

ЕЭК ООН

Руководящие принципы использования регистров и административных данных в целях переписей населения и жилищного фонда



**ОРГАНИЗАЦИЯ
ОБЪЕДИНЕННЫХ
НАЦИЙ**

ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ

Руководящие принципы использования регистров и административных данных в целях переписей населения и жилищного фонда



ОРГАНИЗАЦИЯ ОБЪЕДИНЕННЫХ НАЦИЙ

Нью-Йорк и Женева, 2018 год

Примечание

Употребляемые обозначения и изложение материала в настоящем издании не означают выражения со стороны Секретариата Организации Объединенных Наций какого бы то ни было мнения относительно правового статуса любой страны, территории, города или региона, или их властей, или относительно делимитации их границ.

ECE/CES/STAT/2018/4

ПУБЛИКАЦИЯ ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ
--

e-ISBN: 978-92-1-047485-6

Авторское право © Организация Объединенных Наций, 2018 год
Все права защищены во всем мире
Публикация Организации Объединенных Наций, выпущенная Европейской экономической комиссией
Организации Объединенных Наций

Предисловие

Основная цель настоящей публикации заключается в предоставлении организаторам переписей населения и жилищного фонда рекомендаций по использованию регистров и административных данных в целях переписей, включая оперативные, практические, технические и правовые аспекты. Настоящая публикация предназначена главным образом для национальных статистических управлений и других учреждений, ответственных за проведение переписи, и, как ожидается, также будет полезна для пользователей результатов переписи.

Настоящая публикация была подготовлена Целевой группой, учрежденной Конференцией европейских статистиков (КЕС) и состоящей из экспертов национальных статистических управлений, работа которой координировалась Европейской экономической комиссией Организации Объединенных Наций. КЕС одобрила данную публикацию на своей пленарной сессии 2018 года.

Выражение признательности

Настоящая публикация была подготовлена Целевой группой ЕЭК ООН по регистровым и комбинированным переписям при участии следующих членов: Эрика Шульте Нордхольта (Нидерланды, Председатель Целевой группы), Адельхайда Бауэра (Австрия), Жюли Трепанье (Канада), Дианы Белтадзе (Эстония), Томаса Кернера (Германия), Штефана Дитриха (Германия), Штефана Швайнерта-Альбинуса (Германия), Джона Данна (Ирландия), Яэля Файнштейна (Израиль), Джерардо Галло (Италия), Донателлы Зиндато (Италия), Кристины Байкрофт (Новая Зеландия), Харалда Утне (Норвегия), Янушы Дыгашевича (Польша), Сандры Лагарту (Португалия), Анабелы Дельгаду (Португалия), Паулы Паулину (Португалия), Данило Доленца (Словения), Бекки Тинсли (Соединенное Королевство), Винсента Мьюла (Соединенные Штаты), Адама Вронски (Евростат), Давида Торогуда (Евростат), Сорины Вайю (Евростат) и Паоло Валенте (ЕЭК ООН). Документ был отредактирован Яном Уайтом. Вклад всех этих экспертов достоин самой высокой оценки.

Содержание

Предисловие.....	iii
Выражение признательности.....	iv
Список вставок.....	x
Перечень диаграмм.....	x
Перечень таблиц.....	xi
1. Введение.....	1
1.1 Справочная информация.....	1
1.2 Методы проведения переписей в регионе ЕЭК ООН и их эволюция во времени.....	1
2. Сфера охвата новых руководящих принципов ЕЭК ООН и определения регистровых и комбинированных переписей.....	4
3. Основные характеристики переписи населения и жилищного фонда.....	6
3.1 Индивидуальный учет.....	6
3.2 Одновременность.....	7
3.3 Универсальность (в рамках четко определенной территории страны).....	8
3.4 Данные по малым районам.....	8
3.5 Установленная периодичность.....	9
3.6 Выводы, касающиеся характеристик, определенных КЕС.....	9
4. Соображения, которые следует учитывать при переходе от традиционной к регистровой или комбинированной переписи.....	10
4.1 Выгоды.....	10
4.1.1 Снижение затрат из расчета на душу населения.....	10
4.1.2 Сокращение сроков проведения.....	11
4.1.3 Сокращение коэффициента непредставления ответов и снижение нагрузки на респондентов.....	11
4.1.4 Возможность проведения непрерывной переписи.....	11
4.1.5 Улучшение сотрудничества между подразделениями НСУ.....	11
4.1.6 Больше времени и ресурсов на инновации.....	12
4.1.7 Повышение гибкости и чуткости к новым потребностям в информации.....	12

4.2	Необходимые условия для успешного перехода к регистровой или комбинированной переписи.....	12
4.2.1	Правовая основа.....	12
4.2.2	Общественное одобрение.....	13
4.2.3	Одобрение заинтересованными субъектами.....	13
4.2.4	Сотрудничество между НСУ и другими органами.....	14
4.2.5	Всеобъемлющая и надежная система статистических регистров.....	14
4.2.6	Унифицированная система идентификации.....	15
4.2.7	Знание административных источников.....	15
4.2.8	Прозрачность.....	15
4.3	Возможные трудности.....	15
4.3.1	Зависимость от государственных органов.....	15
4.3.2	Различия в концепциях и определениях.....	16
4.3.3	Своевременность административных регистров.....	16
4.3.4	Расхождения в учетных периодах.....	16
4.3.5	Проблемы конфиденциальности и безопасности.....	16
4.3.6	Трудности с определением подгрупп населения.....	17
4.3.7	Актуализация знаний и ИТ-инфраструктуры.....	17
4.3.8	Снижение интереса.....	17
5.	Общий подход к проведению регистровых и комбинированных переписей.....	18
5.1	Выявление источников данных.....	18
5.2	Процесс преобразования.....	19
5.3	Построение статистических регистров.....	20
5.4	Распространение результатов.....	21
5.5	Измерение/обеспечение качества.....	21
6.	Источники данных и их качество.....	22
7.	Увязка и преобразование.....	26
8.	Статистические регистры.....	27
9.	Качество на выходе.....	28
9.1	Качество продукта.....	28
9.1.1	Качество отдельной переписной переменной.....	30
9.1.2	Качество переписного гиперкуба.....	31
9.2	Охват.....	31
9.3	Качество и конфиденциальность.....	34
9.4	Сопоставление результатов переписи с результатами обследований.....	34
9.5	Отчеты о качестве.....	35
9.6	Группы по обзору качества.....	35

10	Подходы и практические примеры разных стран	36
10.1	Технические подходы	36
10.2	Опыт конкретных стран	37
	Приложение А – Практический пример Ирландии	38
	Общая обстановка.....	38
	Справочная информация о переписи.....	39
	Планы в отношении будущей переписи населения.....	39
	Справочная литература	40
	Приложение В – Практический пример Эстонии	41
	Отказ от традиционной переписи	41
	Подготовка к проведению регистровых переписей в Эстонии	42
	Пилотная перепись 2014 года	43
	План действий на период 2016–2020 годов.....	43
	Пробная перепись 2016 года.....	43
	Резюме.....	44
	Справочная литература	44
	Приложение С – Практический пример Польши	45
	Вводная информация о польской переписи населения и жилищного фонда 2011 года.....	45
	Использование административных источников	45
	Оценка качества данных из административных источников	46
	Другие методы сбора данных, использовавшиеся в ходе переписи.....	47
	Архитектура переписи: ИТ-система переписи.....	48
	Технология ГИС.....	49
	Геостатистический портал.....	49
	Резюме.....	50
	Приложение D – Практический пример Австрии	51
	Решение правительства	51
	Создание необходимых условий.....	51
	Центральный регистр населения (ЦРН).....	51
	Регистр уровня образования, Регистр учащихся и студентов.....	52
	Регистр зданий и жилищ (РЗЖ).....	52
	Отраслевой персональный идентификационный номер для официальной статистики (bPIN OC)	52
	Переписное законодательство	53

Основные принципы регистровой переписи	53
Охват переписи.....	53
Увязка регистров.....	53
Избыточность.....	54
Регистры	54
Анализ места жительства – избыточный охват в Регистре населения	55
Пробная перепись 2006 года.....	56
Ежегодная оценка численности населения по состоянию на 31 октября и регистровая статистика рынка труда	56
Оценка качества	57
Ожидаемые усовершенствования в рамках переписи 2021 года.....	57
Резюме.....	57
Справочная литература	58
Приложение Е – Практический пример Словении	59
История и эволюция переписи населения Словении с 1971 по 2002 год	59
Создание необходимых условий.....	60
Общие предварительные условия.....	60
Переписное законодательство	60
Новая дата переписи.....	61
Новые разработки после переписи 2002 года.....	61
Доперписная оценка качества исходных данных.....	61
Повышение качества данных о номерах жилищ в ЦРН.....	61
Основные принципы регистровой переписи	62
Увязка данных о лицах, домохозяйствах и жилищах.....	62
Качество базовых идентификаторов.....	62
Административные и статистические источники.....	62
Статистический процесс	64
Качество итоговых материалов регистровой переписи	65
Регулярный мониторинг качества.....	65
Специальное обследование охвата.....	66
Заключение	66
Справочная литература	66
Приложение F – Практический пример Португалии.....	67
Переписи в Португалии	67
Построение статистического набора данных о населении в Португалии	67
Источники и переменные административных данных.....	68

Ограничения в процессе построения португальского СНДН: конфиденциальность и защита частной жизни.....	69
Результаты и оценка качества.....	69
Выводы и перспективы.....	71
Приложение G – Практический пример Англии и Уэльса.....	72
Справочная информация.....	72
Что представляет собой перепись на основе административных данных?.....	73
Административные данные уровня единиц, имеющиеся в государственной администрации.....	73
Обследование охвата населения (ОБОН).....	73
Обследование для расчета оценок характеристик населения и жилищного фонда.....	73
Другие источники данных.....	73
Адаптация общей системы к подходу Англии и Уэльса.....	74
Что необходимо сделать для перехода к переписи на основе административных данных?.....	74
Оперативный доступ к существующим и новым источникам данных.....	75
Способность производить эффективно и точно увязку данных.....	75
Методы формирования статистических материалов достаточного качества для удовлетворения приоритетных информационных потребностей пользователей.....	75
Приемлемость для заинтересованных сторон (пользователей, поставщиков, общественности и парламента).....	76
Эффективность затрат.....	76
Как УНС узнает, возможна ли перепись на основе административных данных?.....	76
Справочная литература.....	78
Приложение H – Практический пример Италии.....	79
Справочная информация.....	79
Интегрированная система переписи и социальных обследований.....	80
Справочная литература.....	83
Приложение I – Практический пример Германии.....	84
Введение.....	84
Правовая и институциональная основа.....	84
Комбинированная модель переписи населения и ее внедрение в 2011 году.....	85
Извлеченные уроки и изменения в переписи населения 2021 года.....	87
Справочная литература.....	89
Глоссарий терминов, определений и сокращений.....	90

