ставит перед НСУ серьезные технические проблемы, которые необходимо учитывать и преодолевать. Сложность административных данных может также влиять на доступность и ясность параметров качества результатов для пользователя. То есть пользователям переписи может быть трудно понять, как используются административные данные в ходе переписи и как такое использование влияет на качество результатов переписи.

12. Эти ключевые соображения качества послужат основой для принятия решений об использовании административных данных в целях переписи. В Руководстве подробно рассматривается каждое из этих соображений.

II. Система качества

13. Качество статистики зависит от того, соответствуют ли статистические материалы требованиям своего предполагаемого использования. Например, определение качества ЕСС разработано на основе семейства стандартов ISO 9000 «Качество — степень, в которой ряд присущих объекту характеристик удовлетворяют требованиям» (ISO, 2015). В официальной статистике объект может представлять собой «статистический продукт, услугу, процесс, систему, методологию, организацию, ресурс или вводимые ресурсы [данные]» (Eurostat, 2020, p. 17). В контексте переписи качество используемых административных данных должно, таким образом, рассматриваться с учетом того, как данные собираются и обрабатываются поставщиками и НСУ, вплоть до итоговых материалов переписи.

A. Качество и погрешность в рамках переписи

14. В тех случаях, когда официальная статистика формируется с использованием методики выборочного обследования, вопросы анкеты обследования разрабатываются и тестируются таким образом, чтобы уменьшить погрешности измерения, обеспечивая тем самым максимальную точность и надежность. Таким образом, предполагается, что погрешность полученных оценок вызвана дефектами процессы формирования выборки и, как правило, измеряется и сообщается в виде среднеквадратической ошибки (СКО) и/или доверительных интервалов. Однако такие измерения не фиксируют не связанных с выборкой ошибок, которые особенно важны в контексте переписей, когда цель состоит в том, чтобы охватить всю совокупность населения. Таким образом в более общем плане, как и в случае статистики, сформированной на основе административных данных, основными источниками ошибок в контексте переписей являются не ошибки выборки, а ошибки репрезентативности (охвата) и измерения (Zhang, 2012). Таким образом, общепринятой практикой является корректировка оценок переписи на основе результатов послеперписного обследования, хотя это может привести к спорам, поскольку сложные методы двойной оценки плохо понимаются общественностью⁵.

15. В тех случаях, когда в целях переписи используются административные данные и другие альтернативные источники данных, такие как большие данные, диапазон возможных ошибок является большим, чем при проведении традиционных переписей, поскольку процессы сбора данных не контролируются НСУ. В работе Zhang (2012),

⁵ Например, корректировки, основанные на послеперписном обследовании, проведенном после переписи населения США 1990 года, стали предметом судебного разбирательства и были отклонены Верховным судом США, а корректировки, внесенные в 2000 году, также стали предметом судебного разбирательства (Бюро переписи населения США, 2009 год).

опирающейся на работу Groves et al., 2004, проводится различие между двумя широкими типами ошибок в статистике, сформированной с использованием административных данных: ошибки измерения и ошибки репрезентативности. К первому типу относятся ошибки в измерении характеристик (например, возраста, пола и т. д.), а ко второму — ошибки в репрезентативности единиц или объектов совокупности (например, отдельных лиц или домашних хозяйств в рамках переписи)⁶. Чанг также проводит различие между качеством отдельных источников, обеспечиваемым поставщиками данных, и качеством преобразованных и/или интегрированных источников, после обработки НСУ. Этот подход нашел свое отражение в Руководстве, которое оценивает качество отдельных административных источников и интегрированных источников с уделением особого внимания выявлению ошибок в измерениях и репрезентативности.

16. Кроме того, для оценки качества административных данных был также адаптирован подход Общей погрешности наблюдения (ОПН). В отличие от СКО, ОПН позволяет выявить более широкий спектр ошибок, включая ошибки достоверности, совокупности/охвата, отсутствия ответов, измерения, обработки и моделирования. Таким образом, подходы ОПН стремятся отразить то, как различные ошибки накапливаются в результате статистического плана и методологии, позволяя определить окончательную погрешность любой заданной оценки. Этот подход был адаптирован для сообщения качества статистики, опирающейся на интеграцию административных данных (см. например, Reid, Zabala and Holmberg, 2017, Rogers and Blackwell, 2020). В то же время качество статистики нельзя сводить только к оценке погрешности. В случае интеграции данных из административного источника в план переписи следует оценивать влияние такой интеграции на качество с точки зрения того, в какой степени она добавляет погрешность или неопределенность в конечные результаты, в сопоставлении с преимуществами интеграции, например, уменьшением нагрузки по предоставлению ответов, повышением своевременности, снижением затрат. Исходя из этого в настоящем Руководстве определены дополнительные параметры, которые могут влиять на общее качество результатов переписи, включая институциональную среду и необходимость балансировки параметров качества в целях удовлетворения потребностей пользователей.

17. Следование рекомендациям настоящего Руководства поможет обеспечить, чтобы оценки переписей основывались на наиболее подходящих источниках и методах и не вводили в заблуждение. В то же время следует также рассмотреть вопрос о том, как предполагается использовать административные источники в плане переписи. Учитывая разнообразие возможных видов использования, эти рамочные принципы следует использовать гибко и адаптировать к уровню качества, требуемому для различных видов использования НСУ административных данных и для удовлетворения различных статистических потребностей пользователей статистики переписи, включая общественность, организации, местные и общенациональные органы власти. Поэтому оценка качества неизбежно опирается на квалифицированное профессиональное суждение на протяжении всего процесса статистического

⁶ На основе работы Zhang (2012). В случае вводных данных ошибки измерения обусловлены различиями между предоставленными и целевыми характеристиками (например, пол, возраст, этническая принадлежность, род занятий и т. д.) и включают несколько типов ошибок в переменных, включая релевантность (несовпадение определений), связывание (ошибки в переклассифицированных показателях из-за плохой эквивалентности между предоставленными и целевыми классификациями, которые, следовательно, могут потребовать корректировок, например, методом импутации), и ошибки сопоставимости (ошибки между переклассифицированными и скорректированными показателями). Ошибки репрезентативности связаны с разницей между предоставленными и целевыми единицами. К ним относятся ошибки, связанные с избыточным и недостаточным охватом (несогласованность с целевой совокупностью), идентификацией (ошибки в классификации единицы из-за расхождений между несколькими источниками) и ошибками единиц (ошибки при статистическом создании представляющих интерес статистических единиц в случаях, когда они отсутствуют в каком-либо доступном источнике данных).

производства, от сбора до публикации, в целях удовлетворения потребностей пользователей.

