

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

Двадцать шестое совещание

Женева, 2–4 октября 2024 года

Пункт 2 d) предварительной повестки дня

**Пересмотр рекомендаций Конференции европейских
статистиков по проведению переписей населения
и жилищного фонда для раунда 2030 года:****Технология****Разработка рекомендаций по технологии переписи****Записка Целевой группы Конференции европейских статистиков
по технологии переписи****Резюме*

Настоящий документ содержит проект главы о технологии для Рекомендаций Конференции европейских статистиков (КЕС) по проведению переписей населения и жилищного фонда для раунда 2030 года, а также краткое описание внесенных изменений по сравнению с Рекомендациями для предыдущего раунда 2020 года. Основная цель документа заключается в получении от национальных экспертов по вопросам переписи комментариев и рекомендаций по предложенному тексту, чтобы обеспечить отражение в нем потребностей национальных статистических управлений и последних технологических и методологических разработок.

* В состав Целевой группы Конференции европейских статистиков по технологии переписи входят следующие члены: Андрес Викат (ЕЭК ООН, председатель), Режиc Пламондон (Канада), Джон Конвей (Ирландия), Ана Бинсток (Израиль), Йохан Эмберг (Норвегия), Эспен Андерсен (Норвегия), Галина Шевердова (Российская Федерация), Ханнес Ройтер (Евростат) и Сейффе Тадессе (Статистический отдел Организации Объединенных Наций).

Примечание: Употребляемые в настоящем документе обозначения не означают выражения со стороны Секретариата Организации Объединенных Наций какого бы то ни было мнения относительно правового статуса страны, территории, города или района или их властей или относительно делимитации их границ.



I. Введение

1. Каждые 10 лет Конференция европейских статистиков (КЕС) выпускает Рекомендации, которыми должны руководствоваться страны при проведении переписей населения и жилищного фонда. Эти рекомендации разрабатываются целевыми группами экспертов под эгидой Руководящей группы КЕС по переписям населения и жилищного фонда. При разработке Рекомендаций для раунда переписей 2030 года Целевая группа по технологиям переписи рассмотрела вопросы проведения переписи, связанные с технологией, за исключением географических информационных систем (ГИС), для которых была создана отдельная целевая группа.
2. В разделе II настоящего документа кратко описаны внесенные изменения по сравнению с Рекомендациями для предыдущего раунда 2020 года.
3. В разделе III представлен проект главы о технологии для Рекомендаций КЕС по проведению раунда переписей населения и жилищного фонда 2030 года.
4. Основная цель документа заключается в получении от национальных экспертов по вопросам переписи комментариев и рекомендаций по предложенному тексту, чтобы обеспечить отражение в нем потребностей национальных статистических управлений и последних технологических и методологических разработок.

II. Краткое описание изменений по сравнению с Рекомендациями 2020 года

5. Отправной точкой для составления текста раздела III настоящего документа послужила глава II «Технология» в Рекомендациях КЕС для раунда переписей 2020 года¹. С учетом произошедшего с того времени развития технологии и практики раунда 2020 года для раунда переписей 2030 года были добавлены семь новых частей:
 - a) Рекомендуемые компоненты тестирования (раздел III.B.3);
 - b) Разработка электронного вопросника (III.C.4);
 - c) Технологии для поддержки переписи людей с нарушениями и людей, не пользующихся цифровыми технологиями (III.C.5);
 - d) Интерактивные онлайн-платформы для распространения (III.D.3);
 - e) Хранение и архивирование (III.D.4);
 - f) Облачные технологии (III.D.5);
 - g) Технологии использования административных данных (III.E).
6. Наибольшее внимание уделено технологическим решениям, получившим широкое распространение в практике проведения переписей раунда 2020 года — электронным инструментам сбора данных, облачным технологиям и комплексным решениям для автоматизации всех этапов переписных работ — от планирования переписи до распространения ее результатов.
7. Технологии, связанные с географическими информационными системами (ГИС), предлагаются в качестве отдельной главы Рекомендаций, проект которой представлен в отдельном документе специальной рабочей группой.
8. Вопросам безопасности данных уделяется больше внимания на протяжении этой главы. Они касаются необходимости тестирования интегрированной системы переписи и ее отдельных компонентов, сертификации на предмет кибербезопасности, разделения критических инфраструктур, варианта предоставления ответов через Интернет, использования портативных устройств и хранения данных. Раздел о

¹ ЕЭК ООН, 2015. «Рекомендации Конференции европейских статистиков по проведению переписей населения и жилищного фонда 2020 года». Женева: Организация Объединенных Наций.

безопасности и конфиденциальности включен в новую часть, посвященную технологиям использования административных данных.

9. Хорошо зарекомендовавшие себя технологии, такие как опрос по телефону, системы ввода с клавиатуры и оптические системы считывания, не рассматриваются подробно, поскольку эти системы уже хорошо изучены и документированы. Исходя из опыта проведения раунда 2020 года, не рекомендуется проводить автоматизированный опрос по телефону.

III. Проект текста главы о технологии проведения переписи для Рекомендаций по проведению раунда переписей населения и жилищного фонда 2030 года

A. Введение

10. Технология используется для оказания поддержки на всех этапах переписей населения и жилищного фонда. В этой главе излагаются общие соображения по использованию и тестированию технологий для целей переписи и даются рекомендации по технологиям сбора и обработки данных. В ней также обсуждаются технологии использования данных из административных источников для целей переписи.

11. Хорошо зарекомендовавшие себя технологии, такие как системы ввода с клавиатуры, не рассматриваются подробно, поскольку эти системы уже хорошо изучены и документированы. Наибольшее внимание уделено технологическим решениям, получившим широкое распространение в практике проведения переписей раунда 2020 года — электронным инструментам сбора данных, облачным технологиям и комплексным решениям для автоматизации всех этапов переписных работ — от планирования переписи до распространения ее результатов. При этом понимается, что в годы, предшествующие раунду переписей 2030 года, могут открыться новые возможности для использования технологий, таких как машинное обучение и искусственный интеллект.

B. Общие соображения

1. Факторы, способствующие внедрению технологических инноваций

12. Технология обладает *потенциалом* улучшить охват переписи и качество данных, сократить расходы и быстрее распространять результаты. Однако в краткосрочной перспективе внедрение новых технологий может привести к увеличению расходов. Переписным органам необходимо рассмотреть вопрос о том, как новые возможности, предоставляемые технологическими инновациями, могут способствовать повышению релевантности, точности и своевременности данных переписи². Например, при снижении стоимости обработки данных можно расширить содержание программы переписи или увеличить долю выборки для выборочных вопросов, чтобы повысить качество данных — тематически — в отношении малых групп населения или — пространственно — в отношении малых географических районов и даже километровой прямоугольной сетки. Однако любое расширение содержания программы должно быть сбалансированным с учетом влияния на респондентов.

² Подробное обсуждение факторов, способствующих внедрению технологических инноваций при проведении переписи методом прямой регистрации, см. в разделе В.3 документа «Руководящие принципы использования технологий электронного сбора данных в переписях населения и жилищного фонда» Департамента Организации Объединенных Наций по экономическим и социальным вопросам (2019). Нью-Йорк: Организация Объединенных Наций.

13. Потребность в разработке политики и планировании на основе фактических данных порождает спрос на данные переписи населения со стороны все более широкого круга пользователей. Поэтому системы формирования выходных данных должны быть ориентированы на разных пользователей — начиная с тех, кому нужен быстрый доступ к базовым ведущим показателям, и заканчивая теми, кто рассчитывает провести углубленный анализ.

2. Оценка

14. *Вопрос о принятии на вооружение новых технологий или методологических подходов должен рассматриваться только в том случае, если имеется четкое понимание их преимуществ и если их внедрение может быть управляемым*³.