Список вставок

Вставка 6.1	Система качества в Нидерландах и Австрии.....	24
Вставка 9.1	Австрийская система обеспечения качества	30
Вставка 9.2	Оценка качества переписи с использованием административных баз данных в Эстонии.....	33

Перечень диаграмм

Диаграмма 1.1	Страны – члены ЕЭК ООН в разбивке по методу проведения переписи в ходе раундов переписи 2000–2020 годов.....	3
Диаграмма 5.1	Общий подход к проведению регистровых и комбинированных переписей	18
Диаграмма С.1	Переписные процессы и оценки качества.....	47
Диаграмма С.2	ИТ-система переписи.....	48
Диаграмма D.1	Регистровая модель переписи, базовые регистры и принцип избыточности.....	55
Диаграмма E.1	Идентификаторы, используемые при проведении регистровой переписи в Словении.....	62
Диаграмма F.1	Постоянное население в разбивке по возрастным группам, Португалия, 2011 и 2015 годы.....	70
Диаграмма F.2	Количество муниципалитетов в разбивке по величине расхождений между СНДН и ОЧН 2015 года	70
Диаграмма G.1	Общая система проведения регистровых или комбинированных переписей Соединенного Королевства	74
Диаграмма G.2	Текущая (2017 год) и ожидаемая в будущем (к 2023 году) оценка высокого уровня.....	77
Диаграмма H.1	Общая схема ИСПСО	83
Диаграмма I.1	Немецкая модель переписи населения 2011 года	85
Диаграмма I.2	Модель переписи 2021 года.....	87

Перечень таблиц

Таблица Е.1	Показатели качества ключевых идентификаторов регистровой переписи Словении.....	63
Таблица Е.2	Источники данных об уровне образования, регистровая перепись 2011 года, Словения.....	64
Таблица Е.3	Источники данных по экономическим характеристикам, регистровая перепись 2011 года, Словения.....	64

1. Введение

1.1 Справочная информация

1. В период с 2012 по 2015 год Руководящая группа ЕЭК ООН по переписям населения и жилищного фонда координировала подготовку *Рекомендаций Конференции европейских статистиков (КЕС) по проведению переписей населения и жилищного фонда 2020 года*. Руководящая группа направляла работу девяти тематических целевых групп, созданных для подготовки первоначальных проектов различных глав Рекомендаций. Впоследствии Рекомендации были приняты КЕС в июне 2015 года, и они доступны как в электронном формате на веб-сайте ЕЭК ООН¹, так и в печатном виде на английском, русском и французском языках.

2. В октябре 2015 года Бюро КЕС провело углубленный обзор вопроса диверсификации методологий и источников переписи населения на основе документа, подготовленного Финляндией и Турцией (ECE/CES/BUR/2015/OCT/3), и записки ЕЭК ООН (ECE/CES/BUR/2015/OCT/3Add.1). По итогам этого обзора Бюро поддержало подготовку новых руководящих принципов использования регистров и административных данных в целях переписей населения и жилищного фонда и поручило секретариату подготовить новые круги ведения Руководящей группы по переписям населения и жилищного фонда и Целевой группы по регистровым и комбинированным переписям (Доклад совещания Бюро: ECE/CES/BUR/2015/OCT/21).

3. Проект предлагаемых новых Руководящих принципов, подготовленный Целевой группой, был представлен и рассмотрен на совещании экспертов ЕЭК ООН по переписям населения и жилищного фонда в Женеве в октябре 2017 года и впоследствии был переработан с учетом замечаний, высказанных странами на совещании. В настоящей публикации Руководящие принципы представлены в том виде, в котором они впоследствии были согласованы КЕС.

4. Прежде чем перейти к Руководящим принципам в нижеследующем разделе кратко описываются методы проведения переписей, которые используют страны региона ЕЭК ООН, и их эволюция во времени.

1.2 Методы проведения переписей в регионе ЕЭК ООН и их эволюция во времени

5. Существует много различных способов проведения переписи населения и жилищного фонда. Для упрощения в настоящем документе кратко описываются лишь три основные методики проведения переписи: «традиционная», «регистровая» и «комбинированная». Более подробно различные методологии проведения переписи излагаются в Рекомендациях КЕС.

6. Под **традиционной переписью** здесь понимается перепись, основанная на прямом подсчете численности населения и сборе данных о его характеристиках посредством заполнения переписных листов в бумажной форме или электронном виде. Данные собирают на местах по всей стране в течение сравнительно короткого временного интервала, как правило, не превышающего двух недель. Переписные листы заполняются либо самими домохозяйствами (доставка и сбор анкет осуществляется переписчиками, почтовой службой или иными способами либо в режиме онлайн в случае использования электронных вопросников), либо переписчиками в ходе опроса домохозяйств.

7. Традиционная перепись имеет ряд недостатков. Прежде всего, это мероприятие является очень сложным и дорогостоящим, главным образом из-за необходимости временного найма большого числа людей для сбора данных на местах (переписчиков, контролеров и инспекторов), а также распечатки, распространения и обработки очень большого числа переписных листов. Кроме того, большинство стран испытывают все большие трудности с переписью определенных групп населения, особенно тех, которые

¹ <http://www.unecce.org/publications/2020recomm.html>.

являются высококомобильными и имеют несколько мест проживания, а также сталкиваются с усиливающимся нежеланием по разным причинам респондентов участвовать в переписи. Наконец, традиционная перепись обычно проводится раз в 10 лет (в силу высокой стоимости и сложности), и результаты часто становятся доступными по прошествии сравнительно длительного времени после завершения сбора данных, в то время как многим пользователям хотелось бы иметь более своевременную и чаще обновляемую информацию.

8. Для устранения недостатков традиционной переписи некоторые страны стали использовать выборочный метод (когда большинство домохозяйств заполняют краткий переписной лист с базовой информацией, а подробный переписной лист – выборка домохозяйств, что позволяет сократить общий объем собираемой и обрабатываемой информации) или поощрять самостоятельное заполнение онлайн-вопросников, что позволяет сократить расходы на сбор сведений на местах и повысить качество, но требует тщательного планирования и реализации. Другой подход заключается в том, чтобы собирать данные на местах в течение более длительных периодов времени и проводить выборочные обследования, как это делается во Франции, где применяется методика «скользящей переписи»².

9. Принципиально иным подходом по сравнению с традиционной переписью является **регистрационная перепись**, разработанная скандинавскими странами в 1970-х годах³. В 1981 году Дания стала первой в мире страной, которая провела перепись населения и жилищного фонда, полностью основанную на регистрах. При таком подходе прямой сбор информации от населения не проводится, а традиционная регистрация заменяется использованием административных данных из разных регистров (регистр населения, регистр зданий/адресов, регистр социального обеспечения и т. д.), процесс увязки которых обычно осуществляется на основе личных идентификационных номеров. При наличии качественной системы статистических регистров такой подход позволяет получать переписные данные при гораздо меньших затратах и с использованием относительно ограниченных людских ресурсов.

10. За период с 1990-х годов ряд других европейских стран разработали новаторские методы проведения переписи, которые сочетают использование административных данных с ограниченным опросом населения на местах по конкретным переменным. В рамках этого подхода, называемого **комбинированной переписью**, сбор данных на местах может носить как сплошной, так и выборочный характер. Этот подход часто используется при переходе от традиционной к регистрационной переписи.

11. В ходе раунда переписей 2000 года лишь несколько стран региона ЕЭК ООН⁴ использовали регистрационную или комбинированную методику переписи (три и пять стран соответственно⁴), и самым популярным в регионе по-прежнему оставался традиционный подход к переписи населения (40 стран)⁵. Однако в рамках раунда переписей 2010 года было отмечено значительное увеличение числа стран, проводящих регистрационную (с трех до девяти) или комбинированную перепись (с пяти до десяти), и соответствующее сокращение числа стран, использующих традиционную методику переписи (с 40 до 34 стран)⁴ (см. диаграмму 1.1).

12. Судя по информации о предварительных планах в отношении раунда 2020 года, тенденция отказа от традиционной методики переписи сохраняется: из 48 стран ЕЭК ООН, по которым имеется информация, 14 стран планируют провести регистрационную перепись (29%), 12 стран – комбинированную перепись (25%) и 22 страны – продолжить использовать традиционную перепись (46%). Если брать в расчет только 32 страны – члена ЕС (Европейский союз) и ЕАСТ (Европейская ассоциация свободной торговли), то проведение регистрационной переписи в ходе раунда 2020 года планируют 13 стран (41%), комбинированной переписи – 9 стран (28%), а дальнейшее использование традиционной методики – лишь 10 стран (31%).