В. Измерение качества

18. Качество переписных оценок, полученных с использованием административных источников, особенно трудно оценить и/или измерить из-за сложности и многомерности используемых данных. Как отмечалось выше, многие факторы, влияющие на качество, не поддаются количественной оценке. Кроме того, это то, что представляет собой «соответствие цели», и высококачественная статистика неизменно будет варьироваться от одного пользователя к другому, например, некоторые пользователи могут отдавать предпочтение своевременности, а не точности. Поэтому важно оценивать/измерять качество административных данных по всем ключевым параметрам, которые будут представлять интерес для производителей и пользователей статистических данных. Исходя из этого, то, что подразумевается под оценкой и измерением, нуждается в дополнительном уточнении.

19. В настоящем Руководстве проводится различие между *оценкой качества*, т. е. качественной оценкой, и *измерением качества*, т. е. присвоением количественного параметра данной оценке качества. В тех случаях, когда невозможно разработать показатели для количественной оценки или они еще не разработаны, в настоящем Руководстве рекомендуется проводить качественную оценку их влияния на качество. В дополнение к ним существует ряд дополнительных принципов, служащих руководством для формирования официальной статистики (ЕЭК ООН, 1992) и которые применимы на протяжении всего статистического процесса и в более широком контексте НСУ (например, приверженность принципам качества, независимости, защиты данных, конфиденциальности статистических данных и т. д.). Эти вопросы актуальны для всех статистических процессов и не в полной мере охвачены настоящими Руководством. Однако следует признать, что перепись, в которой используются административные источники, опирается на данные, которые были сформированы за пределами статистической системы, в другой организации, над которой НСУ, как правило, не имеет контроля⁷. По этой причине необходимо тщательно изучить влияние использования этих внешних источников на данные принципы.

С. Этапы оценки качества

20. Для обеспечения простоты использования настоящего Руководства оценка качества административных источников рассматривается в разбивке по четырем широким этапам жизненного цикла переписи. Они применимы независимо от типа переписи. Хотя разработка статистического плана никогда не бывает полностью линейной, аналитическое обоснование того, как проводить оценку качества, позволит производителям статистики оперативно определить ключевые параметры качества, наиболее значимые для их собственного контекста. Этими этапами являются:

- *этап «Источник»*: основанная на метаданных оценка качества новых или предоставленных административных источников, которые будут использоваться в переписи. Этот этап не требует от НСУ обладания фактическими данными, но он имеет решающее значение для последующих этапов;
- *этап «Данные»*: оценка качества первичных административных данных, предоставляемых НСУ административными органами. Это потребует от НСУ

⁷ В некоторых случаях НСУ имеет некоторый контроль над регистром. В Швейцарии, например, Федеральный регистр зданий и жилых помещений или Регистр предприятий являются подразделениями Федерального статистического управления. Поэтому в долгосрочной перспективе, возможно, будет целесообразно интегрировать некоторые соответствующие регистры в структуру НСУ.

валидации предоставленных данных на основе выводов этапа «Источник». Помимо базовой валидации, этот этап включает в себя любую обработку, необходимую для установления качества предоставленных данных по сравнению с тем, что ожидалось, а также сопоставления с альтернативными источниками;

- *этап «Процессы»:* Процессы, часто осуществляемые в отношении административных источников данных, в контексте переписей с целью преобразования данных для использования в рамках переписи и/или повышения качества. Выявленные процессы включают связывание данных; построение статистических регистров и методологию «признаков жизни»; регистрацию с использованием административных данных; методы сопоставления качества переменных в разных источниках; и редактирование и импутацию;
- *этап «Итоговые материалы»:* Общая оценка качества материалов переписи, произведенных с использованием административных данных. Хотя с концептуальной точки зрения это не отличается от оценки итогов традиционной переписи, в настоящем Руководстве предпринята попытка определить возможные отличия.

D. Параметры качества

21. Как отмечалось ранее, под качеством статистики и административных данных понимаются многочисленные параметры, которые не сводятся к ошибкам репрезентативности или измерения, например, статистика, которая является точной, но устаревшей, имеет ограниченное применение. Определенные ЕСС параметры качества включают 1) релевантность, 2) точность и надежность, 3) своевременность и пунктуальность, 4) доступность и ясность и 5) согласованность и сопоставимость⁸. Однако для оценки административных данных эти «стандартные параметры качества не всегда применимы» (Daas et al., 2008, p. 2). С другой стороны, они охватывают все значимые параметры качества административных данных. В нижеследующих таблицах указаны параметры для оценки административных источников, используемые в настоящем Руководстве, на каждом из вышеупомянутых этапов.

Таблица 1

Параметры качества на этапе «Источник»

ПАРАМЕТР КАЧЕСТВА	ОПРЕДЕЛЕНИЕ
Релевантность и точность	Степень, в которой административный источник данных удовлетворяет потребностям переписи. Степень совпадения с целевой совокупностью переписи, переписными концепциями и определениями (релевантность). Степень того, насколько корректно данные описывают явление, которое они призваны измерять (точность).
Своевременность	Промежуток времени между концом отчетного периода, к которому относится информация, и датой, когда информация становится доступной для НСУ.
Согласованность и сопоставимость	Степень, в которой административный источник может быть успешно объединен с другими источниками, используемыми в переписи, в том числе с точки зрения возможности связывания.

⁸ Альтернативные параметры используются различными НСУ (см., например, Statistics Canada, 2017, Australian Bureau of Statistics, 2009). В целом эти альтернативные подходы охватывают примерно одно и то же содержание, хотя и используют различную терминологию или классификации.

<i>ПАРАМЕТР КАЧЕСТВА</i>	<i>ОПРЕДЕЛЕНИЕ</i>
Доступность	Легкость, с которой НСУ может получать административные данные, охватывающая влияние любых ограничений, приемлемость использования для общественности, легкость передачи и получения данных, а также наличие метаданных.
Институциональная среда	Организационные факторы, влияющие на способность поставщика данных обеспечивать ожидаемое качество данных. Охватывает прочность отношений, предыдущий опыт, наличие официальных соглашений, риски, связанные со статусом поставщика, и стандарты качества поставщика.

Таблица 2

Параметры качества на этапе «Данные»

<i>ПАРАМЕТР КАЧЕСТВА</i>	<i>ОПРЕДЕЛЕНИЕ</i>
Валидация и согласование	Файлы данных, предоставленные в НСУ, имеют машиночитаемый формат. После передачи данных в НСУ проводятся дополнительные мероприятия по валидации и согласованию данных для подтверждения того, что ожидаемые переменные/единицы/отчетный период были предоставлены, и для обеспечения согласованности обработки НСУ данных на все этапах использования в целях переписи.
Точность и надежность	Точность, полнота (для переменных и охвата населения) и согласованность предоставленных данных соответствуют требованиям конкретного случая использования в целях переписи, для которого они будут использоваться. Сопоставление с альтернативными источниками позволяет определить приемлемые уровни погрешности измерения или репрезентативности.
Своевременность и пунктуальность	Своевременность и пунктуальность предоставленных данных соответствуют требованиям конкретного случая использования в целях переписи, для которого они будут использоваться.
Возможность связывания	В наличии имеются адекватные связующие переменные (т. е. либо единые уникальные идентификаторы, либо комбинация переменных, которые позволяют идентификацию), и они имеют достаточное качество, чтобы обеспечить связывание данных.