15. *Целесообразность внедрения любой технологии, ранее не опробованной в условиях переписи, должна быть тщательно оценена заранее, с учетом национального контекста, относительной стоимости по сравнению с традиционными решениями, работы, необходимой для разработки и тестирования, потенциальных последствий для общей организации переписных работ, потенциального влияния на качество результатов переписи и воздействия на население в целом.*

16. Переписные органы должны провести такую оценку заблаговременно до начала переписи, чтобы *определить, какие системы и процессы подходят* для их ситуации. Необходимо рассмотреть следующие вопросы:

a) относительные затраты на персонал и процессы ручной обработки по сравнению с затратами на возможные компьютерные системы и сопутствующую инфраструктуру;

b) технологический потенциал и инфраструктура как в национальном статистическом управлении (НСУ), так и в стране в целом;

c) способность НСУ управлять сложными и продвинутыми процессами разработки систем;

d) наличие финансирования и времени для разработки и тестирования технологического решения.

17. В процессе переписи следует учитывать принципы *дизайн-мышления и поведенческих наук*. Дизайн-мышление позволит критически осмыслить текущие процессы и эволюционировать, чтобы удовлетворить будущие потребности клиентов и граждан. Это должно привести к снижению количества отказов в предоставлении ответов, поскольку граждане получают ориентиры для заполнения переписного листа.

18. Прежде чем внедрять новые технологии или методологические подходы, НСУ должно четко понимать связанные с ними *риски*. Как будет работать новая технология? Какова будет реакция респондентов или переписного персонала? Из-за длительных интервалов между циклами переписей (от 5 до 10 лет) возможности для изучения новых подходов на собственном опыте могут быть ограничены.

19. При рассмотрении вопроса о внедрении инноваций НСУ настоятельно рекомендуется рассмотреть вопрос об *изучении опыта других переписных органов в международном масштабе*. Можно также рассмотреть возможность сотрудничества с другими организациями для совместного тестирования новых подходов или технологий до их внедрения.

20. Сложность нового программного обеспечения и инфраструктуры, необходимой для многих новых и появляющихся технологий, может выходить за рамки текущих технических возможностей НСУ. Поэтому вполне вероятно, что некоторые страны

³ Подробное обсуждение процесса принятия решений об использовании той или иной технологии см. в разделе В.4 документа «[Руководящие принципы использования технологий электронного сбора данных в переписях населения и жилищного фонда](#)» Департамента Организации Объединенных Наций по экономическим и социальным вопросам (2019). Нью-Йорк: Организация Объединенных Наций.

захотят рассмотреть возможность *передачи на внешний подряд* значительных компонентов технических решений, связанных с проведением переписи.

3. Тестирование

21. НСУ следует использовать сильную стратегию *тестирования* и рассматривать мероприятия по тестированию как приоритетные для переписи. Подлежащие тестированию аспекты включают в себя функциональную корректность, нефункциональные требования, интерфейсы, обработку ошибок, производительность, надежность, удобство использования и безопасность. Помимо цикла тестирования, внедряемые решения должны быть сертифицированы на предмет кибербезопасности.

22. *Рекомендуется включить в цикл тестирования следующие компоненты:*

- a) тесты, ориентированные на конкретные области или содержание:
 - качественные или когнитивные тесты;
 - тесты для новых онлайн-сервисов;
 - количественный тест содержания или поведенческий тест;
 - А/В-тестирование (сплит-тестирование);
- b) тесты, специфичные для информационных технологий (ИТ):
 - юнит-тесты;
 - тестирование в полевых условиях;
 - тестирование интерфейса;
 - тестирование систем и эксплуатационной готовности;
 - тестирование резервного копирования и восстановления данных;
 - тестирование защиты от атак вымогателей;
 - тестирование объема и производительности;
 - тестирование на стабильность нагрузки;
 - тестирование безопасности, например тесты на проникновение;
 - нефункциональное тестирование (проверка производительности, надежности, полезности и других нефункциональных аспектов программного приложения);
- c) функциональные тесты:
 - системно-интегрированное тестирование: множество итераций в течение цикла переписи;
 - интегрированные сквозные тесты;
 - приемочное тестирование.

23. Тестирование слаженности работы всей системы переписи особенно важно в тех случаях, когда респондентам предоставляются различные варианты заполнения переписных листов, например с помощью мобильного телефона, планшетного или настольного компьютера. Каждый из этих способов должен быть протестирован на доступность, удобство использования, готовность к работе, производительность и безопасность.

24. В условиях стремительного развития информационных технологий и связанных с этим значительных изменений во многих технологических элементах переписи после раунда 2020 года *обязательное тестирование всего технологического решения в рамках пробной переписи приобретает особое значение при подготовке к раунду переписей 2030 года.*

4. Управление проектом

25. Переписные работы включают в себя ряд административных процессов, общих для других *крупномасштабных проектов*. Например, при планировании такой сложной операции, как перепись, целесообразно использовать соответствующее программное обеспечение для планирования проектов. Кроме того, во многих странах могут потребоваться системы и процессы для найма и оплаты труда большого числа временно нанятых переписчиков переписи. НСУ следует рассмотреть вопрос о том, как технологии могут помочь повысить эффективность и результативность этих операций. Это может способствовать как сокращению расходов на проведение переписи, так и повышению общего качества переписи, позволяя сосредоточить ресурсы на выполнении основных задач по регистрации, обработке и распространению данных, а не на административных процессах, таких как оплата труда персонала.

С. Сбор данных методом прямой регистрации

26. Рассматриваются следующие три технологии сбора данных: вариант предоставления ответов через Интернет; опрос по телефону; и использование портативных устройств, таких как ноутбуки, планшетные компьютеры и смартфоны⁴.

1. Интернет

27. При прямой регистрации рекомендуется предлагать вариант предоставления ответов через Интернет в качестве первого или предпочтительного варианта.

28. Ответы на электронные вопросники снижают общую стоимость сбора данных и позволяют получить данные более высокого качества. Данные, собранные через Интернет, будут отформатированы, и респондент сможет самостоятельно их исправить и отредактировать. Для сравнения, сбор данных с помощью бумажных носителей требует множества этапов обработки, таких как сканирование, считывание данных и, что более важно, ввод данных оператором. Тот факт, что система должна интерпретировать различные почерки, усложняет процесс и снижает точность данных. По всем этим причинам вариант предоставления ответов через Интернет явно предпочтительнее.

29. Использование Интернета в качестве способа передачи информации означает, что данные собираются путем *самоисчисления*, а не опроса. Поскольку вариант предоставления ответов через Интернет может быть включен в любой из традиционных методов доставки и сбора переписных листов, существует большое число способов его использования, такие как раздача и сбор и рассылка с возвратом по почте.

30. Ключевым фактором, который необходимо учитывать, является управление операциями по *контролю* за сбором данных, т. е. обеспечение того, чтобы каждое домохозяйство и каждый человек были учтены один и только один раз. Для этого необходимо обеспечить привязку каждого домохозяйства и любого человека в нем к его географическому местоположению. Дополнительной сложностью для тех стран, в которых переписные листы также собираются переписчиками (а не рассылаются по почте), является обеспечение подходящей и своевременной обратной связи с переписчиками для обновления ими информации для контроля сбора, чтобы они не посещали домохозяйства, которые уже вернули переписные листы.

31. Потенциальный уровень использования интернет-варианта следует оценивать по доле населения, имеющего доступ к Интернету из дома, доле населения, пользующегося широкополосными услугами, или общему использованию Интернета

⁴ Для всестороннего ознакомления с этими и некоторыми другими технологиями следует обратиться к документу «Руководящие принципы использования технологий электронного сбора данных в переписях населения и жилищного фонда» Департамента Организации Объединенных Наций по экономическим и социальным вопросам (2019). Нью-Йорк: Организация Объединенных Наций.

для других целей, таких как банковские операции, подача налоговых деклараций или покупки.

32. Необходимо также разработать *системы и процессы*, позволяющие возвращать переписные листы через Интернет. Доказано, что они позволяют сократить расходы за счет снижения нагрузки на переписчиков, а также расходы по вводу данных, печатанию и почтовые расходы.