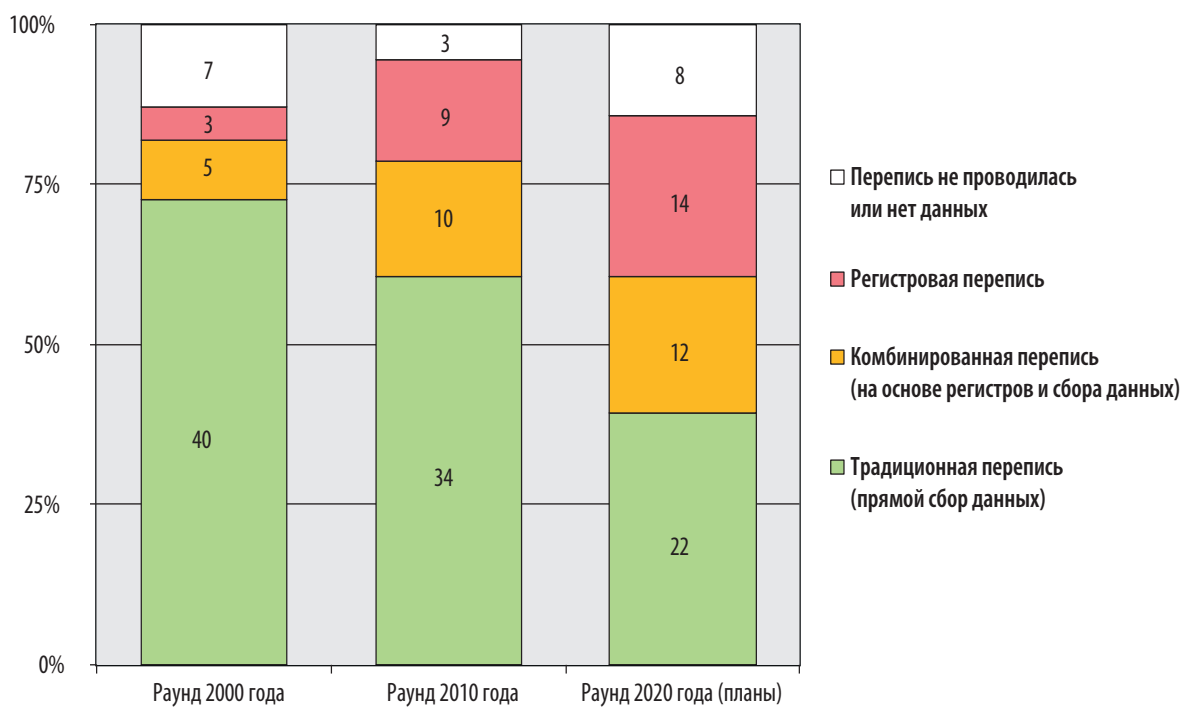
² INSEE, France, The French rolling census, ten years after its launch. Документ, представленный на совещании группы экспертов ЕЭК ООН-Евростата по переписям населения и жилищного фонда, Женева, 30 сентября – 3 октября 2013 года; https://www.unecce.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2013/census_meeting/24_E.pdf.

³ См. главу 10 документа: Register-Based Statistics in the Nordic Countries, UNECE, 2007 http://www.unecce.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf.

⁴ Регион ЕЭК ООН включает страны Европы, Северной Америки, Центральной Азии, а также Турцию и Израиль.

⁵ Источник: Valente, 2015, From the 2010 to the 2020 census round in the UNECE region – Plans by countries on census methodology and technology. Документ, представленный на совещании группы экспертов ЕЭК ООН-Евростата по переписям населения и жилищного фонда, Женева, 30 сентября – 2 октября 2015 года; http://www.unecce.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2015/mtg1/UNECE_paper_Paolo_draft_0925_rev2.pdf.

Диаграмма 1.1 Страны – члены ЕЭК ООН в разбивке по методу проведения переписи в ходе раундов переписи 2000–2020 годов⁶



Источник: Valente (2015)⁶.

⁶ В 2006 году Черногория стала независимой, и число стран ЕЭК увеличилось с 55 до 56. По адресу <https://statswiki.unecsc.org/display/censuses/2020+Population+Census+Round> можно найти обновленный перечень практики стран в рамках раунда переписей населения 2020 года.

2 Сфера охвата новых руководящих принципов ЕЭК ООН и определения регистровых и комбинированных переписей

13. Новые Руководящие принципы ЕЭК ООН посвящены не традиционным, а регистровым и комбинированным переписям. Как следствие, в них приводятся определения только регистровой и комбинированной переписей. Более подробную информацию о традиционной методике переписи можно найти в глобальных Рекомендациях⁷ и Рекомендациях КЕС ЕЭК ООН⁸ для раунда переписей 2020 года. В новых Руководящих принципах ЕЭК ООН упоминаются различные виды регистров (населения и зданий), причем особое внимание уделено регистрам, используемым в переписях.

14. Некоторые определения можно найти в публикации ЕЭК ООН Register-based statistics in the Nordic countries (Регистровая статистика в странах Северной Европы)⁹. В пункте 63 указанной публикации **регистр** определяется как систематизированный набор данных уровня единиц наблюдения, способ организации которых делает возможным их обновление. Обновление – это обработка поддающейся идентификации информации с целью создания, актуализации, исправления или расширения регистра, т.е. отслеживание любых изменений в данных, описывающих единицу наблюдения или ее свойства. Источниками **административных данных** являются массивы данных, содержащие собранную главным образом для административных (не исследовательских или статистических) целей информацию. Такого рода данные собираются государственными ведомствами и другими организациями в целях регистрации, проведения операций и учета, обычно в связи с предоставлением каких-то услуг. Речь идет о данных как административных регистров (имеющих уникальный идентификатор), так и, возможно, других административных данных без такого идентификатора. **Статистические регистры** являются регистрами, созданными для статистических целей. Обычно они создаются путем преобразования данных из регистров и/или других административных источников данных.

15. В некоторых странах термин «административные данные» используется в качестве синонима регистровых данных. В других странах между этими двумя понятиями проводится различие, и к «административным данным» относят данные не только из регистров, но и из других административных источников.

16. Система регистровой переписи населения опирается на набор базовых регистров, содержащих всеобъемлющие данные о единицах, подлежащих описанию в рамках переписи населения и жилищного фонда (см. пункт 123 Рекомендаций КЕС по проведению раунда переписей 2020 года). В некоторых странах, использующих регистровую перепись, некоторые переписные переменные отсутствуют во всех существующих регистрах, что заставляет эти страны дополнять данные переписи сведениями об отдельных единицах наблюдения (микроданными) из уже существующих выборочных обследований. Все страны, проводящие регистровую перепись, объединяет то, что они не используют специально разработанные переписные листы для сбора информации о населении. Поэтому регистровые переписи в целом значительно дешевле комбинированных переписей и тем более традиционных.

⁷ Принципы и рекомендации в отношении переписей населения и жилого фонда, Третье пересмотренное издание (United Nations, 2017), см. https://unstats.un.org/unsd/publication/seriesM/Series_M67Rev3en.pdf.

⁸ Рекомендации Конференции европейских статистиков по проведению переписей населения и жилищного фонда 2020 года (Организация Объединенных Наций, 2015 год), см. <http://www.unece.org/publications/2020recomm.html>.

⁹ http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf.

17. В основе формирования статистики комбинированной переписи лежат регистры и иные административные источники в сочетании с выборочным сбором данных на местах или сплошной регистрацией по избранным переменным¹⁰.

18. Остальная часть Руководящих принципов структурирована следующим образом. В главе 3 описываются основные характеристики переписи и то, насколько им удовлетворяет регистровая или комбинированная перепись. В главе 4 описывается ряд аспектов, которые необходимо учитывать при планировании перехода от традиционной к регистровой или комбинированной переписи. В главе 5 излагается общий подход, описывающий процесс проведения этих нетрадиционных переписей. В главах 6–9 приводится более подробная информация о процессах и методах, связанных с каждым этапом этого общего подхода, и описывается роль обеспечения качества на каждом этапе. В заключительной главе 10 и приложениях A-I приводятся практические примеры ряда стран, перешедших или планирующих перейти от традиционной сплошной к регистровой или комбинированной переписи.

¹⁰ Рекомендации по проведению переписей населения и жилищного фонда 2020 года (Организация Объединенных Наций, 2015 год), см. <http://www.unecsc.org/publications/2020recomm.html>, пункты 52 и 116.

3. Основные характеристики переписи населения и жилищного фонда

19. Основные характеристики переписи населения и жилищного фонда первоначально были определены еще в 1853 году на Международной конференции статистиков в Брюсселе, и их использование позволяло странам проводить переписи, результаты которых были сопоставимы на международном уровне с точки зрения методологии и качества. В настоящее время эти пять основных характеристик были переработаны и опубликованы КЕС¹¹ с целью обеспечения согласованности данных переписей, проводимых странами с разным уровнем технического развития и разными культурами. Использование всех этих характеристик позволяет НСУ (национальным статистическим управлениям) независимо от применяемой методологии собирать демографические данные, сопоставимые между странами и используемые для принятия решений и прогнозирования демографических процессов.

20. Пятью основными характеристиками переписи являются:

- ↑ индивидуальный учет;
- ↑ одновременность;
- ↑ универсальность (в рамках четко определенной территории страны);
- ↑ данные по малым районам;
- ↑ установленная периодичность.

21. В ниже следующих разделах рассматривается вопрос о том, как организовать регистровую или комбинированную перепись так, чтобы она отвечала каждой из этих характеристик.

3.1 Индивидуальный учет

22. Принцип индивидуального учета является одним из основополагающих для любой переписи населения. Традиционно это достигается за счет включения в переписной лист вопросов обо всех членах домохозяйства. В случае регистровой переписи применяется другой подход, при котором используются данные из административных регистров. В этом случае важно, чтобы по каждой единице переписи в используемых регистрах имелась отдельная, однозначно идентифицируемая информация. При проведении комбинированной переписи применяется тот же подход, однако информация из административных источников используется только в отношении некоторых переменных.

23. Если конкретная единица учета в разных регистрах не имеет единого идентификатора, необходимо создать новый статистический идентификатор (на основе группы идентифицирующих переменных), с тем чтобы обеспечить увязку переменных в соответствующих регистрах, и тщательно проверить его качество (на предмет ошибок и уникальности).

24. Иногда приходится рассчитывать данные по переписным переменным путем комбинирования данных из нескольких административных регистров с использованием специальных алгоритмов. Это возможно, если единицы учета во всех таких регистрах имеют один и тот же идентификатор. В этом случае созданная таким образом переменная будет идентифицироваться однозначно и сохраняться в статистическом регистре.

25. К основным учетным единицам переписи населения и жилищного фонда относятся не только лица, но также и домохозяйства, семьи и жилища (занятые или вакантные). Все эти единицы требуют идентификации, однако нет необходимости использовать для этого все различные идентификационные переменные. Минимальные необходимые идентификаторы – это идентификаторы лиц (идентификационный код лица)

¹¹ Рекомендации по проведению переписей населения и жилищного фонда 2020 года (Организация Объединенных Наций, 2015 год), см. <http://www.unecsc.org/publications/2020recomm.html>, пункты 23–28.