Таблица 3

Параметры качества на этапе «Процессы»

<i>ПАРАМЕТР КАЧЕСТВА</i>	<i>ОПРЕДЕЛЕНИЕ</i>
Точность связывания записей	В тех случаях, когда производится связывание нескольких источников (друг с другом или с ответами на вопросы переписи), это связывание является точным и объективным, повышая тем самым общее качество переписной методологии и/или набора переписных данных.

<i>ПАРАМЕТР КАЧЕСТВА</i>	<i>ОПРЕДЕЛЕНИЕ</i>
Охват и согласованность статистических регистров и переписей на основе административных данных	В тех случаях, когда имеются регистры переписной совокупности (или подсовокупностей) или когда административная информация используется в дополнение к данным переписи, они в достаточной степени охватывают целевую совокупность/переменные, повышая тем самым общее качество переписной методологии и/или набора переписных данных.
Точность устранения расхождений	В тех случаях, когда различные источники связываются друг с другом и в них имеются одни и те же атрибуты, методы выбора между источниками повышают общее качество переписной методологии и/или набора переписных данных.
Точность редактирования и импутации	В тех случаях, когда переменные/единицы переписи получены/созданы с помощью методов импутации или моделирования, полученный производный результат является точным и объективным, тем самым повышая общее качество переписной методологии и/или набора переписных данных.

Таблица 4

Параметры качества на этапе «Итоговые материалы»

<i>ПАРАМЕТР КАЧЕСТВА</i>	<i>ОПРЕДЕЛЕНИЕ</i>
Релевантность	Степень, в которой статистические материалы удовлетворяют текущие и потенциальные потребности пользователей.
Точность и надежность	Близость оценочного результата к неизвестному истинному значению — насколько они надежны во времени и по географии.
Своевременность и пунктуальность	Промежуток времени между публикацией и периодом, к которому относятся данные, а также временной лаг между фактической и запланированной датой публикации.
Доступность и ясность	Доступность и ясность означают действия, предпринятые для того, чтобы помочь пользователю найти и понять данные, которые его или её интересуют.
Согласованность и сопоставимость	Степень, в которой данные могут быть сопоставимы во времени и по различным темам. Степень схожести данных, полученных из разных источников или с помощью разных методов, но относящихся к одному и тому же явлению.

Источник: Eurostat, 2013 и 2018.

III. Этап «Источник»

22. В данной главе приводятся рекомендации по ключевым параметрам качества, процессу оценки и соответствующим инструментам и показателям для оценки качества источников административных данных, которые будут использоваться при проведении переписи, в случае как их первого применения, так и регулярного их повторного предоставления НСУ. Обычно на этом этапе данные недоступны. Однако, поиск информации об источниках административных данных начинается, скорее

всего, в ходе обменов и ознакомительных встреч между НСУ и поставщиком источника административных данных.

23. Оценка на этом этапе должна привести к выработке рекомендации относительно того, следует ли продолжать осуществление инициативы по получению данных (или же продолжить использование этого источника данных). Если будет принято решение о дальнейших действиях, то предоставленные административные данные будут подвергнуты более подробной оценке на этапе «Данные».

24. Необходимо оценивать качество источников как при первом получении административных данных, так и в каждом случае, когда они вновь предоставляются НСУ. Это связано с тем, что характеристики любого ранее предоставленного набора данных могут претерпевать изменения с точки зрения концепций, классификации, методов сбора и т. д.

A. Параметры качества источника

25. К параметрам качества данных, которые должны быть изучены на данном этапе оценки качества, относятся релевантность; своевременность; согласованность и сопоставимость; доступность и интерпретируемость; и институциональная среда. Эти параметры описываются ниже, а процессы, инструменты и показатели для оценки представлены в последующих разделах. Следует отметить, что невозможность достичь минимально приемлемого качества по одному из параметров не может быть компенсирована качеством других параметров.

1. Релевантность и точность

26. Релевантность отражает степень, в которой источник административных данных удовлетворяет потребностям НСУ с точки зрения предполагаемого использования. Для того чтобы считаться релевантным, источник административных данных должен, например, отвечать целям его использования. Это может быть связано с сокращением расходов или нагрузки на респондентов; повышением качества материалов переписи; или подготовкой расширенных или новых материалов переписи. Для этого административный источник должен быть репрезентативным в плане представляющей интерес для переписи совокупности (целевая совокупность), а данные измерения, полученные от населения, должны соответствовать потребностям переписи. Ключевой частью оценки релевантности является понимание контекста, в котором собирались административные данные.

27. В рамках оценки релевантности рассматриваются также вопросы точности административных данных. Под точностью понимается степень того, насколько корректно данные описывают явление, которое они призваны измерять. Важно понять, как сбор, обработка и обеспечение качества, осуществляемые административной организацией, могут влиять на точность получаемых данных и их полезность.

2. Своевременность

28. Своевременность означает период времени между датой, к которой информация относится, и датой, когда информация становится доступной для НСУ. Своевременность информации влияет на ее релевантность.

3. Согласованность и сопоставимость

29. Согласованность отражает степень, в которой административные данные могут успешно комбинироваться с данными из других источников, используемых НСУ, т. е. с данными переписи, в рамках широкой аналитической основы, во времени. Использование стандартных концепций, классификаций и целевых совокупностей способствует согласованности как внутри переписей, так и между ними. Таким образом, требуется четкое понимание оперативных определений, используемых административным поставщиком, цели сбора данных и влияния на сопоставимость

изменений в административном источнике во времени и по предметным областям при оценке согласованности.

30. Часто требуется связать административный источник на уровне статистической единицы переписи, чтобы интегрировать данные в план переписи. Поэтому вопрос о сопоставимости идентификаторов, используемых в различных источниках данных, которые должны быть связаны между собой, является одним из параметров согласованности.

4. Доступность и интерпретируемость

31. Доступность означает легкость, с которой НСУ может получать (и понимать) соответствующие административные данные во всей их полноте. Это включает в себя понимание любых ограничений (правовых и тех, которые налагаются поставщиком); приемлемость для общественности; легкость передачи и получения данных (пригодность формата или носителя для передачи данных и затрат); а также наличие и ясность документации и метаданных. Крайне важно, чтобы использование источника административных данных основывалось на правовой базе, которая дает НСУ безусловное право на доступ и использование данных и метаданных для статистических целей.