33. Безопасность данных — очень важный вопрос, который должен быть одним из ключевых соображений при разработке инфраструктуры. *Для сбора и обработки переписной информации должны быть созданы физически отдельные инфраструктуры. Заполненные индивидуальные переписные листы после их сбора и ввода данных должны быть перемещены в безопасную инфраструктуру обработки данных, отделенную от инфраструктуры сбора.*

34. Стандартный переписной лист, который можно скачать из Интернета, требует гораздо меньше инфраструктуры, чем бланк, заполняемый в режиме онлайн. Однако загружаемые бланки обычно требуют более высокого уровня компьютерной грамотности, чем онлайн-бланки. Они не обязательно будут работать на разных конфигурациях компьютеров, и будет существовать ожидание того, что НСУ сможет справиться с каждой отдельной проблемой. С точки зрения респондентов недавний опыт показывает, что они с гораздо большей вероятностью предпочтут заполнить бланк в режиме онлайн. *По этим причинам рекомендуется использовать бланки для заполнения в режиме онлайн.*

35. Принятие к использованию варианта ответа через Интернет требует предоставления респондентам *верительных данных* и способов доставки логинов и паролей, необходимых для доступа к онлайн-бланку, которые могут включать:

- a) рассылку бумажных бланков или писем по почте;
- b) доставку переписчиком непосредственно по адресу респондента;
- c) рассылку по электронной почте;
- d) рассылку по электронной почте;
- e) использование логинов онлайн-порталов государственных услуг или других онлайн-сервисов, требующих ввода персонального идентификационного номера.

36. *Онлайн-бланк* предоставляет возможность интерактивного редактирования для улучшения качества ответов, что невозможно при использовании бумажного бланка. Люди, пользующиеся электронными бланками, ожидают, что им будет предложено определенное руководство — как минимум, что они будут последовательно заполнять бланк и им не будут задавать вопросы, которые не относятся к их ситуации. Для обеспечения высокого качества данных, собранных через Интернет, важно предусмотреть механизмы контроля ошибок при заполнении бланка. Такой контроль должен осуществляться в режиме реального времени, и респондент должен иметь возможность немедленно изменить любые неверные данные. Если в ответах респондента обнаруживаются противоречия, онлайн-бланк должен сообщить об этом и предоставить респонденту возможность исправить один или несколько ответов, удалить их или подтвердить, что заявленная ситуация существует в реальной жизни, хотя это и не предусмотрено разработчиками бланка.

37. Предоставление варианта ответа через Интернет может способствовать повышению качества переписи, поскольку некоторым *группам населения, с трудом поддающимся переписи*, будет проще принять в ней участие. Большинство стран сообщают о трудностях при переписи, например, молодых взрослых и людей, проживающих в жилых единицах с повышенным уровнем безопасности, доступ в которые ограничен. Некоторым людям с инвалидностью может быть проще заполнить интернет-бланк, чем бумажный переписной лист. Кроме того, эти группы с большей вероятностью используют Интернет для других целей, поэтому, если есть возможность, этот вариант следует продвигать среди этих групп в качестве средства поощрения участия в переписи.

38. Обеспечение достаточной *инфраструктуры* является одной из основных вызовов, связанных с предоставлением варианта ответа через Интернет. Перепись проводится в течение относительно короткого периода времени и охватывает все население страны, и маловероятно, что НСУ уже располагает необходимой инфраструктурой, чтобы справиться с пиковыми нагрузками, связанными с переписью. Поэтому вполне вероятно, что, по крайней мере, этот компонент интернет-решения можно с полным основанием отдать на внешний подряд. Для ограничения числа обращений может потребоваться изменение процедур сбора. Например, разнесение во времени сроков доставки переписных листов или писем с приглашениями или требованием к лицам, находящимся за пределами заранее определенных целевых групп/районов, связаться с НСУ, прежде чем они смогут воспользоваться интернет-бланком, может быть одним из способов ограничения использования интернет-бланка.

39. Поэтому переписным органам следует оценить, как они хотят *продвигать* использование интернет-варианта. Продвижение интернет-варианта должно определяться возможностями сервиса по обработке ожидаемой нагрузки и должно быть скоординировано с другими процедурами сбора данных. Стратегия связей с общественностью должна включать в себя гарантии безопасности и конфиденциальности информации, предоставляемой через Интернет. Если предположить, что интернет-вариант предназначен для всего населения, то стратегия связей с общественностью должна также включать в себя управление ожиданиями общественности относительно возможности доступа к сайту в периоды пиковой нагрузки. Следует подготовить простые сообщения с так называемыми «вежливыми рекомендациями», советуя людям использовать Интернет в «непиковое» время, и при необходимости использовать их на самом интернет-сайте переписи, в любой телефонной справочной службе переписи, а также в любой рекламе в СМИ.

40. Можно ожидать, что использование варианта ответа через Интернет будет выше уровня, наблюдавшегося в ходе раунда переписей 2020 года. Во время сбора данных переписные органы должны постоянно следить за уровнями предоставления ответов населением и при необходимости *стараться повысить* уровень предоставления ответов в режиме онлайн.

2. Портативные устройства

41. Растущая сложность и снижение удельных расходов на связь с помощью портативных, планшетных компьютеров и смартфонов означает, что они могут стать *затратоэффективным решением* для некоторых аспектов сбора данных переписи. Возможные цели применения таких устройств — замена бумажных карт переписчиков, регистров и списков адресов, а также в качестве средства сбора данных на местах. Они могут применяться во всех методологиях сбора переписи, начиная с раздачи и заканчивая сбором переписных листов.

42. Преимущество портативных устройств заключается в том, что они могут предоставлять *двустороннюю управленческую информацию в режиме реального времени*. Руководители переписи могут получать информацию о ходе сбора данных на основе данных о доставке переписчиками переписных листов и сборе ими заполненных бланков. Кроме того, руководители переписи могут предоставлять переписчикам через портативное устройство информацию о полученных бланках и о том, какие домохозяйства необходимо посетить повторно. Кроме того, можно обмениваться геопространственной информацией для целей сбора (например, отсутствующие адреса, новые жилища), что позволит эффективно использовать ресурсы. Руководители переписи могут в режиме реального времени выявлять участки, где перепись отстает от графика или не соответствует стандартам качества, и принимать соответствующие меры.

43. Использование портативных устройств должно создать больше возможностей для повышения эффективности сбора данных. Однако при использовании таких устройств необходимо учитывать ряд технических вопросов:

a) размер экрана может влиять на способность переписчика точно записывать и проверять ответы. По этой же причине при работе с мобильными устройствами через Интернет существует риск фрагментации данных из-за малого размера экранов;

b) компактные и легкие устройства с достаточно большим объемом памяти наиболее удобны для полевой работы переписчиков. Яркость и контрастность экрана должны регулироваться, чтобы можно было пользоваться устройством как при ярком, так и при слабом освещении;

c) чтобы обеспечить сохранность данных, введенная информация должна храниться в устройствах как можно меньше по времени — предпочтительно не более 24 часов;

d) устройства должны быть способны работать в автономном режиме в течение определенных периодов времени. Вопрос о времени работы от батареи следует рассматривать в зависимости от ежедневной нагрузки на полевой персонал. Возможно, стоит приобрести дополнительный банк питания для устройства;

e) если обновление системы и программного обеспечения происходит на этапе сбора данных, необходимо исключить риски потери ранее собранных данных или их несоответствия данным, собранным после обновления;

f) в некоторых районах страны точность GPS (например, в плотной городской застройке) и прием мобильного сигнала (например, в горных или лесных районах) могут быть неудовлетворительными. Необходимо оценить возможности веб-подключения портативного устройства, особенно если портативное устройство используется для веб-сбора данных.