и жилища (идентификационный код жилища). Эти идентификационные коды должны быть связаны друг с другом, с тем чтобы идентификационные коды жилища были присвоены каждому лицу. Это обеспечивает привязку каждого занятого жилища к списку идентификационных кодов всех проживающих в нем лиц. Для идентификации жилого помещения используется код адреса, который также может содержать пространственные координаты.

26. Страны, проводящие традиционную перепись, обычно собирают информацию о домохозяйствах, основываясь на концепции совместного ведения хозяйства¹². Соблюдения этой концепции можно добиться, задавая соответствующие вопросы в рамках обследования или переписи, однако для стран, проводящих регистровые переписи, данная задача является более сложной. Многие из этих стран используют иную концепцию, а именно концепцию «совместного проживания», в соответствии с которой все лица, проживающие в одной жилищной единице, считаются членами одного и того же домохозяйства. Хотя применение подобного определения отражается на общем количестве частных домохозяйств в минимальной степени, оно может оказывать значительное влияние на некоторые типы домохозяйств, такие как домохозяйства, состоящие из одного лица. Данная погрешность в количестве частных домохозяйств и в их предполагаемой структуре зависит от традиций страны и условий жизни. Для стран, проводящих регистровую перепись, проблему также представляет вычленение семей внутри домохозяйств на основе информации о связях между их членами.

27. В некоторых странах (например, в Словении) существует регистр домохозяйств. Наличие такого регистра облегчает организацию регистровой переписи, особенно если в регистре содержатся идентификационные данные домохозяйства. В таком случае имеется точная информация о том, идентификационные коды какого физического лица относятся к идентификационным кодам какого домохозяйства. Регистр домохозяйств, таким образом, может значительно повысить качество регистровой или комбинированной переписи. Однако практика Словении является уникальной. Ирландия изучает возможность использования алгоритма дерева принятия решений для определения связей между людьми, проживающими в одном жилище, с тем чтобы бы продолжить использовать нынешнее определение домохозяйства, основанное на концепции совместного ведения хозяйства. (см. приложение А). С целью изучения последствий перехода к так называемой «переписи на основе административных данных» (см. приложение Г) Управление национальной статистики (УНС) Соединенного Королевства изучает возможное влияние на пользователей замены концепции совместного ведения хозяйства концепцией совместного проживания.

28. Порой идентификационные коды целесообразно использовать и в отношении других единиц, таких как предприятия и организации. Если они связаны с идентификаторами лиц или жилых помещений, то они становятся ценным инструментом для получения другой статистики, например статистики ежедневных поездок на работу и обратно.

3.2 Одновременность

29. Условием, определяющим одновременность данных переписи, является выбор фиксированного критического момента переписи. Традиционно для соблюдения данного условия перепись проводится в течение очень короткого периода времени, в идеале – в течение только одних суток. Хотя сегодня большинство переписей проводится в течение двух-трехнедельного периода, все собранные данные должны относиться к конкретному учетному периоду. Это важнейшее условие должно соблюдаться и в случае регистровой и комбинированной переписей.

30. Если используемые административные регистры регулярно обновляются, то необходимо установить переписной период и брать данные из всех регистров именно за этот период. Иногда регистры регулярно обновляются в одно и то же время, например в начале года, и именно эту дату следует выбирать для проведения переписи, гарантируя тем самым одновременность данных переписи.

31. В случае комбинированной переписи важно, чтобы учетный период переписи, указанный в переписных листах, и учетный период информации, взятой из регистров, совпадали или были как можно ближе друг к другу.

¹² См. пункты 768–769 <http://www.unec.org/publications/2020recomm.html>.

32. При использовании в ходе переписи нескольких административных регистров важно, чтобы все данные, взятые из них, относились к одному и тому же учетному периоду. Обычно для расчета переписных переменных с использованием специальных алгоритмов требуется некоторое время; таким образом, данные по этим переписным переменным могут быть готовы для опубликования лишь спустя некоторое время после переписи. В случае некоторых конкретных переменных в комбинированных или регистровых переписях устанавливаются разные учетные периоды по причине конкретной административной цели регистра. Демографические данные из регистров населения обычно можно брать в начале года. Однако данные о рабочей силе могут оказаться более актуальными за некоторое время до конца года, поскольку в период Рождества и Нового года занятость, как правило, снижается. В случае некоторых административных регистров, возможно, не удастся использовать день переписи в качестве учетного. Так, регистры образования часто содержат релевантные данные на дату (например, по состоянию на начало учебного года), отличающуюся от выбранного дня переписи. В таких случаях НСУ могут в качестве компромисса использовать образовательные данные по состоянию на дату, расположенную как можно ближе к дню переписи.

3.3 Универсальность (в рамках четко определенной территории страны)

33. Для обеспечения универсальности традиционной переписи переписные листы, используемые для регистрации населения, являются одинаковыми для всех домохозяйств и лиц. При использовании переписных листов на разных языках важно обеспечить, что содержание и смысл всех вопросов точно совпадали.

34. Если административные регистры, используемые в ходе переписи, являются едиными для всей страны и всех групп населения, то это условие можно считать выполненным. Однако при наличии различных административных данных в разных областях или по разным группам населения (например, в случае существования регистра городского населения и регистра сельского населения или различий в административном учете между городами) необходимо проанализировать возможные расхождения между административными источниками и найти способ определить единые переписные переменные на основе этих различных административных источников. В этом случае правдоподобные результаты могут быть получены с помощью новых переменных (определенных посредством соответствующего алгоритма) в статистическом регистре.

3.4 Данные по малым районам

35. Получение обширной информации по малым географическим районам и малым подгруппам населения (для которых здесь используется собирательное понятие «данные по малым районам») является ключевой целью переписи любого рода, поскольку другого единого источника сопоставимых данных, как правило, не существует. В последние годы значительно возрос спрос на геокодированные данные о населении, а также о зданиях и жилищах. Такое увеличение спроса привело к включению в Рекомендации КЕС по проведению раунда переписей населения 2020 года нового дополнительного признака «картограмма населения».

36. В контексте регистровых или комбинированных переписей данные по малым районам могут быть получены из административных источников при условии, что они имеют широкий, желателен сплошной охват. В случае недостаточного охвата некоторых малых районов и, как следствие, отсутствия некоторых административных данных (недостаточная универсальность) необходимо постараться улучшить набор административных данных, прежде чем он может быть использован в качестве источника для переписи. Подобный недостаточный охват способен создать проблемы с повседневным использованием административных данных, и поэтому данная проблема в любом случае требует решения. Усовершенствования статистического регистра иногда можно достичь путем добавления в него информации из другого источника при условии, что возможна увязка через единые идентификационные коды.

37. Иногда могут существовать специальные административные данные по некоторым малым районам (особенно по некоторым малым группам людей). В таком случае эти разные административные источники необходимо увязать между собой (см. раздел 3.2). При получении удовлетворительного результата такие объединенные данные вполне можно использовать. В случае проведения комбинированной переписи отсутствующую в административных источниках информацию можно также восполнить при помощи

обследования, в ходе которого, при необходимости, можно использовать различные методологии сбора данных, такие как опросы на дому или по телефону или самостоятельное заполнение бумажных или электронных вопросников с учетом специфики различных районов. Однако использование выборочного обследования не устраняет проблемы с охватом данных по малым районам. В подобных ситуациях до принятия любого решения относительно размера выборки обследования необходимо проанализировать, какой уровень детализации результатов переписи требуется пользователям.

3.5 Установленная периодичность

38. В настоящее время переписи, как правило, проводятся в мире с десятилетней периодичностью. Организация Объединенных Наций рекомендует странам проводить по меньшей мере одну перепись каждые десять лет (в период с 2015 по 2024 год в рамках раунда переписей 2020 года). В соответствии с требованиями Европейского союза страны-члены должны были (на основании соответствующего регламента ЕС) провести перепись в 2011 году, а следующую – в 2021 году. В то же время в некоторых странах (таких, как Австралия, Ирландия, Канада, Новая Зеландия и Словения) переписи проводятся чаще. Для обеспечения международной сопоставимости при проведении всех переписей, вне зависимости от методологии, следует придерживаться одного и того же базового десятилетнего цикла. При использовании пятилетнего цикла переписи государством – членом ЕС один из двух годов переписи должен совпадать с годом переписи, установленным Евростатом.

39. Одним из преимуществ регистровой переписи является возможность проводить ее чаще, чем раз в десять лет, поскольку регистровые данные доступны постоянно и обновляются чаще. Целесообразно также разработать программное обеспечение переписи таким образом, чтобы оно было постоянно готово для использования на любую учетную дату. В этом случае переписи можно проводить раз в десять лет, раз в пять лет, раз в два года или даже ежегодно. Ежегодное обновление демографических данных является одной из ключевых целей для многих европейских стран, которые в настоящее время используют разные методологии переписи.

40. Кроме того, можно готовить некоторые «обновления» итогов переписи, используя сокращенный перечень переменных (но достаточно полный для удовлетворения потребностей пользователей). Из этого также следует, что в странах, которые продолжают проводить регулярную перепись каждые десять лет с использованием комбинированной методологии, сокращенный перечень переменных, рассчитываемых на основе административных источников данных, можно будет обновлять чаще.

3.6 Выводы, касающиеся характеристик, определенных КЕС

41. Из вышесказанного можно сделать вывод о том, что если в стране имеется система административных регистров, являющихся взаимосвязанными, простыми в использовании и высококачественными (что означает, что все единицы идентифицируются однозначно с использованием единого идентификатора), или если источники административных данных позволяют создать статистические регистры аналогичного качества, то страна вполне может организовать перепись населения и жилищного фонда, которая будет удовлетворять всем требуемым КЕС характеристикам.

42. Если имеющийся набор административных источников данных не способен обеспечить требуемое качество в отношении всего спектра переписных переменных, использование комбинированной или регистровой методологии переписи все же делает возможным и целесообразным сбор данных по сокращенному перечню переменных (в том числе по тем, которые охватываются административными регистрами) в период между регулярными переписями.