5. Институциональная среда

32. Под институциональной средой понимаются организационные или институциональные факторы, которые могут влиять на способность поставщика данных предоставлять данные ожидаемого качества и в соответствии с согласованным графиком (пунктуальность). Это включает в себя прочность отношений с поставщиком данных, включая эффективность каналов связи и то, насколько чутко поставщик реагирует на запросы НСУ. Это также включает в себя наличие (или возможность наличия) официальных соглашений и риски, связанные со статусом и сложностью организации-поставщика. Кроме того, это включает в себя стандарты и процедуры качества, принятые административной организацией (административными организациями).

В. Инструменты и показатели

33. Качество административного источника должно оцениваться по параметрам качества, указанным в разделе выше. Ниже приводятся рекомендации по процессу оценки, включая инструменты и показатели для оценки административного источника для использования в целях переписи.

1. Релевантность и точность

34. Понимание различий между административной совокупностью и требуемой переписной совокупностью, а также между параметрами/переменными в административном источнике и требуемыми переписными характеристиками, имеет важное значение для оценки релевантности и точности. Погрешность, возникающая в результате этих различий, называется, соответственно, погрешностью репрезентативности и измерения (Zhang, 2012). На этапе оценки «Источник» можно получить некоторое представление об этих ошибках и их влиянии на релевантность на основе метаданных о цели и методах сбора данных поставщика. Влияние погрешностей репрезентативности и измерения на точность и надежность также учитывается на этапах «Данные» и «Процессы».

а) Целевая переписная совокупность (репрезентативность)

35. Для оценки релевантности НСУ должно определить, соответствует ли набор объектов в источнике административных данных единицам совокупности, представляющей интерес для переписи (целевая совокупность). Объектом является базовый элемент совокупности, по которому собирается информация; это может быть человек, домохозяйство, жилище, событие или операции и т. д. Для установления

релевантности в плане репрезентативности предлагаются следующие показатели. По каждому показателю приводится ряд вопросов, которые могут помочь в проведении оценки:

- **Согласованность (объектов) с целевыми единицами переписи:**
 - i) Насколько административные объекты сопоставимы с целевыми единицами переписи?
 - ii) Какие определения, методы и процессы используются для идентификации и включения объекта в источник?
 - iii) Существуют ли какие-либо законы или правила, определяющие объекты?
 - iv) Проводятся ли держателем данных какие-либо проверки для обеспечения актуальности определений?
 - v) В случае несовпадения с переписными единицами возможно ли преобразование, способное обеспечить удовлетворение потребностей переписи?
- **Охват (набора объектов) в сопоставлении с целевой совокупностью переписи:**
 - i) Соответствует ли охват объектов потребностям переписи населения?
 - ii) Имеются ли доказательства недостаточного охвата (объекты, которые отсутствуют в источнике, но являются частью целевой совокупности переписи) и чрезмерного охвата (объекты, которые имеются в источнике, но не являются частью целевой совокупности переписи), которые способны повлиять на полезность источника?
 - iii) Существуют ли различия между географическими зонами, обусловленные различиями в практике держателя данных или законодательством, которые необходимо учитывать?
 - iv) Существуют ли какие-либо правила, законодательные или нормативные требования, включая штрафы за несоблюдение, которые могут влиять на включение и исключение объектов из источника?
 - v) Какие методы и процессы используются держателем данных для включения новых объектов, которые отвечают требуемым критериям/определениям включения (например, процедуры постановки на учет), и для исключения объектов, которые больше не соответствуют целевой совокупности для административного источника (например, процедуры снятия с учета)?
 - vi) В случае наличия ошибок охвата, существуют ли другие источники данных, которые можно использовать в сочетании с источником для преодоления проблемы, например, недостаточного или избыточного охвата источником?

b) Переписные переменные/характеристики (измерение)

36. Для оценки релевантности НСУ должно также определить, соответствует ли информация, собранная от объектов в источнике административных данных, потребностям переписи в плане целевых концепций (например, этническая принадлежность, статус занятости, размер домохозяйства, статус пользования жильем и т. д.). Для определения релевантности в плане измерения предлагаются следующие показатели:

- **Наличие целевых переменных/характеристик:**
 - i) Содержит ли административный источник необходимые переменные исходя из предполагаемого использования данных в целях переписи?
 - ii) Охватывают ли в широком смысле переменные/характеристики соответствующий учетный период переписи?

- Соответствие концепций, определений и классификации переменных потребностям переписи:
 - i) Соответствуют ли административные концепции, определения и классификации потребностям переписи?
 - ii) Существуют ли различия между идеальными целевыми концепциями держателя данных и тем, что на самом деле достигается с помощью их оперативного целевого показателя, используемого при сборе данных?
 - iii) В случае наличия расхождений в концепциях, определениях и классификациях переписи, возможно ли преобразование для удовлетворения потребностей переписи?
- Погрешность согласования /измерения по отношению к учетному периоду переписи:
 - i) Какова частота сбора данных по переменной/характеристике?
 - ii) Существуют ли известные лаги между событием или явлением, происходящим или фиксируемым в административном источнике (например, родители могут не регистрировать рождение в течение нескольких недель в национальном регистре рождений)?
 - iii) Фиксируются ли метки времени в источнике данных для указания того, к какому периоду относится тот или иной элемент данных?
 - iv) Существуют ли какие-либо стимулы или сдерживающие факторы для субъекта данных в плане обновления информации о себе по мере изменения своих обстоятельств/изменения информации в административном источнике (например, льготы или штрафы в связи с выполнением/невыполнением соответствующих требований)?
- Качество сбора и потенциальная погрешность измерения в сопоставлении с концепциями переписи:
 - i) Какую цель преследует держатель данных при сборе данных и как это может влиять на качество данных?
 - ii) Существуют ли какие-либо правовые обязательства, цели или стимулы (или их отсутствие), которые могут влиять на качество сбора данных?
 - iii) Возникают ли в процессе сбора данных держателем данных какие-либо проблемы в плане качества переменных, включая возможность любого смещения? Речь может идти о том, что данные регистрируются опосредованно и, следовательно, не сообщаются непосредственно субъектом данных (что увеличивает вероятность ошибок при регистрации);
 - iv) Какие процедуры существуют для валидации и проверки данных при вводе держателем данных?
 - v) Существуют ли какие-либо стимулы или сдерживающие факторы для субъектов данных в плане предоставления держателю данных полной и точной информации?
- Качество обработки данных и возможность ошибки при обработке держателем данных:
 - i) Можно ли считать, что проводимая держателем данных обработка обеспечивает соответствие качества полученных данных потребностям переписи?
 - ii) Какие проверки проводит держатель данных для обеспечения качества?
 - iii) Подвергаются ли данные редактированию или импутации? Если да, то когда и как и имеется ли в источнике данных показатель, позволяющий определить, когда производились редактирование и импутация?

iv) Существуют ли какие-либо правила, нормы или стимулы для держателя данных, которые могут влиять на способ обработки данных?