44. Решения, основанные на портативных устройствах, должны быть всесторонне протестированы до начала переписи, как индивидуально, так и во взаимодействии с другими элементами технологии переписи, в которых не используются портативные устройства.

45. Использование портативных устройств также связано с целым рядом проблем безопасности:

a) по сравнению с пачками бумажных переписных листов существует больший риск их кражи или потери. Однако регулярная загрузка данных с таких устройств должна свести к минимуму необходимость повторной регистрации участков в случае их утери;

b) необходимо принять меры для защиты конфиденциальности любых данных, находящихся на устройстве, в случае потери устройства или при передаче данных. Данные, хранящиеся на устройствах, должны быть зашифрованы и доступны только с помощью специальных мер защиты (например, паролей, отпечатков пальцев);

c) передача данных также должна быть защищена с помощью шифрования и использования безопасных каналов между начальным и конечным пунктом передачи;

d) программное обеспечение безопасности должно быть загружено на устройство и должно быть совместимо с другими приложениями на устройстве. Однако программное обеспечение безопасности и пароли добавляют дополнительный уровень сложности при использовании. Эти меры безопасности увеличат расходы на поддержку;

e) важным вопросом является обучение и техническая поддержка персонала, занимающегося переписью населения. Не следует полагать, что люди, которые будут наняты для выполнения функций переписчика, будут технически компетентны. Эти факторы становятся все более сложными и трудноуправляемыми по мере увеличения численности переписчиков и физических расстояний. В крупных странах переписчики могут рассчитывать на обучение и техническую поддержку, предоставляемую дистанционно через Интернет или по телефону.

46. Средства обучения использованию портативных устройств должны быть загружены в устройство для удобства их использования переписчиками в целях обучения и во время полевой работы, охватывать все элементы работы переписчика, быть интерактивными, иметь удобную навигацию, содержать наглядные примеры реагирования переписчика во всех возможных ситуациях использования этого устройства.

47. Переписным органам следует заранее подумать о том, как использовать большое количество устройств после переписи. Хранить устройства для следующей переписи нецелесообразно, поскольку за 5–10 лет без использования и подзарядки они могут технологически устареть и стать непригодными для использования. Переписные органы могут передать часть этих устройств другим пользователям (например, государственным организациям), а часть оставить себе.

3. Телефон

48. В прошлом автоматизированный опрос по телефону предлагался в качестве потенциально затратноэффективного решения для стран, в которых существует «короткий» переписной лист, требующий только ввода базовой демографической информации. Однако ни одна страна не применяла его в ходе раунда переписей 2020 года. *Автоматизированный опрос по телефону использовать не рекомендуется.*

49. Метод опроса по телефону с использованием компьютерных технологий (САТИ) может быть использован для сбора данных с помощью переписного листа и/или для проверки и восполнения недостающих данных, собранных с помощью длинного переписного листа. Однако странам следует помнить, что удобство использования таких систем значительно снижается по мере увеличения количества и сложности вопросов или роста числа лиц в домохозяйстве.

4. Разработка электронного вопросника

50. Разработка электронного вопросника — очень важная часть технологического решения, когда ответы предоставляются онлайн или собираются с помощью портативных устройств.

51. При разработке электронного вопросника должны учитываться следующие требования. Он должен:

a) содержать полный набор и четкую последовательность вопросов, которые делятся на открытые и закрытые;

b) предлагать варианты ответов с возможностью выбрать только один вариант или несколько вариантов ответов на закрытые вопросы, а при выборе варианта «другое» — возможность ввести собственный ответ респондента;

c) обеспечивать полное отображение всего вопроса и ответа на него на экране устройства без прокрутки или перехода к следующему экрану, если это возможно, потому что скрытая часть вопроса или вариантов ответа может быть пропущена при заполнении;

d) предоставлять опцию помощи по вопросам, например текстовую подсказку или переход к соответствующему элементу метаданных или учебным материалам;

e) обеспечивать удобную навигацию по вопросам для одного респондента, между членами одного домохозяйства и между различными разделами вопросника (например, о жилищных условиях, о домохозяйстве, о лице);

f) использовать встроенные средства контроля достоверности введенных данных, принимая во внимание ранее введенную информацию об этом лице и других членах этого домохозяйства;

g) отображать индикатор продвижения в заполнении вопросника, а также общие количественные характеристики заполненного вопросника, такие как количество лиц в домохозяйстве и полнота ответов в отношении каждого лица.

52. В случае использования электронного вопросника как для самостоятельного заполнения респондентом в режиме онлайн, так и на устройстве переписчика, дизайн вопросника может отличаться, поскольку респонденты не имеют представления о методологии переписи, тогда как переписчик предварительно обучен и ознакомлен с терминологией переписи и метаданными, встроенными в вопросник.

53. *Онлайн-бланк должен дополнительно содержать краткое описание основных требований к его заполнению респондентом. В их число входят:*

- a) описание общей структуры вопросника и последовательности его заполнения;
- b) расчетное время заполнения ответов на вопросы на одного респондента или одно домохозяйство;
- c) описание способов вызова справочной информации и реагирования на сообщения об ошибках;
- d) описание возможности исправить, удалить или добавить какую-либо информацию в ранее заполненный вопросник, если это необходимо;
- e) признаки успешного и неуспешного завершения процесса переписи и дальнейшие действия респондента (например, получение подтверждения участия в переписи);
- f) способ обратной связи с НСУ, например номер телефона, адрес электронной почты горячей линии переписи или НСУ, чтобы оценить качество онлайн-услуг или задать вопросы, на которые не удалось получить ответы при заполнении бланка;
- g) возможность при необходимости получить перевод бланка на самые широко используемые в стране языки;
- h) также может быть полезно предоставить ответы на часто задаваемые вопросы с использованием терминологии, доступной для респондентов, и дать гиперссылку на страницу сайта НСУ, где описаны правовые, методологические и организационные принципы переписи.

5. Технологии для поддержки переписи людей с нарушениями и людей, не пользующихся цифровыми технологиями

54. При внедрении новых технологий необходимо помнить, что переписью должно быть охвачено *все население*, независимо от используемого технического оборудования и уровня компьютерной грамотности респондента.

55. Технология может помочь в проведении переписи людей с нарушениями и людей, не пользующихся цифровыми технологиями, *двумя основными способами:*

- a) обеспечения охвата респондентов, не имеющих необходимого подключения к Интернету;
- b) обеспечения того, чтобы как можно больше людей могли предоставить свои ответы в электронном режиме.

56. *Люди, не пользующиеся цифровыми технологиями.* Чтобы охватить как можно больше людей, придется подумать о местах, где нет надлежащего доступа к Интернету. Один из аспектов, о котором мы должны подумать в таких случаях, — это, к сожалению, бумажный вариант. Однако следует рассмотреть и другие варианты, например, чтобы люди, имеющие доступ к спутниковому Интернету, могли собирать информацию с помощью портативных устройств. Кроме того, дайте респонденту номер телефона, по которому можно позвонить, чтобы он мог заполнить свой бланк переписного листа с помощью оператора.

57. *Доступный вариант предоставления ответов через Интернет.* Еще один аспект, который мы должны рассмотреть, — это то, как люди с ограниченными возможностями смогут заполнить переписной лист через Интернет. Вариант предоставления ответов через Интернет должны соответствовать стандартам доступности, определенным в «Руководство по обеспечению доступности веб-контента» (WCAG) 2.2. Придерживаясь этих стандартов в техническом отношении,

важно учитывать принцип доступности уже на этапе разработки содержания. Перед разработкой любых новых функций, которые могут повлиять на доступность, необходимо провести консультации с экспертным центром по вопросам доступности и группой пользователей, чтобы убедиться, что новые функции разработаны таким образом, чтобы они надлежащим образом доступны. В следующих пунктах описаны примеры того, что можно сделать.