4. Соображения, которые следует учитывать при переходе от традиционной к регистровой или комбинированной переписи

43. Аргументом в пользу принятия решения о переходе к комбинированной или регистровой переписи является то, что это сулит ряд выгод. Однако переход от традиционной к комбинированной или регистровой переписи требует тщательного подхода, и его успех зависит от выполнения определенных условий, касающихся вопросов данных, технологии, правовых вопросов, а также заинтересованности сторон. Это говорит о том, что в процессе перехода могут возникнуть определенные проблемы. Эти проблемы, о которых говорится в разделе 4.2, могут служить значительными препятствиями для некоторых стран.

4.1 Выгоды

44. Проведение комбинированной или регистровой переписи предполагает ряд выгод и возможностей, о которых речь пойдет ниже.

4.1.1 Снижение затрат из расчета на душу населения

45. Традиционные переписи являются очень дорогостоящими. Во многих странах, проводящих традиционную перепись населения, расходы на нее нередко эквивалентны примерно двум годовым бюджетам НСУ. Поэтому неудивительно, что правительства оказывают давление на управления, заставляя их сокращать эти расходы, особенно в тех случаях, когда имеются другие источники данных.

46. При проведении комбинированной переписи со сплошной регистрацией по некоторым переменным экономия средств, полученная за счет сокращения числа вопросов в переписном листе, частично нивелируется затратами на объединение данных сплошной регистрации с административными источниками данных. Хотя экономия может быть небольшой, такой подход все же может оказаться предпочтительным, особенно в тех случаях, когда процесс перехода позволяет использовать в будущих переписях больше данных из административных источников.

47. Проведение комбинированной переписи без сплошной регистрации по некоторым переменным позволяет добиться значительно большей экономии. Практика показывает, что переход к комбинированной переписи может привести к сокращению расходов на 22% по сравнению с традиционной переписью¹³. Кроме того, в случае проведения регистровой переписи никаких специальных обследований не требуется и экономия может оказаться очень существенной. Данные, представленные некоторыми странами, показывают, что таким образом в среднем можно сэкономить до 98% расходов на традиционную перепись¹⁴. Однако следует понимать, что такое сокращение затрат на проведение переписи может быть достигнуто только после того, как будет обеспечен требуемый доступ к соответствующим регистрам.

48. Некоторые страны перешли от традиционной к регистровой переписи в рамках одного цикла переписей. Этот шаг предполагает, что все расходы на внесение необходимых изменений придутся на один десятилетний период. Однако на практике такой переход, как правило, осуществляется в несколько этапов, и переходу к регистровой переписи зачастую предшествует внедрение промежуточной комбинированной модели переписи. Это позволяет распределить расходы между двумя–тремя циклами.

¹³ Расчеты на основе данных о ППС в таблице 7.2 документа http://www.unece.org/fileadmin/DAM/stats/publications/2013/Measuring_population_and_housing_2010.pdf.

¹⁴ Там же.

49. Переход к комбинированной и, в особенности, к регистровой модели способствует проведению более эффективной с точки зрения затрат переписи. Очевидно, что этот шаг может быть продиктован бюджетными ограничениями, налагаемыми правительством, даже если регистры являются неполными или недостаточно качественными для использования в качестве источников данных для переписи. В таких случаях с самого начала должно быть ясно, что стране не стоит пытаться перейти к новой системе незамедлительно и следует продолжать в той или иной форме проводить традиционную перепись. Однако даже в этих условиях инновации, требующие более широкого использования административных данных, способны повысить эффективность работы НСУ. Полезным шагом может оказаться решение государственных органов открыть статистическим управлениям доступ к административным источникам данных для формирования экспериментальной регистровой статистики. Правительство само может оказать большую помощь путем устранения любых юридических препятствий для обмена данными и благодаря субсидированию процесса перехода.

4.1.2 Сокращение сроков проведения

50. Страны, в которых создана статистическая система на основе регистров, тратят на проведение регистровой переписи гораздо меньше времени, чем страны, использующие другие виды переписи, главным образом по причине отсутствия необходимости проводить регистрацию данных переписчиками. Это, разумеется, не относится к комбинированным переписям, при проведении которых определенная работа на местах все же требуется. Без регистрации данных переписчиками меньше времени, как правило, уходит на планирование переписи (что позволяет сэкономить как время, так и деньги); при этом на подготовку результатов к публикации тратится не больше, а иногда и меньше времени. Однако следует понимать, что проведение первой регистровой переписи может потребовать больше времени, чем в последующие циклы, поскольку планировать ее придется с чистого листа.

4.1.3 Сокращение коэффициента непредставления ответов и снижение нагрузки на респондентов

51. В отсутствие необходимости проведения какой-либо регистрации переписчиками, если используются только – или в основном – административные источники данных, которые носят всеобъемлющий характер и охватывают все население, можно ожидать меньшего числа проблем, связанных с непредставлением ответов, и сведения нагрузки по предоставлению данных к нулю. В связи с тем что в международных переписях и обследованиях отмечается все более высокий уровень непредставления ответов, это становится все более важным преимуществом. Даже при переходе к комбинированной переписи нагрузка на население по предоставлению данных будет все равно ниже, особенно если регистрация данных переписчиками будет осуществляться только по выборке населения.

4.1.4 Возможность проведения непрерывной переписи

52. Чем больше административных данных использует страна при проведении переписи, тем шире ее возможности по проведению переписи ежегодно или еще чаще. Теоретически при наличии качественных административных данных обновлять результаты можно было бы на ежедневной основе. Такие переписи в режиме реального времени, возможно, будут проводиться в будущем, однако в настоящее время пользователи ожидают, что переписная статистика будет распространяться более регулярно, чем раз в десять лет. По мере того как данные из других источников смогут обеспечить получение такой информации в режиме реального времени, переписи населения, как ожидается, будут идти в ногу со временем. Ежегодное проведение переписи могло бы косвенно способствовать обновлению НСУ своих знаний и ИТ-инфраструктуры.

4.1.5 Улучшение сотрудничества между подразделениями НСУ

53. В некоторых НСУ различные отделы или подразделения часто структурно организованы так, что вынуждены работать обособленно, без активного взаимодействия. При переходе к регистровой статистической системе НСУ может отказаться от подобной разобщенной структуры. В результате перехода от статистической организации, основанной на обследованиях, к системе, основанной на регистрах, традиционная взаимосвязь «один к одному» между источниками и статистикой заменяется взаимосвязью «многие ко многим» между всеми отраслями статистики. В связи с этим существенно более значение приобретает улучшение сотрудничества подразделений в рамках НСУ. Кроме того, за счет более эффективной интеграции статистики повышается согласованность статистической структуры в рамках НСУ.

4.1.6 Больше времени и ресурсов на инновации

54. Инновации имеют решающее значение для долгосрочного развития НСУ. Поскольку включение статистических регистров в процесс производства официальной статистики экономит как время, так и деньги, становится легче проводить инновации. Сэкономленные таким образом ресурсы можно было бы использовать для стимулирования административных и технологических инноваций, с тем чтобы процессы производства данных продолжали идти в ногу со временем.

4.1.7 Повышение гибкости и чуткости к новым потребностям в информации

55. При надлежащем хранении всех данных можно не только чаще готовить регулярную статистику, но и создавать ее новые виды для удовлетворения меняющихся потребностей пользователей. Это позволяет НСУ быть более гибкими и чуткими к новым потребностям в информации и повышать свою полезность в глазах общества. Хотя это может и не являться самоцелью, однако способно привести к более высокому уровню удовлетворенности пользователей.

4.2 Необходимые условия для успешного перехода к регистровой или комбинированной переписи

56. Если страна желает перейти к комбинированной или регистровой переписи, она должна выполнить ряд условий, прежде чем информация из административных регистров (и других источников) может быть успешно интегрирована для создания базового статистического регистра. Выполнение некоторых из них может создать для НСУ определенные проблемы, которые рассматриваются ниже.

4.2.1 Правовая основа

57. Независимо от типа переписи, которую проводит НСУ, она должна проводиться в рамках установленной правовой основы. В частности, для проведения комбинированной или регистровой переписи необходимы правовые нормы, обеспечивающие доступ к административным данным и их защиту. Подобные правовые нормы обычно закреплены в Законе о статистике или в Законе о переписи.

58. Закон должен наделять НСУ правом доступа, в идеале бесплатного, к соответствующим административным данным в любых государственных источниках, предпочтительно и к персональным идентификаторам. Во избежание правовой неопределенности или споров следует оговорить, что право доступа не применимо в судебных разбирательствах, связанных с охраной общественного порядка или безопасностью страны. Еще один важнейший вопрос, который должен быть урегулирован законодательно, связан с необходимостью наделить НСУ определенным влиянием или полномочиями в областях создания, пересмотра или удаления тех административных данных, которые предполагается использовать в статистических регистрах.

59. В свою очередь НСУ должны быть обязаны по закону защищать конфиденциальность административных данных, которые они получают, и придерживаться принципа «одностороннего движения», за исключением особых обстоятельств, предусмотренных в законе. Действительно, соответствующее законодательство могло бы пойти дальше, полностью запретив другим управляющим данными доступ к данным в статистических регистрах НСУ.

60. В ряде стран порядок проведения переписи регулируется законодательно¹⁵. В некоторых странах НСУ начали с изучения административных источников данных и лишь затем определили правовую основу, сделавшую регистровую статистику возможной. В других странах сначала была создана правовая база, на основе которой затем была налажена подготовка и публикация регистровой статистики. Для того чтобы накопить опыт для постепенного перехода к регистровой переписи, НСУ зачастую проще начать с формирования регистровой статистики лишь по некоторым переменным, данные по которым собираются в ходе традиционной переписи, хотя следует признать, что юридических барьеров на этом пути может оказаться ничуть не меньше.

¹⁵ Особенно в тех случаях, когда представленность в национальном законодательном органе зависит от результатов переписи.