37. На этапе «Источник» оценка по показателям обычно основывается на качественной оценке (например, путем указания того, насколько удовлетворяет требованиям каждый показатель (полностью, частично или не удовлетворяет, с объяснением причин на основе ответов на заданный вопрос). Количественная оценка погрешности репрезентативности и измерения осуществляется на этапе «Данные» (на основе анализа данных) по параметру точности и надежности.

38. Оценка по показателям должна служить основой для принятия решения (часто с опорой на опыт и экспертное суждение) об использовании (или дальнейшем использовании) источника в целях переписи. При принятии решения следует учитывать, может ли источник данных содействовать удовлетворению потребностей переписи (например, сокращению расходов и нагрузки на респондентов, улучшению и совершенствованию материалов переписи), с учетом любых расходов или рисков (см. ниже параметры институциональной среды и доступности).

39. В литературе описаны различные системы качества, которые предлагают аналогичные показатели, как в настоящей главе, с использованием различных параметров качества, а также наборы вопросов и системы балльной оценки для информационного обеспечения оценки (см. например, работы Daas et. al., 2009; SN-MAID, 2014; Iwig et. al., 2013; Statistics Canada's Administrative Data Evaluation Guide (Lavigne & Nadeau, 2014); Statistics Austria's Quality Assessment of Administrative Data, Documentation of Methods Framework (Statistics Austria, 2019)).

2. Своевременность

40. Административный источник может охватывать соответствующий период времени для целей переписи, но для того, чтобы он был полезен, он также должен быть доступен своевременно с учетом графика переписи. Для оценки своевременности предлагается следующий показатель:

- Своевременность и периодичность предоставления данных в сопоставлении с потребностями переписи:
 - i) Каков лаг между моментом события или явления и его отражением в административном источнике?
 - ii) Каков лаг между датой окончания учетного периода административных данных и датой, когда данные могут быть предоставлены НСУ?
 - iii) Как часто данные могут предоставляться в НСУ в сопоставлении с потребностями переписи?
 - iv) Существуют ли какие-либо требования с точки зрения метода передачи и требуемых форматов и структур данных, используемых НСУ, которые могут повлиять на сроки поставщика данных?
 - v) Может ли НСУ структурировать и обрабатывать данные своевременно для нужд переписи начиная с момента, когда данные становятся доступными?

41. В тех случаях, когда данные вряд ли будут доступны вовремя, НСУ, возможно, пожелает выяснить, может ли предварительная версия набора данных быть предоставлена досрочно. В таких случаях набор данных может быть неполным и подверженным более высоким уровням погрешности. Таким образом, может быть установлен компромисс между своевременностью данных и их точностью.

42. Как указано ниже в отношении параметра институциональной среды, важно прописать даты передачи данных и учетные периоды данных в официальных соглашениях с поставщиком данных. Несмотря на то, что данные могут быть доступны своевременно исходя из требований держателя данных, они не обязательно будут переданы своевременно НСУ, в то время как последнее несет официальную ответственность за своевременное предоставление данных переписи.

3. **Согласованность и сопоставимость**

43. Важно оценить, в какой степени административный источник может успешно комбинироваться с другими источниками данных для использования в целях переписи. Информация, собранная по показателям, определенным для оценки релевантности, может быть использована для оценки согласованности. Это включает в себя информацию о различиях между основополагающими концепциями, определениями, классификациями и методами источника административных данных и других источников данных для комбинированного использования в рамках переписи.

44. Для проведения полностью регистровой переписи важно проанализировать характеристики переписи и источник административных данных; провести сопоставление и определение степени, в которой информация, содержащаяся в источнике административных данных, облегчает разработку соответствующих характеристик переписи. В частности, НСУ должно выяснить, соответствуют ли данные, содержащиеся в регистрах, определению характеристик переписи. В случае частичного или полного несоответствия, НСУ следует изучить причины несоответствия между характеристиками переписи и информацией, доступной в источнике административных данных.

а) **Сопоставимость**

45. Административные данные подвержены изменениям и различиям с течением времени и по предметным областям в связи с изменениями в законодательстве, регулировании и процедурах, которые могут влиять на концепции, определения, классификации и охват источника. В более общем плане изменения могут влиять на все показатели, касающиеся репрезентативности и измерения, как указано в разделе, посвященном параметру релевантности. Это имеет особое значение для переписи, для которой стабильность во времени может быть ключевым соображением. Для оценки сопоставимости предлагается следующий показатель:

- Сопоставимость во времени и по предметным областям:
Существуют ли какие-либо изменения во времени или различия между предметными областями (например, географическими зонами), влияющими на них:
 - i) Являются ли определение объекта и охват объектов в административном источнике релевантными для целей переписи?
 - ii) Являются ли концепции, определения и классификации, связанные с переменными административного источника, релевантными для целей переписи?
 - iii) Имеются ли процедуры сбора, обработки и обеспечения качества данных, которые способны повлиять на качество источника для целей переписи?

б) **Возможность связывания**

46. Одним из соображений, касающихся согласованности и сопоставимости, является легкость связывания источника административных данных с другими соответствующими наборами данных в целях переписи. Для оценки возможности связывания источника предлагаются следующие показатели:

- Наличие уникального ключа для связывания:
 - i) Содержит ли источник уникальный идентификатор, совпадающий с уникальными ключами, необходимыми для связывания в целях переписи?
 - ii) Доступен ли идентификатор по всем соответствующим объектам в источнике?

- Наличие уникальной комбинации переменных для связывания:
 - i) Содержит ли источник уникальную комбинацию переменных (например, имя, возраст и адрес), которые можно было бы использовать для связывания в целях переписи?
 - ii) Имеется ли уникальная комбинация переменных для каждого объекта в источнике?

47. Качество связующих переменных также оценивается на этапе «Данные», а качество процесса связывания рассматривается в рамках этапа «Процессы».

4. Доступность

48. Для оценки параметра доступности предлагаются следующие показатели:

- ограничения на доступ к данным и их использование;
- приемлемость для общественности;
- легкость передачи и получения данных;
- интерпретируемость источника — ясные и полные метаданные.

49. В нижеследующих разделах приводится подробная информация, необходимая для оценки по каждому из показателей.

а) Ограничения на доступ к данным и их использование

50. Важно выявить любые ограничения, которые могут влиять на способность НСУ получать доступ к административному источнику и использовать его. Например, существующие ограничения в целях защиты данных, прописанные в правовых актах, могут накладывать определенные ограничения на сбор и обработку данных (особенно, если данные защищены дополнительными мерами безопасности или законами на государственном уровне). Такие правовые акты могут касаться конкретных источников данных или могут быть более общими, разрешающими НСУ доступ к таким источникам данных по мере необходимости, при условии согласия держателя данных. Держатель данных может также наложить дополнительные ограничения на предоставление данных и разрешенное использование. Речь может идти о следующих ограничениях:

- исключение записей или переменных;
- процедуры обработки для предупреждения идентификации (до предоставления), такие как шифрование идентификаторов, помехи, объединение или цензурирование по верхней границе предоставленных данных;
- ограничения на использование данных;
- ограничения на сохранение данных и правила исключения/удаления;
- правила о методах раскрытия информации, которые должны применяться НСУ, влияющие на результаты переписи.