58. *Скрытый текст для автоматически генерируемых по маске символов полей.* В их число входят многочисленные типы масок ввода, которые помогают обеспечить соответствие содержимого ожидаемым форматам, а также автоматическое генерирование разделительных символов, таких как дефисы в поле даты. Скрытый текст инструкций, видимый только экранным дикторам, информирует пользователей нарушениями зрения о том, что необходимо ввести пользователю, а что будет сделано автоматически с помощью JavaScript.

59. *Требуемый цветовой контраст для просмотра текстового содержимого.* Для того чтобы пользователи могли просматривать текстовое содержимое, все тексты анализируются на предмет того, что шестнадцатеричные цветовые различия в RGB-формате достаточны для соответствия стандартам WCAG 2.0 уровня AA по коэффициенту контрастности не менее 4,5:1 для обычного текста и 3:1 для крупного текста (14 пунктов (обычно 18,6брх) и жирного или крупнее, или 18 пунктов (обычно 24рх) или крупнее).

60. *Метки, табуляторные ограничители и ссылки aria-described by usage, для обеспечения контекста иерархии полей.* Поскольку некоторые макеты вопросов в обследованиях могут содержать несколько слоев текста, следование стандартам WCAG для значимых последовательностей привело к автоматическому созданию табуляторных ограничителей (т. е. синих прямоугольников), которые соответствуют странице в порядке чтения (слева направо, сверху вниз). Соблюдение требования WCAG к Меткам или Инструкциям привело к созданию текста меток для всех вводимых данных (например, зависимой клавиши «Да» и полей для галочки⁵), который считывается экранными дикторами, когда фокус наводится на сам вводимый текст. Наконец, чтобы соответствовать критерию успеха WCAG «Информация и связи», ARIA-атрибуты могут быть созданы на входных данных, где, например, при фокусировке на флажке в зеленом прямоугольнике ниже, приложение будет читать текст последовательно. Это обеспечивает контекст, который поможет пользователю решить, стоит ли ставить галочку в поле.

6. Программное обеспечение переписи

61. На базовом уровне мультимодальные операции по сбору данных требуют *своевременного* предоставления информации переписчикам, чтобы они не посещали домохозяйства, которые уже представили переписной лист. Это и вопрос эффективности, и вопрос связей с общественностью. Современные технологии позволяют улучшить управление полевыми операциями, а значит, и качество самой переписи.

62. Хотя ключевым вопросом является своевременное поступление информации к переписчику, эти же системы должны обеспечивать *двусторонний обмен информацией между руководителями переписи и переписчиками в режиме, близком к реальному времени.* Такой мониторинг работы переписчиков позволит более своевременно принимать меры в тех случаях, когда процесс сбора данных отстает от графика или возникают проблемы с качеством собранных данных.

63. Возможно, НСУ придется прибегнуть к услугам сторонних организаций для разработки ключевых компонентов решения. Независимо от того, являются ли эти системы внутренними или внешними, они должны соответствовать международным

⁵ W3C 2023. [Руководство по обеспечению доступности веб-контента \(WCAG\) 2.2](#). Рекомендация W3C от 5 октября 2023 года.

(например, семейство ISO27000), региональным (например, NIS2⁶ в Европейском союзе) или национальным *стандартам кибербезопасности*. Это особенно важно, поскольку программное обеспечение для различных внешних и внутренних операций (сбор, обработка, географические информационные системы, импутация, распространение) предоставляется различными решениями (например, универсальными, собственными, индивидуальными разработками).

64. Интегрированная система полевой связи может использовать уже существующую ИТ-инфраструктуру и опираться на нее. Это подразумевает, например, прямой доступ к национальной инфраструктуре геокодирования, предоставляющей данные об индивидуальных адресах. *Геопространственная информация (адреса, здания, кадастровые участки) должна опираться на использование постоянных идентификаторов (PID)*.

65. *Инфраструктура управления переписью должна содержать следующие элементы:*

a) регистр адресов жилых помещений с геопространственными координатами, в котором все адреса распределены по переписным участкам;

b) регистр переписчиков и их контактная информация с возможностью привязки к каждому из них определенного переписного участка (из регистра жилых помещений) и в том числе адресов — если используется данный метод сбора данных;

c) регистр устройств для сбора данных (например, планшетных компьютеров или смартфонов) и их уникальных серийных номеров с возможностью их привязки к переписчикам (из регистра переписчиков) — если используется данный метод сбора данных;

d) центральное хранилище данных переписи, используемое для сбора, обработки и накопления всех данных, связанных с адресом проживания респондентов (из регистра жилых помещений), полученных всеми методами сбора данных, используемыми в переписи (ответы в режиме онлайн, данные, полученные от переписчиков, данные из административных источников), при этом метод сбора данных должен быть указан в хранилище;

e) коммуникационное программное обеспечение, обеспечивающее своевременный обмен информацией между переписчиками, инструкторами и группой управления переписью.

66. Все элементы этой инфраструктуры должны быть взаимосвязаны и управляться централизованно с помощью программных и технологических средств, специально разработанных для целей переписи. *Рекомендуется осуществить установку, настройку и проверку работоспособности устройств в центре управления переписью перед передачей устройств персоналу на местах.*

67. Централизованное управление портативными устройствами для сбора данных переписи включает в себя *автоматизацию следующих функций* (если используется этот метод сбора данных):

a) установка на устройства следующего: специального программного обеспечения для сбора данных переписи переписчиком (электронный вопросник); список адресов жилых помещений и карты переписного участка; метаданные и классификаторы, используемые в электронном вопроснике; инструменты для контроля работы устройства/переписчика; учебные материалы для переписчиков/пользователей устройств;

b) однозначная связь между устройством, переписчиком и переписным участком в регистрах жилищ, переписчиков и устройств, обновление регистров и связей между их элементами в случае замены переписчиков или устройств;

⁶ Директива (ЕС) 2022/2555 Европейского парламента и Совета от 14 декабря 2022 года [о мерах по обеспечению высокого общего уровня кибербезопасности в странах Союза](#).

с) онлайн-управление и мониторинг работы каждого устройства после его инициализации персоналом на местах;

d) удаление с устройства всей информации и программного обеспечения, использовавшегося для целей переписи, после успешной передачи собранных данных в центральное хранилище данных по окончании переписи и подготовка устройств к длительному хранению (консервации) или использованию в других целях (например, передача другому агентству).

68. *Онлайн-управление и мониторинг работы устройств* включают в себя:

a) получение информации о времени включения и выключения устройства;

b) получение геокоординат для заполнения электронного вопросника по каждому адресу и статистика заполнения, такая как продолжительность заполнения и количество заполненных вопросов, ошибки и исправления в вопросниках;

с) передачу на устройства информации об адресах онлайн-ответов и контрольных данных по этим респондентам, чтобы переписчик мог проверить полноту переписи по этому адресу или при необходимости внести исправления в данные в соответствии с запросом респондента;

d) удаленную установку обновлений программного обеспечения на устройства в экстренных случаях. Рекомендуется использовать этот способ только в случае обнаружения критических ошибок в программном обеспечении в ходе переписи, поскольку важно обеспечить согласованность данных, собранных с использованием различных версий программного обеспечения;

e) блокировку устройства в случае, если сотрудник на месте заявляет о потере или краже устройства, чтобы предотвратить незаконное использование устройства или утечку информации с него;

f) обеспечение возможности дистанционного консультирования переписчика с центральным бюро переписи.

69. *Средства управления центральным хранилищем данных переписи* обеспечивают следующие функции:

a) загрузка бланков, заполненных респондентами в режиме онлайн (если используется этот метод сбора данных);

b) проверка их пригодности для обработки (наличие ответов, необходимых для идентификации респондента);

с) привязка заполненных бланков к адресам жилых помещений;

d) создание подтверждения, например в виде QR-кода, успешного участия в переписи для обратной связи с респондентом (для направления в зависимости от способа получения заполненного бланка в личный кабинет респондента онлайн-переписи, по электронной почте или другим способом) и для передачи переписчику — если используется этот метод сбора данных;

e) загрузка заполненных электронных вопросников с устройств переписчиков, из административных источников и других каналов сбора данных, используемых в переписи, с привязкой к адресам жилых помещений;

f) объединение версий заполненных переписных листов, полученных по всем каналам сбора данных, используемым в переписи, относящихся к одному и тому же адресу жилого помещения — выбор эталонной версии, обогащение ее данными из других версий, удаление дубликатов при необходимости;

g) подсчет статистики хода переписи и визуализация мониторинга работы для выявления проблемных переписных участков, на которые необходимо воздействовать, чтобы исправить ситуацию в ходе переписи.