61. Переход на методику переписи, в которой административные источники данных играют важную роль, всегда требует тщательной подготовки, в том числе проведения пилотных исследований. НСУ должны понимать, что после полного или частичного отказа от регистрации данных переписчиками возобновить ее будет довольно сложно. По истечении некоторого времени опыт проведения традиционной переписи утрачивается, особенно в случае перехода на регистровую перепись, не предполагающую никакой регистрации данных переписчиками.

62. Законодательство, регулирующее доступ к административным данным, возможно, потребуется дополнить внутренними мерами политики и директивами НСУ, которые обеспечат воплощение требований законодательства и политики и директив центрального правительства в требования и обязанности для руководителей и сотрудников НСУ.

4.2.2 Общественное одобрение

63. Хотя закон может наделять НСУ правовыми полномочиями на проведение комбинированной или регистровой переписи, необходимо также заручиться поддержкой общественности для обеспечения того, чтобы такая перепись была приемлемой. Эта задача может оказаться более сложной, чем создание правовой основы. В то время как в некоторых странах в ходе даже традиционной переписи у людей может создаться впечатление, что «большой брат следит за ними», в других странах использование и увязка административных данных, собранных в нестатистических целях, могут рассматриваться в качестве еще более серьезного посягательства на конфиденциальность, поскольку общественность не имеет никакого контроля над информацией о себе, подлежащей раскрытию.

64. В рамках традиционной переписи озабоченность по поводу конфиденциальности может привести к снижению уровня участия в переписи или преднамеренному предоставлению неверной информации. Становится все труднее проводить корректировки на отсутствие ответа по подобным единицам и отдельным позициям. Таким образом, с одной стороны, общественность может предпочесть, чтобы ей задавали меньше вопросов, если аналогичная информация уже имеется в наличии. С другой стороны, часть населения может предпочесть самостоятельное заполнение переписного листа извлечению и комбинированию информации о нем из нескольких административных источников.

65. При проведении регистровой или комбинированной переписи люди могут испытывать беспокойство по поводу того, что информация из различных административных источников будет повторно использоваться и увязываться в целях переписи, или даже возражать против этого. Люди могут не понимать, что в рамках переписи информация используется исключительно в статистических целях. В отсутствие переписных листов и при использовании только регистров общественность, как правило, будет менее осведомлена о проведении переписи. Однако отсутствие какой-либо реакции общественности не следует истолковывать как одобрение с ее стороны.

66. Поэтому при проведении регистровой или комбинированной переписи желательно подготовиться к возможным конкретным вопросам, касающимся проблем неприкосновенности частной жизни, конфиденциальности и безопасности. При проведении комбинированных переписей можно ожидать дискуссии на тему того, какие переменные должны включаться в переписные листы, а какие – братья из административных источников.

4.2.3 Одобрение заинтересованными субъектами

67. Заинтересованные субъекты, а точнее, пользователи данных, обычно ожидают от каждой переписи получения не менее детальной информации, чем от предыдущей. Однако при изменении методологии переписи это не всегда возможно.

68. Важно заблаговременно информировать заинтересованных субъектов и проконсультироваться с ними. Когда ожидания пользователей не оправдываются, они вполне могут занять критическую позицию. Даже хорошо информированные пользователи могут превратиться в ярых критиков, если они поймут, что могут лишиться части информации, к которой они имели доступ благодаря предыдущей переписи. Тем не менее удовлетворить всех пользователей, как правило, не представляется возможным, и разочарование некоторых из них от перехода к новой методологии переписи зачастую неизбежно.

69. Поэтому важно иметь коммуникационную стратегию для взаимодействия с заинтересованными субъектами, которая должна охватывать следующие цели или хотя бы некоторые из них:

- ↑ создать обстановку прозрачности вокруг планов НСУ;
- ↑ заверить пользователей в том, что их потребности будут учтены;
- ↑ убедить заинтересованных субъектов в выгодах использования административных данных и продемонстрировать им, что информация будет по-прежнему защищена;
- ↑ укреплять партнерские отношения с заинтересованными субъектами, с тем чтобы НСУ могло задействовать потенциал внешних экспертов;
- ↑ обеспечить участие заинтересованных субъектов в успешном переходе на новую методологию проведения переписи.

70. Открытость и четкая демонстрация заинтересованным субъектам новых возможностей и выгод позволят заручиться их поддержкой. Особенно важно проводить достаточные консультации по любым изменениям в предоставлении тех статистических данных, которые имеют финансовые последствия для заинтересованных субъектов (например, денежные трансферты муниципалитетам).

4.2.4 Сотрудничество между НСУ и другими органами

71. Для использования административных источников данных в целях переписи чрезвычайно важное значение имеет эффективное сотрудничество между НСУ и другими (главным образом правительственными) органами. Прежде чем приступить к формированию регистровой статистики НСУ, необходимо знать, когда они смогут получить микроданные (данные уровня административных единиц) и соответствующие метаданные. В рамках комбинированной и тем более регистровой переписи выполнение НСУ своих договорных или законодательно установленных обязательств по своевременному предоставлению данных надлежащего качества в значительной степени зависит от держателей административных данных. Если держатели данных не предоставят свои данные, ответственность за несвоевременную публикацию переписной статистики будет нести НСУ.

72. Крайне важно информировать держателей административных данных о том, насколько важны их данные для НСУ и как их данные будут использоваться. Помимо правовой основы (см. раздел 4.2.1) и хороших отношений с другими органами, содействовать процессу переписи также может подписание договоров о сотрудничестве или соглашений о гарантированном уровне обслуживания. Теоретически в целях переписи могут использоваться и негосударственные источники административных данных, однако нередко это сопряжено с проблемами обеспечения конфиденциальности и качества данных и с соображениями коммерческого характера; использование частных данных во многих случаях обходится НСУ очень недешево.

4.2.5 Всеобъемлющая и надежная система статистических регистров

73. Всеобъемлющая и надежная система статистических регистров (содержащая точные и актуальные данные) имеет чрезвычайно важное значение для проведения комбинированной или регистровой переписи. Административные источники данных, включая административные регистры, например регистр населения, обычно создаются не для статистических целей, таких как проведение переписи. Поэтому создание надежной системы статистических регистров требует процесса преобразования (см. главу 7).

74. Для обеспечения использования регистровых данных в целях официальной статистики важно иметь хорошие рабочие отношения с держателями административных данных. В зависимости от положений правовой основы и в том случае, если административные органы будут расположены к этому, в некоторых странах существуют также возможности для улучшения отношений между держателями данных и НСУ благодаря созданию новой или расширению существующей регистровой статистики, например за счет продольных исследований, позволяющих оценить процесс реализации политики. Разумеется, всякий раз, когда в качестве нового источника начинает использоваться административный источник данных, НСУ должно связываться с соответствующим держателем административных данных. Вместе с тем важнейшее значение имеют постоянные контакты, например через менеджеров по работе с клиентами, которые позволяют информировать держателей административных данных о той важной роли, которую играют их данные, а НСУ –

о любых изменениях в получаемых ими микроданных и метаданных. Залогом успешного использования регистровой статистики являются регулярные контакты между НСУ и держателями административных данных.

75. Во многих странах, проводящих регистровые переписи, система административных источников данных используется самыми разными государственными органами. Чем больше пользователей у этой системы, тем выше качество, на которое можно рассчитывать. При использовании такой системы в целях переписи главным соображением является качество результирующего статистического регистра, а не задействованных источников административных данных: достаточно ли высоко качество этих данных для получения достоверных результатов переписи?

4.2.6 Унифицированная система идентификации

76. Использование унифицированной системы идентификации в различных административных источниках данных значительно облегчает проведение регистровых переписей. Предпочтительно иметь уникальные идентификационные коды на уровне отдельных единиц, являющиеся едиными для всех регистров. Для стран, в которых не существует уникальных личных идентификационных кодов, задача эффективной и точной увязки данных выглядит особенно трудной.

4.2.7 Знание административных источников

77. Когда страна желает перейти от традиционной к комбинированной или регистровой переписи, до начала такого перехода ей важно всесторонне и подробно выяснить, какие данные имеются в административных источниках. Хотя получение таких знаний может происходить на раннем этапе планирования переписи, нельзя недооценивать те усилия, которые могут потребоваться для обеспечения успеха этого процесса. Хотя важные уроки можно извлечь из успешного или неудачного опыта стран, проводящих комбинированные или регистровые переписи, следует всегда учитывать и национальный контекст. НСУ ни в коем случае не рекомендуется слепо копировать методологию другой страны при разработке комбинированной или регистровой переписи. Однако извлечение уроков из опыта других может значительно сократить процесс перехода.

4.2.8 Прозрачность

78. При планировании перехода на другую методологию проведения переписи рекомендуется действовать прозрачно и максимально делиться с заинтересованными субъектами информацией о планах и пробных мероприятиях. Как отмечалось выше в разделе 4.2.3, чрезвычайно важно информировать пользователей о любых решениях по переходу к регистровой переписи, поскольку такое важное изменение в методологии, вероятно, скажется на содержании и доступности результатов. Прозрачность и открытость облегчают внешний контроль новых процессов и обратную связь.

4.3 Возможные трудности

79. Несмотря на выгоды, отмеченные выше в разделе 4.1, проведение комбинированной или регистровой переписи сопряжено с рядом недостатков и рисков, которые описываются ниже.

4.3.1 Зависимость от государственных органов

80. Переходя к комбинированной или регистровой переписи, НСУ попадают в сильную зависимость от государственных органов, являющихся держателями используемых административных данных. НСУ должны осознавать, что для таких органов подготовка статистических данных не является основным видом деятельности, которому они обычно отдают приоритет. Любые ошибки или недочеты в административных регистрах будут отрицательно влиять на качество официальной статистики НСУ, за которую оно будет нести ответственность.