51. НСУ следует выявить и описать любые применимые ограничения, по которым можно провести оценку воздействия (и рисков воздействия) ограничений на использование административного источника в целях переписи. В рамках оценки НСУ следует также рассмотреть вопрос о том, способно ли оно обеспечить соблюдение ограничений. Это может включать в себя технические и процедурные меры защиты, которые должно внедрить НСУ. Эти меры защиты, как правило, являются частью Меморандума о взаимопонимании (MoV) или Соглашения о безопасности данных с владельцем/поставщиком данных. В частности, в MoV может быть прописано, как будет защищена позволяющая установить личность информация (ПУЛИ).

б) Приемлемость для общественности

52. Предоставление НСУ доступа к источнику данных для использования в целях переписи также может зависеть от приемлемости такого шага для общественности. Общественность должна понимать и поддерживать новые подходы и использование своей информации или, по крайней мере, не относиться враждебно к этому. Если общественность выступает против использования административного источника данных, существует риск для качества. Например, это может отрицательно сказаться на сотрудничестве общественности с переписью или административным источником, используемым в целях переписи. Поэтому НСУ следует обеспечивать прозрачность в отношении использования административных источников в целях переписи, подчеркивая преимущества для общественности и одновременно обеспечивая гарантии неприкосновенности частной жизни и безопасности.

53. Для оценки приемлемости для общественности можно использовать следующие инструменты или процессы:

- консультации с общественностью или вовлечение общественности в обсуждения;
- оценка воздействия на неприкосновенность частной жизни (ОВНЧЖ);
- оценка этики данных.

54. НСУ может проводить консультации с общественностью или привлекать её к обсуждениям по вопросам использования административных данных в целях переписи (или проведения других статистических исследований или подготовки материалов). Это может принимать различные формы, включая официальные консультации, анкетирование (через опросы или путем приглашения НСУ высказаться по соответствующим вопросам через его веб-сайт), качественное исследование общественного мнения или использование представительных групп граждан. Представительные группы граждан имеют целью собрать вместе представителей общественности (быть репрезентативными для определенной совокупности или отражать различные группы интересов) для оценки их взглядов и мнений.

55. Оценка воздействия на неприкосновенность частной жизни (ОВНЧЖ) представляет собой формальный процесс, результатом которого является документ, описывающий процесс, выводы и результаты, который помогает НСУ изучить воздействие новой программы или услуги (или предлагаемых программ и планов) на частную жизнь людей. В качестве инструмента управления рисками, используемого на этапе планирования программы или инициативы по оказанию услуг, ОВНЧЖ помогают организациям более полно учитывать последствия заданного предложения для конфиденциальности. ОВНЧЖ используются также для того, чтобы контроллеры данных могли выполнять свои обязательства в соответствии с Общим регламентом защиты данных (в соответствии с европейским законодательством). ОВНЧЖ может применяться к различным видам предполагаемого НСУ использования источника данных в плане переписи.

56. Оценка этики данных проводится для установления того, является ли доступ к общедоступным данным, их использование и обмен ими в исследовательских и статистических целях этичным и отвечает ли это принципам общественного блага. НСУ могут использовать инструмент самооценки этики (см. например, UKSA, 2020), но могут также использовать официальный орган для предоставления экспертных заключений или одобрения, например, Консультативный комитет по этике данных (например, Национальный консультативный комитет по этике данных при Государственном статистике)⁹.

⁹ Дополнительную информацию можно получить по адресу <https://uksa.statisticsauthority.gov.uk/about-the-authority/committees/national-statisticians-data-ethics-advisory-committee/#:~:text=The%20National%20Statistician%E2%80%99s%20Data%20Ethics%20Advisory%20Committee%20%28NSDEC%29,advise%20the%20National%20Statistician%20on%20the%20ethical%20>.

57. Выводы, сделанные по итогам консультаций и обсуждений с общественностью, ОВНЧЖ и оценок этики, могут помочь НСУ оценить приемлемость использования административных источников в целях переписи (и для формирования другой статистики НСУ).

с) Легкость передачи и получения данных

58. Поставщик данных может использовать весьма различные модели данных, форматы, схемы, программное и аппаратное обеспечение, знакомые НСУ. Речь в частности идет о методах хранения и передачи данных (включая меры безопасности при их передаче). Структуры данных также могут быть сложными, а размеры файлов — чрезвычайно большими (особенно в случае данных об операциях). Важно, чтобы НСУ понимало такие различия и сложности для того, чтобы оценить, смогут ли системы НСУ принять и «проглотить» такие наборы данных. Этот процесс может также включать переговоры с поставщиком о разработке процессов и систем для облегчения передачи наборов данных в формате, отвечающем потребностям НСУ. Однако это может быть трудоемким и дорогостоящим процессом.

59. В более общем плане затраты являются ключевым фактором, который следует учитывать при оценке легкости доступа. Сюда могут входить затраты поставщика данных или затраты НСУ на развитие своего потенциала для получения административного набора данных (например, если необходимо приобрести новое программное обеспечение или аппаратное оборудование). Важно оценивать любые затраты в сопоставлении с ожидаемой полезностью, которую обеспечит новый административный источник.

60. На практике подробная информация об условиях передачи данных в НСУ, включая структуру и формат файлов (например, неструктурированные файлы, реляционная база данных); SAS, Excel или текстовые форматы и т. д.), переменные, частоту предоставления и даты передачи, стандарты данных и согласованные расходы, будет включаться в соглашения об обмене данными или соглашения о передаче между НСУ и поставщиком. Такие соглашения будут подписываться уполномоченными руководителями каждой из организаций.

d) Интерпретируемость источника — ясные и полные метаданные

61. Оценка интерпретируемости касается наличия и доступности всеобъемлющих и четких метаданных и документации об административном источнике. Без этого невозможно понять и оценить административный источник на предмет предполагаемого использования. Метаданные должны включать в себя подробную информацию о:

- административной организации и цели сбора;
- используемых концепциях, определениях, классификациях и протоколах;
- методах и процедурах сбора, обработки, валидации и обеспечения качества;
- отчетных единицах и переменных, включая словари данных, структуры файлов, форматы и связи внутри данных.

62. Эта информация важна для оценки по другим параметрам качества, описанными в данной главе. Часто бывает так, что четкие и полные метаданные не будут существовать по всем аспектам административного источника на начальном этапе изучения источника для использования НСУ. Поэтому необходимо работать с поставщиком данных в целях построения соответствующих метаданных. Это зависит от хорошей связи с поставщиком и готовности поставщика работать с НСУ (см. ниже «Институциональная среда»). В зависимости от сложности административного источника, НСУ может принять решение о командировании сотрудников для работы в административной организации в целях углубления понимания источника.