70. *Поведенческие науки* также могут помочь в понимании мотивов, ощущений и опыта респондента. Примеры мер по увеличению доли предоставивших ответы респондентов включают своевременные меры по исправлению положения, улучшение

коммуникации и снижение когнитивного уровня, необходимого для возврата заполненного бланка в режиме онлайн или бумажном виде.

71. Еще одним ключевым фактором снижения риска являются отношения с ключевыми *технологическими партнерами*. Для успешной разработки, внедрения и реализации систем переписи решающее значение имеет эффективное управление, в котором центральное место занимает внутренняя группа управления переписью.

72. Сегодня доступны технологические решения, которые позволяют *объединить множественные функции управления на местах*. Управление взаимоотношениями с клиентами (ОВК) может предоставить решения для управления запросами на местах, внешнего веб-сайта, улучшения коммуникации с помощью функции чат-бота, а также службы помощи гражданам для управления системой данных.

73. НСУ могут рассмотреть возможность использования *технологии удаленного доступа* для обеспечения гибкого графика работы сотрудников, обрабатывающих данные переписи, обеспечивая при этом безопасность и конфиденциальность данных.

D. Обработка данных

74. В последние годы произошли значительные *усовершенствования в технологиях обработки данных*, которые позволили снизить стоимость обработки данных переписи и повысить качество данных.

75. Основываясь на выводах, сделанных по результатам анализа практики стран в ходе раунда переписей 2020 года, можно предположить, что большинство стран, которые будут использовать прямую регистрацию в ходе следующей переписи, будут использовать *вариант предоставления ответов через Интернет*. Можно предположить, что использование бумажных бланков и технологии OCR/OMR будет ограниченным.

76. Потребуется всестороннее планирование *интеграции между ключевыми системами переписи*, чтобы обеспечить более бесперебойное функционирование по мере того, как предоставленные ответы на вопросы переписи будут поступать с этапа сбора на этап обработки. Привитие культуры эффективного управления различным партнерам, которые могут быть задействованы, также будет иметь решающее значение для успешной интеграции между полевыми системами, системами обработки и распространения информации.

1. Обработка бумажных вопросников

77. Сбор данных с помощью бумажных вопросников может потребоваться по причине предпочтений респондентов или отсутствия доступа к Интернету. *Для обработки бумажных вопросников рекомендуется использовать автоматизированные процессы, такие как интеллектуальное распознавание символов (ICR)*.

78. Оптическое распознавание меток (OMR) может быть затратоэффективным вариантом, если бланк переписи содержит только вопросы с *помечаемыми вариантами ответов*. Для обработки ответов в рукописном виде требуются дополнительные средства ввода данных или компьютерного кодирования. Однако OMR в значительной степени вытеснены технологиями ICR.

79. Наиболее затратоэффективным вариантом, скорее всего, будет *сочетание методов цифровой визуализации, ICR, корректировки и автоматизированного кодирования*. Пример этого процесса кратко описан ниже.

а) Бланки переписи обрабатываются сканером, в результате чего получается изображение. Программное обеспечение для распознавания используется для идентификации ответов в полях с помечаемыми вариантами ответов и перевода рукописных ответов в текстовые значения. Уровни доверия устанавливаются для определения того, какие ответы имеют приемлемое качество, а какие требуют дальнейшей корректировки или проверки;

b) автоматизированная корректировка призвана уменьшить необходимость вмешательства оператора и обычно включает в себя использование словарных справочных таблиц и контекстное редактирование. Словари подбираются в зависимости от обрабатываемого вопроса переписи. Так, например, словарь для вопроса о стране рождения будет содержать только названия стран. Подготовительная работа по созданию словарей терминов на естественном языке значительно повысит эффективность кодирования;

c) оператор может откорректировать нераспознанные изображения. Это является эффективным с точки зрения затрат только для тех вопросов, в случае которых высока вероятность того, что скорректированные данные могут быть автоматически закодированы;

d) автоматическое кодирование использует автоматизированные алгоритмы для сопоставления введенных ответов с индексами. Ответы, которые не поддаются увязке, затем передаются в процесс компьютерного кодирования. Для ответов, которые не могут быть автоматически закодированы, рекомендуется использовать алгоритм машинного обучения, который может заменить человека-кодировщика с таким же или даже лучшим качеством данных и значительным снижением затрат. Для обучения алгоритма машинного обучения можно использовать данные предыдущей переписи. Можно также использовать данные из текущего цикла пробной переписи, особенно если необходимо закодировать новые переменные. Чтобы еще больше ограничить расходы и повысить качество, ответы, которые не были закодированы, следует проанализировать на предмет общих ответов. Эти ответы могут быть добавлены в индексы кодирования и повторно представлены с помощью автоматизированного кодирования, либо может быть проведена другая форма массового кодирования.

80. Дальнейшие соображения по использованию цифровых изображений, ICR, корректировки, автоматизированного кодирования, оптического распознавания знаков (OMR) и оптического распознавания символов (OCR) представлены в *Рекомендациях КЕС для переписей раунда 2020 года*⁷. Можно ожидать, что генеративный искусственный интеллект приведет к появлению новых возможностей и заменит кодировщиков. Однако это требует инвестиций для построения надлежащих моделей, чтобы поддерживать высокое качество данных.

2. Формирование выходных материалов

81. Традиционно выходные материалы переписи включают в себя агрегированные таблицы, статистические данные, иллюстративные материалы, карты и т. д. с соответствующими метаданными. (См. [перекрестная ссылка на соответствующую главу].)

82. *Распространение в режиме онлайн* через Интернет дополнительно позволяет разрабатывать продукты, лучше отвечающие потребностям различных типов пользователей данных переписи, затратноэффективно распространять гораздо более широкий спектр данных переписи и повышать удобство использования данных. Интерфейсы программирования приложений (API) должны быть доступны для повышения удобства использования предоставляемых данных.

83. Функции и содержание данных могут определяться с учетом потребностей *пользователей различного уровня*. Эти функции должны быть эффективно интегрированы, начиная с простых и кончая самыми сложными, с тем чтобы служить ориентиром для пользователя в зависимости от характера запроса и анализа, который он желает осуществить с использованием различных продуктов.

84. Одной из главных целей переписи является формирование информации о *малых географических районах и о малых группах населения* (одновременно с социальной и

⁷ ЕЭК ООН, 2015. [Рекомендации Конференции европейских статистиков по проведению переписей населения и жилищного фонда 2020 года](#). Женева: Организация Объединенных Наций.

экономической точек зрения). Распространение через Интернет может быть ориентировано на поддержку этих двух видов использования данных. Что касается малых географических районов, то технология ГИС может использоваться одновременно в качестве средства определения представляющих интерес территорий в целях поиска данных и средства формирования картографических изображений на основе результатов поиска. Существуют разнообразные пакеты программного обеспечения, которые могут использоваться для выборки совокупностей из крупных заранее определенных матричных таблиц.

85. Система распространения информации через Интернет должна обеспечивать *гибкость*, позволяя пользователям экспортировать результаты в ряд общедоступных пакетов статистического анализа, составления таблиц или картографирования.