4.3.2 Различия в концепциях и определениях

81. В регистрах и других источниках административных данных часто используются концепции и определения демографических переменных, отличные от тех, которые обычно применяются в традиционных переписях. НСУ должны осознавать возможность существования таких различий и определять, являются ли они приемлемыми, при переходе от традиционной к комбинированной или регистровой переписи. Когда на одной чаше весов лежит преимущество и согласованность результирующей статистики, а на другой – сокращение расходов на регистрацию данных переписчиками, то различия, которые в одних странах могут считаться допустимыми, в других странах могут восприниматься пользователями как абсолютно неприемлемые. НСУ должны взвесить все за и против, прежде чем принимать решение о том, желают ли они заплатить эту цену за переход к регистровой или комбинированной переписи, отказавшись от сплошной регистрации данных переписчиками по отдельным переменным. Иногда добиться достаточно точного соответствия исходным определениям и концепциям позволяет использование производных из различных источников переменных или редактирование информации из новых источников, используемых в целях переписи. Однако этого можно добиться не всегда, и в таких случаях НСУ следует выбирать – закрывать глаза на существующие различия или продолжать использовать метод сплошной регистрации в отношении отдельных переменных с соответствующими затратами. В целях будущих (после 2020 года) раундов переписей, когда будет производиться пересмотр базовых концепций переписи (например, категорий переписного населения), следует изучить возможность получения необходимых данных из регистров. Это могло бы облегчить использование регистров и в то же время гарантировать согласованность на международном уровне.

4.3.3 Своевременность административных регистров

82. Государственные органы, отвечающие за ведение административных регистров, не хранят данные в статистических целях и имеют другие приоритеты, что может привести к задержкам в предоставлении НСУ соответствующих административных данных и метаданных. Это может создать проблемы для НСУ с точки зрения своевременности их регистровой статистики, особенно в тех случаях, когда сроки представления данных из различных источников сильно варьируются.

4.3.4 Расхождения в учетных периодах

83. Конкретная проблема, с которой сталкиваются НСУ при переходе к комбинированной или регистровой переписи, заключается в том, что различные источники административных данных часто имеют разные учетные даты. Иногда источник позволяет провести четкое различие между учетной датой и датой события, однако подобная полезная практика применяется не всегда. Без должного решения этой проблемы существует риск того, что не все источники будут приведены к единой учетной дате. В этой связи встает вопрос: «Каково допустимое расхождение в учетных датах?». Ответ на этот вопрос зависит от того, о какой переменной идет речь. Некоторые переменные остаются довольно стабильными в течение длительного времени, и поэтому небольшие различия в учетных датах, как правило, не являются проблемой. Значительные различия в учетных датах всегда являются нежелательными. Наконец, важно понимать, что даже при традиционной переписи не вся информация на практике может относиться к одной контрольной дате переписи. В тех случаях, когда предоставляемая для переписи информация касается отдаленных событий, в действие вступает фактор памяти, и респонденты не всегда предоставляют ответы, относящиеся к контрольной дате переписи.

4.3.5 Проблемы конфиденциальности и безопасности

84. Использование административных данных в целях, отличных от тех, для которых первоначально была получена данная информация, неизбежно вызывает проблемы конфиденциальности и безопасности. Эти проблемы часто касаются увязки персональных данных из разных источников. Некоторые страны внесли исходя из этого специальные изменения в нормативно-правовую основу, из которых можно заключить, что их общество одобряет (или по меньшей мере допускает) использование административных данных в целях официальной статистики. В других странах подобный консенсус еще не достигнут.

4.3.6 Трудности с определением подгрупп населения

85. При проведении переписи важно обеспечить универсальность. Однако диапазон и степень детализации результатов регистровой переписи ограничиваются теми переменными, которые могут быть рассчитаны на основе существующих источников. Не все они могут относиться ко всему населению. Кроме того, даже тем странам, которые для сбора информации, не имеющейся в административных источниках, используют выборочные обследования, иногда бывает трудно или даже невозможно получить точные результаты по малым районам или конкретным подгруппам населения из-за размера выборки.

4.3.7 Актуализация знаний и ИТ-инфраструктуры

86. Странам, которые проводят переписи с большими перерывами, может быть непросто удерживать в НСУ персонал, обладающий необходимым опытом и знаниями, для того чтобы в периоды между переписями обеспечивать актуализацию знаний и ИТ-инфраструктуры. Однако ежегодное обновление результатов переписи позволяет свести эту проблему к минимуму.

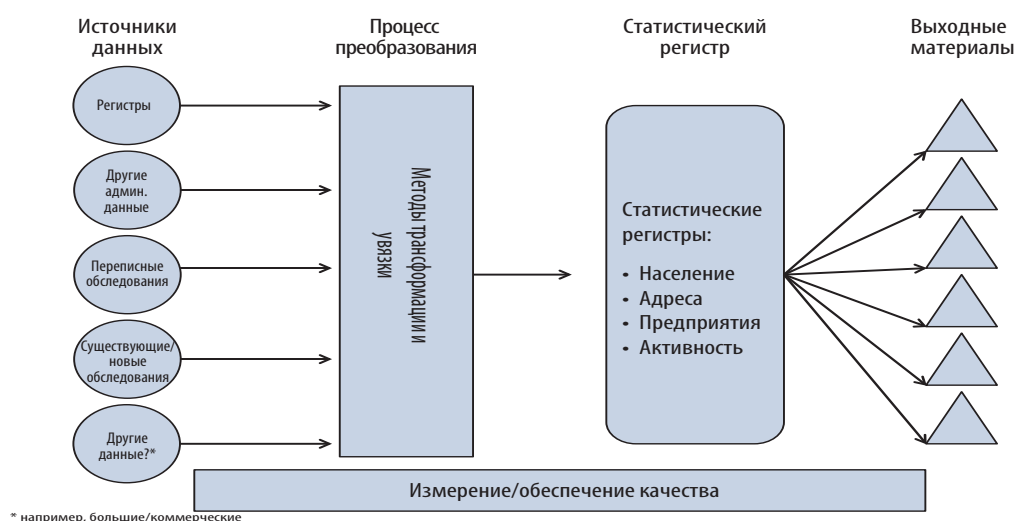
4.3.8 Снижение интереса

87. В странах, где переписи проводятся с использованием переписных листов, не только пользователи, но и широкая общественность будут интересоваться их результатами. Однако в странах, перешедших на регистровые переписи, где люди больше не заполняют переписные листы, интерес общественности к результатам переписи часто снижается. Многие люди даже не подозревают о проведении переписи, и, как следствие, общенациональный интерес к переписям значительно ослабевает. Пользователи продолжают интересоваться статистическими результатами, хотя информация из стран, проводящих регистровые переписи, свидетельствует о том, что интерес пользователей к выбору исходных источников и методологии переписи со временем снижается. Это также связано с тем, что в этих странах другие результаты зачастую доступны раньше, чем результаты переписи.

5 Общий подход к проведению регистровых и комбинированных переписей

88. Как отмечалось в главе 1 настоящих Руководящих принципов, все больше стран переходят к регистровой или комбинированной переписи. Однако методы и процессы, которые каждая страна может использовать для проведения такой переписи, могут различаться. В этой связи представляется полезным рассмотреть общий подход, который мог бы применяться к их проведению, с указанием требуемых ключевых этапов, указанных на диаграмме 5.1.

Диаграмма 5.1 Общий подход к проведению регистровых и комбинированных переписей



89. Этот подход состоит из пяти основных этапов:

- ↑ выявление источников данных;
- ↑ процесс преобразования;
- ↑ построение статистических регистров;
- ↑ распространение результатов;
- ↑ измерение/обеспечение качества.

90. Далее в настоящей главе дается краткое описание каждого из этих этапов.

5.1 Выявление источников данных

91. Регистровые переписи, как правило, проводятся на основе регистров населения, однако в последнее время страны, не имеющие отлаженного общенационального регистра населения, изучают возможности комбинирования других административных источников для создания эквивалентного статистического регистра населения. Страны, как правило, используют целый ряд других источников для повышения качества или охвата результатов, которые могут быть получены в рамках регистровой переписи. К их числу могут

относиться административные данные, выборочные обследования и другие источники данных, такие как большие данные или коммерческие данные.

92. Например, в рамках своей работы по изучению возможности перехода к переписи на основе административных данных (см. приложение G) Управление национальной статистики (УНС) Соединенного Королевства разработало набор данных статистики народонаселения (НДСН)¹⁶, на основе которого оно будет рассчитывать результаты, схожие с итогами переписи. НДСН призван обеспечить создание единого, согласованного набора данных, который будет служить основой для оценки численности обычно проживающего населения. Он формируется путем увязки записей из множественных административных источников данных и применения набора правил включения и распределения для достижения желаемых результатов. Аналогичная работа проводится в настоящее время в Центральном статистическом управлении Польши (ЦСУ) (см. приложение С).

93. При проведении переписи населения и жилищного фонда с использованием множественных источников данных, включая широкое использование административных регистров, важно рассмотреть вопрос о том, возможно ли интегрировать эти источники. Самым простым способом увязки различных источников данных является использование уникальных идентификаторов в отношении физических лиц и адресов (как указано в разделе 3.1). В некоторых странах существуют также и другие уникальные идентификаторы, например предприятий, которые являются весьма полезными для увязки данных. Наличие идентификаторов, представляющих собой интегрирующие переменные, позволяет использовать данные из различных источников. Однако отсутствие таких идентификаторов не обязательно будет препятствовать интеграции многочисленных источников данных. Процесс объединения данных из нескольких источников, относящихся к одним и тем же единицам, может осуществляться с использованием различных детерминистических и стохастических методов. В случае детерминистических методов ключевой переменной, объединяющей все наборы, служит идентификатор, появляющийся во всех источниках. В случае стохастических методов должен быть создан соответствующий идентификатор с использованием информации атрибута, который является общим для всех комбинируемых источников.

94. Набор источников, которые могут быть включены для удовлетворения основным характеристикам, изложенным в главе 3, вероятно, будет разным для каждой страны. В случае тех, кто проводит комбинированную перепись, ключевым отличием является включение переписных данных либо выборочной регистрации переписчиками по избранным переменным, либо сплошной регистрации переписчиками по избранным переменным.

5.2 Процесс преобразования

95. Поскольку данные в регистрах населения и других административных источниках не собираются в первую очередь для целей переписи, создание надежной системы статистических регистров требует процесса преобразования. В странах, которые используют такие данные в рамках своих регистровых или комбинированных переписей, обычно существует необходимость увязки данных из целого ряда источников и преобразования этих увязанных данных в статистические регистры.