5. Институциональная среда

63. НСУ полностью зависит от поставщика данных в плане сбора, обработки и предоставления административных данных в ожидаемом качестве и в соответствии с согласованным графиком. НСУ также зависит от качества информации, предоставляемой поставщиком о данных и о любых прогнозируемых изменениях в данных. Поэтому важно оценить уверенность в способности поставщика данных удовлетворять эти требования. Для оценки институциональной среды предлагаются следующие показатели:

- прочность отношений с поставщиком;
- предыдущий опыт работы с поставщиком;
- наличие официальных соглашений;
- риски, связанные со статусом поставщика;
- стандарты качества поставщика.

64. *Прочность отношений.* Должны быть созданы процессы для управления отношениями с поставщиком данных, обеспечивающие постоянный диалог. Они должны включать в себя механизмы для:

- сообщения требований НСУ поставщику;
- своевременное информирование (поставщиком) о любых изменениях, которые могут повлиять на источник данных (например, об изменениях в законодательной базе в отношении данных, в концепциях и классификациях, а также в процессах и процедурах сбора, управления и сообщения данных);
- направления поставщику любых вопросов об источнике данных;
- обеспечения обратной связи с поставщиком, которая может привести к улучшению источника.

65. *Предыдущий опыт.* Речь идет о том, насколько чутко поставщик реагировал на запросы НСУ, возникали ли какие-либо проблемы с предыдущими поставками данных (например, несвоевременная передача, непредвиденные ошибки), предоставлял ли поставщик точную информацию об источнике данных в прошлом (это могло быть установлено в результате проверок, проведенных НСУ на более позднем этапе).

66. *Официальные соглашения.* Речь идет о наличии или возможной разработке письменных соглашений (юридических или иных), охватывающих следующие вопросы:

- роли и обязанности НСУ и поставщика. Речь может также идти о том, играет ли НСУ определенную роль в утверждении любых изменений в административном источнике, используемом (или планируемом к использованию) в целях переписи;
- правовая основа для предоставления данных и любые требования безопасности/конфиденциальности;
- спецификация требований в соответствии с Соглашением об обмене/передаче данных, о котором шла речь в разделе «Легкость передачи и получения данных» выше.

67. *Статус поставщика.* Риски, связанные со статусом поставщика, должны оцениваться НСУ путем выяснения того, является ли поставщик признанной, стабильной и авторитетной организацией. При этом следует рассмотреть вопрос о том, существует ли какая-либо правовая или нормативная основа для выполнения поставщиком административной функции, которая придавала бы уверенность в устойчивости и качестве источника. Следует также учитывать риски, связанные со сложностью организации(ий)-поставщика(ов), участвующего(их) в сборе, обработке и сообщении данных источника. Например, может быть задействовано несколько органов или организаций, каждая из которых влияет на качество передаваемых окончательных данных.

68. *Стандарты качества поставщика.* Должна быть проведена оценка того, сможет ли поставщик соответствовать ожиданиям НСУ в отношении качества. При этом следует учитывать информацию о принципах, стандартах и руководствах, используемых поставщиком для обеспечения качества, включая действующую процедуру, охватывающую сбор, обработку и передачу данных НСУ. Информация о том, как поставщик проверяет, соблюдаются ли стандарты, является ценной; это могут быть результаты внутреннего или внешнего аудита, проводимого регулирующими органами или профессиональными организациями. Поставщик также может составлять отчеты о качестве, которые должны рассматриваться НСУ. Более подробная оценка, основанная на ключевых аспектах административного источника, включена в приведенный выше параметр качества «релевантность».

69. После оценки поставщика данных в соответствии с вышеизложенными критериями, НСУ может оценить риски, связанные со своевременной передачей административных данных поставщиком, в соответствии с требуемыми критериями качества.

IV. Выводы и рекомендации

70. Административные данные могут использоваться в рамках различных методологий проведения переписи и для поддержки всех этапов процесса переписи, включая создание генеральной совокупности адресов, поддержку полевых операций, регистрацию населения, сбор данных по переменным переписи, обеспечение качества, редактирование и импутацию, моделирование и оценку. Их использование может обеспечить более частое и своевременное получение статистических данных о населении; повышение точности и надежности и значительное сокращение расходов и нагрузки на респондентов.

71. Однако до того, как административный источник сможет использоваться в целях переписи, необходимо оценить и преодолеть серьезные проблемы качества. Наиболее важной из них является то, что административные данные, как правило, не собираются для целей переписи. Исходя из этого НСУ может иметь малое влияние на используемые концепции и определения; целевую совокупность; процедуры сбора, обработки и обеспечения качества; и используемые методы, структуры и системы данных.

72. В Руководстве определены этапы оценки качества с учетом ряда параметров качества, а также предлагаются соответствующие инструменты и показатели, которыми пользователь может руководствоваться в процессе оценки. Применение Руководства призвано помочь читателям в принятии решений об использовании административных данных в целях переписи, поддерживая при этом процесс постоянной оценки и совершенствования. В Руководстве сформулирован ряд предложений и рекомендаций, которые кратко излагаются ниже.

A. Рекомендации

73. Были сформулированы следующие рекомендации:

- i. НСУ следует *выявить административные источники*, которые могут быть релевантными для целей переписи с учетом различных возможностей использования. Важно *определить ожидаемые или требуемые результаты использования источника*, на основании которых можно будет провести оценку релевантности. Они могут включать в себя повышение эффективности проведения переписи с точки зрения сокращения расходов и нагрузки на респондентов; повышение качества переписи; или подготовку новых или усовершенствованных материалов переписи. Центральное место в такой оценке отводится определению того, что административный источник должен предоставлять с точки зрения целевой совокупности и требуемых характеристик этой совокупности для использования в целях переписи. В полном варианте Руководства, включая практические примеры, приводятся

примеры использования административных данных в нескольких разных странах.