3. Интерактивные онлайн-платформы для распространения

86. По мере увеличения объема данных, собираемых на пространственном, временном или семантическом уровнях (например, европейская километровая прямоугольная сетка), стали появляться *продукты визуализации данных*, помогающие людям, владельцам предприятий, ученым и руководству всех уровней понимать ключевую информацию, извлекаемую из этих данных. Визуализация данных может использоваться для оперативного информирования, упрощения представления больших объемов данных, выявления закономерностей и взаимосвязей, а также для отслеживания изменений в переменных во времени⁸.

87. *Рекомендуется распространять данные на интерактивных онлайн-платформах, где различные группы пользователей могли бы непосредственно создавать таблицы и визуализации (например, графики, карты) в соответствии со своими потребностями.*

88. Основное внимание должно быть уделено применению подготовленных на основании подхода FAIR (находимость, доступность, совместимость и повторное использование) Руководящих принципов по управлению научными данными и их рациональному использованию⁹. Эти принципы подчеркивают способность вычислительных систем находить, получать доступ, взаимодействовать и повторно использовать данные без вмешательства или с минимальным вмешательством человека. Аналитические решения на платформах могут оставаться ограниченными, поскольку это не является основной целью переписи.

89. Пользователи данных переписи имеют очень разный уровень статистической грамотности или владения компьютером. *Рекомендуется учитывать следующие группы пользователей:*

a) непрофессиональные пользователи, которые в основном просматривают визуализации данных или основные моменты отчетов;

b) промежуточные пользователи, которые настраивают визуализации или выбирают конкретные данные;

c) технически продвинутые пользователи, которые будут загружать данные для проведения собственного анализа и использовать интерфейсы программирования приложений (API).

90. *Международные стандарты*, такие как SDMX, должны считаться приоритетными для форматирования выходных данных и метаданных, которые предоставляются пользователям.

91. Независимо от средств доступа или распространения, *защита статистической конфиденциальности* данных переписи является первоочередной задачей в любых таких системах, и необходимо предусмотреть процедуры статистического контроля для предотвращения идентификации.

⁸ Например, [Eurostat Census Hub](#).

⁹ [Инициатива GO FAIR](#).

4. Хранение и архивирование

92. Технологии хранения данных развиваются в соответствии с новыми технологическими тенденциями. Исторически сложилось так, что данные хранились в локальной среде, на традиционных ленточных носителях, иногда дополняемых дисковыми накопителями. Однако, поскольку объем данных продолжает расти в геометрической прогрессии, возможности решений для хранения данных пришлось адаптировать. Многие организации сейчас рассматривают возможность внедрения *гибридного подхода к хранению данных*, используя сочетание облачных и локальных решений для хранения данных в зависимости от индивидуальных потребностей.

93. При определении того, какая политика хранения подходит для отдельных статистических организаций, необходимо учитывать определенные факторы. В их число входят:

a) риск использования дополнительных решений для хранения данных, которые не являются локальными;

b) для разработки и обслуживания передовых решений в области хранения данных, таких как гибридный подход, потребуются новые навыки. Таких специалистов трудно получить и дорого удержать;

c) каждая статистическая организация должна убедиться в том, что она понимает свои данные с точки зрения их классификации. Это поможет в принятии решений, необходимых для создания возможностей для проектирования хранилищ;

d) по мере того как организации рассматривают новые решения для хранения данных, возникает необходимость в обеспечении кибербезопасности для выявления, обнаружения киберрисков и угроз, защиты от них, реагирования на них и восстановления после их возникновения. Стратегия киберзащиты должна обеспечивать защиту хранилищ в гибридной среде;

e) существует множество облачных систем хранения данных, в том числе ледниковое (Glacier) хранилище для длительного хранения данных, которые не используются часто. Организациям, которые исторически полагались на резервное копирование всех данных с использованием полного, инкрементного и/или дифференциального методов, потребуется более глубокое понимание своих данных, чтобы использовать возможности этих новых подходов к хранению;

f) сотрудничая с внешними поставщиками услуг хранения данных, организациям необходимо обеспечить заключение соответствующих соглашений об уровне обслуживания, которые будут соответствовать их ожиданиям в отношении предоставления услуг;

g) хранение и архивирование должны осуществляться в соответствии с правовыми документами и устоявшимися правилами. Например, в Европейском союзе любые данные научных статей должны храниться в течение 10 лет¹⁰;

h) процедуры резервного копирования и восстановления должны быть реализованы и протестированы для выполнения любых действий, необходимых для перезапуска систем и сервисов, проверки их работы и целостности данных;

i) план обеспечения бесперебойного функционирования рекомендуется разрабатывать и тестировать для предотвращения сбоев и восстановления систем и связанных с ними данных в случае таких угроз, как стихийные бедствия или кибератаки.

5. Облачные технологии

94. Облачные вычисления развиваются очень быстро, и ожидается, что они станут *самой распространенной ИТ-инфраструктурой в компаниях по всему миру*.

¹⁰ Решение Комиссии (ЕС) 2021/2121 от 6 июля 2020 года [об управлении записями и архивами](#).

Стратегии внедрения облачных технологий варьируются от стратегий «только облако», «облако в первую очередь» до, в последнее время, «умное облако»¹¹.

95. Переход на облачные технологии становится все более заметным и в случае государственных организаций по всему миру, поскольку они признают преобразующий потенциал облачных вычислений в повышении операционной эффективности, масштабируемости и сфере предоставления услуг. Использование облачных технологий позволяет им оптимизировать использование ресурсов, повысить доступность данных и стимулировать инновации в различных областях.

96. Развитие облачных технологий включает в себя множество платформ и моделей обслуживания. Модели обслуживания включают инфраструктуру как услугу (IAAS), платформу как услугу (PAAS) и программное обеспечение как услугу (SAAS). Платформы включают в себя государственные, частные и общественные облака. Каждая из этих отдельных моделей может предоставлять статистическим организациям возможности и создавать проблемы при их внедрении и развертывании.

97. Необходимо проанализировать многие соображения, связанные с внедрением облачных технологий статистическими организациями. Например, глобальные поставщики облачных услуг (гиперскейлеры) опираются на глобальные контракты, которые могут не соответствовать местным законам и законодательству. Одна из областей, требующих рассмотрения, включает юрисдикцию в случае возникновения спора. Для этого потребуются локальные соглашения, согласованные на национальном уровне. Это поможет снизить значительный барьер на пути внедрения облачных технологий. Суверенитет данных также является важным соображением и требует согласования.

98. В Руководстве ЕЭК ООН по внедрению облачных технологий¹² 2024 года рассматриваются многие ключевые темы, возможности и проблемы, связанные с облачными технологиями. Они могут помочь статистическим организациям глубже понять актуальные для них вопросы.

99. Для многих статистических организаций может подойти промежуточный подход к внедрению облачных технологий с постепенными приращениями. Если начать с менее чувствительных приложений и систем и изучить их опыт, это может обеспечить большую степень комфорта для ключевых заинтересованных сторон, принимающих решение о переходе на облачные технологии. Обеспечение оценки и измерения риска на постоянной основе также поможет в переходе.

100. Облачная инфраструктура может дать значительные преимущества, если перепись проводится с циклом от 5 до 10 лет. Преимущество использования облачных технологий заключается в возможности легко увеличивать и уменьшать масштаб при необходимости. Это может привести к экономии средств на программу переписи, но, тем не менее, требует осторожности и рассмотрения более широких последствий для организации. Например, локальная инфраструктура, созданная для проведения переписи, может быть использована для других целей организации. Так, с точки зрения программы переписи существуют возможности экономии, но их следует оценивать на более высоком уровне. Для более короткого цикла, такого как ежегодная перепись, такой подход может оказаться не самым лучшим.

101. Дополнительные преимущества связаны с инвестициями в новые облачные технологии, такие как Cloud native и Paas. Кроме того, следует рассмотреть возможность использования многооблачной архитектуры, чтобы воспользоваться преимуществами различных провайдеров облачных услуг. Например, может быть одна облачная инфраструктура для приложения, а другая — для базы данных.

102. Облачная инфраструктура обеспечивает большую гибкость, но ею необходимо строго управлять, чтобы не платить за инфраструктуру, которая не используется. В противном случае использование облака может оказаться более дорогостоящим

¹¹ ЕЭК ООН, 2024. «Облачные технологии для официальной статистики». Женева: Организация Объединенных Наций.

¹² Там же.

вариантом. Облачные серверы следует уменьшать по масштабу, когда они не используются, и увеличивать, когда это необходимо.

103. Процесс миграции в облако должен быть хорошо проанализирован и тщательно спланирован. Просто перенос в облако для использования другого центра обработки данных не является рекомендуемым подходом. В некоторых странах было доказано, что такой подход обходится дороже, чем оставаться в локальной среде. Страны должны уделять внимание тому, чтобы стать «облачными умниками», т. е. перейти в облако, перестроив приложение в облаке или используя аспекты технологии Cloud native (контейнеризация, Paas, непрерывная интеграция и непрерывное развертывание).

104. *Использование облачных технологий может дать важные преимущества при проведении переписи, но не всегда является оптимальным подходом. Перед переходом на облачные технологии рекомендуется провести тщательную оценку.*

Е. Технологии использования административных данных

1. Сфера охвата

105. Технологии, применяемые для использования административных данных, отличаются от технологий сбора данных на местах. Развитие *новых информационно-коммуникационных технологий* (ИКТ) и повышение их доступности позволяют более широкое использование административных регистров в целях переписей населения и жилищного фонда. Учитывая развитие современных технологий и стремление организаций внедрять инновационные решения в переписи раунда 2030 года, необходимо будет создать или модернизировать программно-аппаратную инфраструктуру для сбора, хранения и увязки данных из административных источников и хранения метаданных о процессах и продуктах.

106. Качество исходных данных оказывает большое влияние на качество выходных продуктов. По этой причине существенное значение имеет методология *повышения качества данных из административных источников*, например путем их корректировки в соответствии со статистическими требованиями. Современные ИКТ могут оказаться крайне полезными и сыграть ключевую роль в повышении эффективности и результативности этих операций. Для оценки качества административных источников, используемых в переписях, следует обратиться к Руководству ЕЭК ООН по этому вопросу, опубликованному в 2021 году¹³.

107. В рамках работ по подготовке переписи, особенно процесса планирования, необходимые *технические требования*, связанные с использованием данных из административных источников, которые могут повлиять на необходимость модернизации инфраструктуры, должны быть специфицированы в следующих областях:

- a) сбор данных;
- b) хранение данных;
- c) связывание данных;
- d) хранение метаданных или информации о процессах и продуктах.

108. Применение нескольких способов сбора данных из административных регистров и данных из других источников для использования в целях переписей населения и жилищного фонда, потребует более комплексных *процессов организации и управления*, а также более сложных систем. Современные технологии и в этом случае открывают возможности для совершенствования. Процесс сбора данных из административных регистров должен включать в себя подготовку стратегии сбора данных с использованием множественных способов.

¹³ ЕЭК ООН, 2021. [Руководство по оценке качества административных источников для использования в целях переписи](#). Женева: Организация Объединенных Наций.

2. Безопасность и конфиденциальность

109. Очень важно учитывать, что в обществе *все больше внимания* уделяется безопасности, конфиденциальности и защите данных. Это видно из законодательной и статистической базы многих стран. Кроме того, страны — члены ЕС должны соблюдать Общий регламент по защите данных (GDPR)¹⁴. В связи с этим повышаются требования к тому, как органы, проводящие перепись населения и жилищного фонда на основе регистра, должны получать и обрабатывать данные. Для удовлетворения этих требований рекомендуется уделять особое внимание принципу «нужной информации» и обработке обезличенных данных.

110. Безопасная ИТ-инфраструктура является необходимым условием для сбора данных из административных регистров. Одним из критически важных вопросов, связанных со сбором данных из административных источников, является защита данных. Независимо от применяемой технологии стратегия сбора данных должна обеспечивать *информационную безопасность*. Данное требование следует учитывать уже на ранних этапах планирования процесса получения и сбора данных из административных регистров и проектирования надлежащей программной и аппаратной инфраструктуры. Следует подробно рассмотреть технические вопросы, касающиеся шифрования передаваемых данных наряду с использованием безопасных каналов передачи.

111. Необходимо обеспечить безопасность от потери данных, применяя *соответствующие технические и организационные меры* для защиты данных от случайного или незаконного уничтожения или случайной потери (включая резервное копирование), изменения, несанкционированного раскрытия или доступа.

112. Безопасность должна быть реализована на *нескольких уровнях*. Владельцы регистров, НСУ, собирающие данные, и другие партнеры должны создать защищенные каналы передачи данных с соответствующими сертификатами для контроля доступа, предоставляемого сборщикам.

113. Организация данных по различным «состояниям» — таким как исходные данные, подготовленные данные, статистические данные и выходные данные — позволяет эффективно реализовать *управление авторизацией* для различных ролей и групп в НСУ. Доступ к данным должен осуществляться по принципу «нужной информации», с использованием анонимизированных идентификаторов и средств предотвращения идентификации на всех уровнях агрегирования.

3. Связывание, передача и хранение данных

114. Новые технологии могут оказаться полезными и для процесса связывания записей и данных. После определения административных источников, которые будут использоваться в переписи, необходимо будет составить их карту и создать интерфейсы программирования приложений (API) для *автоматической передачи данных* в центральную систему.

115. Административные источники отличаются друг от друга как в рамках различных предметных областей внутри страны, так и между странами. По возможности рекомендуется использовать стандартизированные API. Независимо от используемого интерфейса или встречающихся различий, *хорошие метаданные и хорошее понимание данных* очень важны для их включения в перепись.

116. В настоящее время многие страны переходят на облачные системы хранения данных и вычислений. По сравнению с традиционными локальными решениями этот переход открывает новые возможности, в том числе потенциал для создания более гибкой, управляемой переменными вселенной данных, или *озер данных*, а не традиционно хранимых, структурированных по темам и обновляемых данных. Эти озера данных будут предъявлять высокие требования к метаданным и точной периодизации, что станет возможным благодаря увеличению вычислительной

¹⁴ Регламент (ЕС) 2016/679 Европейского парламента и Совета от 27 апреля 2016 года [о защите физических лиц при обработке персональных данных и о свободном обращении таких данных](#).

мощности облачных вычислений. Централизация административных данных в озере данных поддержит бизнес-процессы и обеспечит мощную и надежную инфраструктуру для хранения, создания метаданных и синхронизации различных источников данных.

4. Улучшение административных данных

117. Для преобразования административных данных в статистические можно использовать различные методы. Благодаря процедуре *автоматической очистки данных* можно устранить ошибки в исходных данных из административных регистров и эффективно редактировать данные.

118. *Модели машинного обучения* могут быть использованы, например, для определения «переписного адреса» лица или типа частного домохозяйства. Возможно построение глубокой нейронной сети (DNN) на основе обучающей модельной системы искусственных нейронных сетей (ANN), которая обучает модель.

119. *Искусственный интеллект* (ИИ) может быть использован для определения классификации занятости экономической отрасли и профессионального занятия лица. Для разработки моделей нейро-лингвистического программирования (НЛП) и DNN можно создать ИТ-группу, в которую войдут инженеры по данным и специалисты по ИИ. Эти модели классификации НЛП нуждаются в постоянном обновлении и качественном анализе для оценки результатов. Классификация с использованием ИИ может быть более эффективной, давать качественные результаты и экономить время: примерно 90 % случаев классифицируются таким образом. При этом понимается, что специфические требования к безопасности и конфиденциальности не всегда позволяют работать с такими данными и моделями в «облаке». Также известны ограничения, связанные с использованием моделей классификации, разработанных на английском языке, в контексте других языков.

IV. Заключение

120. Настоящий проект рекомендаций по технологии переписи для раунда переписей и жилищного фонда 2030 года представляется для комментариев и обсуждения.