ii. *Отношения между НСУ и поставщиком административных данных имеют крайне важное значение. Они должны подкрепляться надежными механизмами связи, письменными соглашениями и отличным пониманием потребностей обеих сторон. Кроме того, должна существовать согласованная правовая основа для передачи и использования данных. Чтобы помочь выстроить такие отношения и обеспечить надежность передачи данных, НСУ следует определить области, которые принесут выгоды поставщику. Это могут быть механизмы обратной связи, помогающие поставщику лучше понять свои данные, путем сотрудничества в областях, представляющих общий интерес, или оказания поставщику помощи (путем использования его данных в ходе переписи) в поддержку расширения обеспечиваемого им общественного блага. Разумеется, обратная связь в отношении возможных проблем, связанных с качеством данных, несет в себе дополнительное преимущество, заключающееся в содействии постоянному повышению качества.*

iii. *НСУ следует взаимодействовать с поставщиком, чтобы получить углубленное представление об источнике данных. Это должно привести к созданию четких и всеобъемлющих метаданных об административном источнике. Метаданные обеспечат полезную справочную базу как для переписи, так и для любой другой статистики, которая может извлечь выгоду из использования источника. Полный вариант Руководства содержит подробные сведения о метаданных, которые должны быть собраны, а также различные ссылки на соответствующую литературу.*

iv. *Поскольку административные данные, как правило, не собираются для нужд переписи, важно, чтобы НСУ понимали и оценивали различия между целевыми совокупностями, концепциями, определениями и временными параметрами. В более общем плане необходимо тщательно оценивать согласованность и сопоставимость административного источника, а также его ограничения по различным параметрам качества. Это включает в себя возможность связывания источника, если это является требованием для использования в целях переписи. Эта оценка будет служить информационным подспорьем для этапов обработки, включая сопоставительный анализ и расчет, редактирование и импутацию, а также связывание и интеграцию источников (когда решения принимаются между источниками и по ним на основе их качества).*

v. *НСУ должно понимать любые ограничения и проблемы, связанные с получением и интеграцией данных административного источника в перепись. Это может касаться ресурсов и расходов; рисков, связанных со способностью поставщика осуществлять передачу данных своевременно и требуемого качества, а также приемлемости использования источника для общественности и пользователей данных переписи. В этом отношении существуют важные компромиссы, которые НСУ должно учитывать. В частности, ценность административного источника должна оцениваться с точки зрения его полезности для проведения переписи, с учетом усилий и рисков, связанных с получением и использованием данных.*

vi. *НСУ имеет ограниченный контроль над сбором и обработкой административных данных, которые могут быть подвержены изменениям в плане охвата и характеристик населения с течением времени. Это может быть обусловлено, например, правовыми, политическими, процедурными или системными изменениями, влияющими на данные и/или их передачу. Поэтому НСУ должно оценивать уровень риска и управлять им. Управление риском должно осуществляться путем работы с поставщиком данных в отношении потенциальных или планируемых изменений; гибкого и чуткого реагирования на изменения; а также уменьшения зависимости от какого-либо одного источника или элемента данных, когда это возможно, как за счет использования других источников данных, так и за счет адаптации процессов/методологий.*

vii. *Важно, чтобы общественность и пользователи данных понимали, как и почему в целях переписи используются административные данные.* Поэтому НСУ следует обеспечивать прозрачность в вопросах использования; предоставляя четкое обоснование выгод, определенных с учетом любых рисков и затрат (т. е. явное превышение выгод над издержками). Этого можно добиться с помощью *хорошей коммуникации, включая публикацию действующих процедур и политики, способствующих эффективному использованию и защите данных.*

viii. Включению источников административных данных в процесс формирования итогов переписи должно предшествовать проведение адекватно обеспеченного ресурсами исследования целесообразности, которое предоставит «апробацию концепции» планируемой интеграции административных данных в процесс формирования итогов переписи. Рекомендуется провести ряд пробных тестов (с использованием реальных данных) заблаговременно до начала основной переписи, чтобы убедиться в том, что непредвиденные проблемы выявлены, что даст достаточно времени для исправления или корректировки методов, процессов или систем.

ix. *Экспертный обзор* (работа с поставщиками данных и профильными экспертами) и *сопоставления между источниками и во времени имеют важное значение для выявления любых проблем с качеством* в источнике или регистре. Использование тщательно спланированных обследований (связанных с административными данными или регистрами) может иметь особенно важное значение для выявления и корректировки погрешностей охвата и измерения.

x. НСУ *следует фиксировать и публиковать результаты оценки и обеспечения качества на протяжении всего процесса переписи*, включая этапы «Данные», «Процессы» и «Итоговые материалы». Это позволит производителям и пользователям переписи оценивать и обеспечивать обратную связь, поддерживая постоянный диалог. Это важно для того, чтобы пользователи понимали сильные и слабые стороны и могли определить, был ли достигнут верный баланс между параметрами качества.

xi. НСУ *следует разработать свою собственную систему и стратегию обеспечения качества*, поддерживаемую четкой и исчерпывающей документацией и процедурами обучения. Настоящее Руководство служит полезной основой для решения этих задач, наряду со справочными материалами и практическими примерами, приводимыми в нем. Стратегия должна опираться на постоянную оценку и совершенствование административных данных в рамках планов и процедур проведения переписи. Это должно включать в себя коммуникационные связи между НСУ, пользователями и поставщиками данных.

В. Области дальнейших разработок

74. В Руководстве основное внимание уделяется оценке качества административных источников для использования в целях переписей, при этом в нем содержится некоторая информация о процессах, используемых для интеграции и преобразования данных в целях повышения их качества. Кратко освещается также качество итогов переписей, в которых используются административные данные. Тем не менее в Руководстве не описывается *более широкая концепция общей погрешности* или модель того, как погрешность из каждого источника превращается в погрешность в окончательных оценках переписи с учетом изменений в качестве, связанных с обработкой (которые могут уменьшить или увеличить погрешность).

75. Разработка такой модели, которая учитывает все источники погрешности, частично рассматривается в рамках модели общей погрешности, принятой Статистическим управлением Новой Зеландии (Reid et al., 2017). Эта модель основана на предложенном Личунь Чжаном (Zhang, 2012) расширении парадигмы Общей погрешности наблюдения (ОПН) (см. Groves and Lyberg, 2010; Biemer, 2010). Она

состоит из трех этапов, охватывающих: 1) оценку отдельных источников, 2) оценку интегрированного набора данных; и 3) расчет и оценку результата. Работа ESSnet KUMUSO по качеству многоисточниковой статистики (WP1 Quality, 2019) также служит полезной основой для оценки качества статистических материалов, основанных на множественных источниках (данные обследований и административные данные).

76. Это могло бы стать областью для дальнейших разработок и международного сотрудничества *с конкретным акцентом на то, как такая модель может быть применены к переписям*. Это может включать в себя изучение того, как можно разработать концепцию или модель общей погрешности и использовать их для оценки качества итогов переписи, опирающихся на множественные источники. Она также могла бы включать в себя работу по изучению того, каким образом воздействие (и смешанное воздействие) различных погрешностей на всех этапах переписи может способствовать принятию решений о наилучшем общем статистическом плане переписи.

77. И наконец, в Руководстве основное внимание уделяется оценке административных данных, однако существуют и *другие источники коммерческих данных, которые открывают возможности для использования в целях совершенствования или повышения качества переписной статистики* (например, геопространственные данные, данные мобильной телефонии). Предлагаемые в настоящем Руководстве этапы оценки, параметры, инструменты и показатели качества в значительной степени применимы к источникам, выходящим за рамки административных данных. Это также может быть областью, требующей проведения дальнейшей международной работы, с особым акцентом на то, отличаются ли, и если да, то каким образом, инструменты и методы оценки качества таких источников для использования в целях переписи от тех, которые обсуждаются в данном Руководстве.