96. Первый шаг заключается в увязке различных источников данных, которые будут использоваться. В странах, где существует унифицированная система идентификации (т.е. где все соответствующие регистры и источники данных используют один и тот же уникальный идентификатор для людей, предпочтительно идентификационные коды), этот процесс является относительно простым и позволяет свести к минимуму ошибки в процессе увязки.

97. Это является более сложной задачей для стран, которые не имеют такой унифицированной системы идентификации или которые используют административные источники данных без уникального идентификатора. Однако иногда это может быть достигнуто путем проведения увязки с помощью набора идентификаторов (таких, как имя, пол, дата рождения, цифровой адрес), дающей совпадения, которые могут быть использованы для получения результатов с той же точностью, что и результаты, полученные путем увязки регистров, которые содержат уникальный идентификатор. Некоторые из этих методов кратко рассматриваются в главе 7. Страны, находящиеся в такой ситуации, могут рассмотреть вопрос о присвоении

¹⁶ Новейшую методологию см. по адресу <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/methodology/methodologyofstatisticalpopulationdatasetv20>.

уникального идентификатора в рамках создания статистического регистра для обеспечения возможности интеграции и эффективного использования данных. Если такие успешные связи (т.е. связи, которые имеют хорошее качество и, следовательно, не обременены большим количеством ошибок) установить невозможно, то страна вряд ли не может перейти к комбинированной или регистровой переписи, но, тем не менее, административные данные могут затем использоваться в целях сравнительного анализа или обеспечения качества (см. в приложении А пример Ирландии).

98. После увязки данных может потребоваться их дальнейшее преобразование для создания или повышения качества статистического регистра. В главе 2 мы определили административные источники данных как «массивы данных, которые содержат информацию, собранную главным образом для административных (не исследовательских или статистических) целей». В странах, которые используют такие данные в рамках своих регистровых или комбинированных переписей, обычно существует необходимость преобразования увязанных данных в статистические регистры. Существуют две основные цели построения статистических регистров в НСУ. Первая – очистка и редактирование данных; вторая – формирование переменных переписи с использованием специальных алгоритмов. Иногда для достижения этой второй цели данные из различных регистров могут быть эффективно объединены в статистический регистр. Этот процесс будет варьироваться в зависимости от страны и может включать в себя добавление информации из дополнительных источников данных (через увязку), а также проведение некоторых статистических процедур. Такие примеры могут включать очистку, удаление ошибочных значений, устранение расхождений между источниками (например, адресная информация может различаться в целом ряде источников), редактирование и кодирование данных, поиск и замещение отсутствующих значений (возможно, методом импутации) и выбор записей, отвечающих критериям представляющей интерес группы населения (например, проживающие на территории страны на момент переписи).

99. Статистические регистры необходимы для всех НСУ, проводящих регистровые переписи, однако способность преобразовать многочисленные административные источники данных в статистические регистры часто является ключевой проблемой для НСУ, не имеющих регистров населения.

5.3 Построение статистических регистров

100. Всеобъемлющая и надежная система статистических регистров является необходимым условием для проведения комбинированной или регистровой переписи. Обычно полезно иметь систему, состоящую из нескольких регистров, которые могут быть соединены между собой с помощью связей между идентификаторами каждой из первичных единиц регистра, показывающих, каким образом эти единицы связаны между собой. В работе Wallgren и Wallgren (2014) говорится о создании четырех отдельных связанных базовых статистических регистров:

- ↑ регистр населения – регистр жителей страны. Он может существовать во многих версиях, отражая: а) текущую численность населения, б) численность населения в определенный момент времени (например, на 31 декабря), в) все изменения в течение определенного периода (например, календарного года) и г) численность населения, постоянно проживающего в течение определенного периода (например, более одного календарного года);
- ↑ регистр адресов/жилищ – регистр адресов/жилищ;
- ↑ коммерческий регистр – регистр предприятий;
- ↑ регистр активности – регистр, содержащий информацию о различных видах активности резидентов. Этот регистр обычно состоит из трех разделов: а) занятость или трудовая деятельность, б) учебная деятельность и в) другая деятельность, связанная с рынком труда (например, периоды безработицы, военной службы, выплаты пособий и пенсии).

101. В эстонской системе регистров, например, имеется четыре регистра, но их содержание несколько отличается:

- ↑ регистр населения – регистр жителей страны;
- ↑ регистр адресов/жилищ – регистр адресов/жилищ;
- ↑ коммерческий регистр – регистр предприятий;
- ↑ регистр фермерских хозяйств – регистр фермерских домохозяйств.

102. Все регистры связаны между собой с помощью различных идентификаторов (личный ИН, ИН адреса, ИН предприятия и ИН фермерского хозяйства). Эстония осуществляет переход от комбинированной к регистровой переписи (см. приложение В).

103. Польский опыт свидетельствует о настоятельной необходимости того, чтобы регистр адресов/жилищ содержал географический компонент для определения пространственного местоположения каждого жилища (здания) с максимально возможной точностью координат по оси x–y. Это позволит, в частности, использовать технологии ГИС для поддержки основных этапов регистрации данных переписчиками и последующего пространственного анализа. Польша проводит комбинированную перепись (см. приложение С).

5.4 Распространение результатов

104. Одной из важнейших составляющих перехода к регистровой переписи является демонстрация способности готовить набор стандартных статистических данных переписного типа при соответствующих показателях качества до отказа от традиционной переписи. Наличие эффективной и совместимой системы регистров (включая статистические регистры) позволяет распространять на периодической основе переписные данные чаще, чем раз в десять лет. Как отмечалось ранее, становится возможным расчет ежегодных таблиц переписного типа, которые уже существуют в Нидерландах и ряде других стран.

5.5 Измерение/обеспечение качества

105. Кроме того, измерение и обеспечение качества должны осуществляться на всех этапах рассматриваемого подхода, хотя масштабы и методы проведения процессов измерения и обеспечения качества в разных странах будут разными.

106. В широком смысле эти процессы можно разделить на качество на входе (качество входных данных), качество процесса (изменения в качестве по мере добавления каждого процесса) и качество на выходе (качество результирующей статистики). Некоторые страны, возможно, пожелают добавить дополнительные процедуры измерения качества; так, например, в практическом примере Англии и Уэльса (см. приложение G) описывается расширение данного подхода с целью включения в него обследования охвата населения не только для измерения уровня охвата, но и для корректировки выходных данных на неполноту охвата. Страны, которые планируют переход к регистровой или комбинированной переписи, возможно, пожелают провести оценку качества новой модели по сравнению со старой для того, чтобы удостовериться в целесообразности перехода (см. практический пример Англии и Уэльса).

107. В остальных главах описываются процессы и методы, а также роль, которую играет обеспечение качества на каждом этапе.

6 Источники данных и их качество

108. Качество данных, используемых в процессе составления переписной статистики, сильно влияет на качество конечных статистических материалов. Таким образом, качество данных из административных регистров и других административных источников является одним из ключевых элементов, которые следует учитывать в процессе принятия решения об использовании административных данных для формирования статистики. Поэтому необходимо разработать и внедрить стандартный метод оценки качества административных данных в качестве потенциальных источников для переписи. Степень эффективной интеграции данных из различных административных источников является важным фактором при проведении такой оценки.

109. Качество административных данных, как правило, трудно измерить из-за их сложности и многомерности, и многие факторы, влияющие на качество, не поддаются измерению. Методология оценки такого качества опирается на многочисленные аспекты, критерии и показатели. Она может иметь форму контрольного перечня или обследования (в случае отсутствия других стандартов для проверки соответствия) и должна охватывать информацию об общих характеристиках регистра (таких, как уровень национального охвата, частота обновления и средства доступа), а также информацию о ведущихся переменных (такую, как определения и идентификаторы). Ключевые аспекты оценки качества регистра включают в себя наличие, своевременность, ясность, полезность, релевантность и непротиворечивость, а также стоимость доступа к регистру. Ключевыми аспектами оценки качества данных, содержащихся в регистре, являются точность и сопоставимость.

110. Оценка качества административных данных может использоваться для оценки пригодности источника данных. С этой целью необходимо разработать набор показателей, с помощью которых может быть проанализирован каждый источник. На основе этих показателей НСУ может принять решение об использовании конкретного источника. Кроме того, качество административных данных влияет на качество результатов переписи. Показатели источников данных могут быть интегрированы в систему качества, которая оценивает входную информацию (административные источники), качество процесса и качество продукта регистровой переписи. Проект ESS.VIP ADMIN¹⁷ направлен на разработку надлежащих методов обеспечения качества и облегчение доступа.

111. Такая информация об административных данных должна дать ответы на следующие вопросы:

- ↑ Как составляются данные и для чего?
- ↑ Предусмотрено ли их составление законном?
- ↑ Какова целевая группа населения?
- ↑ Регулярно ли обновляются данные?
- ↑ Проводятся ли проверки на достоверность?
- ↑ Как определяются переменные и сопоставимы ли эти определения со статистическими концепциями?

112. В качестве первого шага требуется получить информацию от соответствующего административного органа. Этого можно добиться путем изучения руководств, формуляров и вспомогательной документации. Рекомендуется также консультироваться с сотрудниками административного органа, ответственными за ведение источника. Для измерения заранее определенных параметров качества источников и метаданных следует использовать стандартный контрольный перечень или вопросник.

¹⁷ ESS.VIP ADMIN (проект по реализации концепции Европейской статистической системы в отношении административных источников данных) – это проект, рассчитанный на период 2015–2019 годов. Данный проект призван содействовать использованию административных источников в рамках ЕСС благодаря улучшению доступа к административным источникам, углублению методологических знаний, необходимых для интеграции административных данных в статистическое производство, и предоставлению инструментов для оценки качества материалов, опирающихся на административные источники. Он призван оказать государствам-членам поддержку в достижении этих теоретических результатов в конкретных областях статистики. Для получения дополнительной информации см. https://ec.europa.eu/eurostat/cros/content/essvip-admin-administrative-data-sources_en.

113. В качестве второго шага следует проанализировать микроданные из источника (общее количество записей, отсутствующие значения, нестандартные значения, дубликаты, количество записей без ключа). На данном этапе увязка записей из источника с данными обследования на уровне единиц (если такая увязка допустима и возможна) может дать ответы на дополнительные вопросы, например о сопоставимости переменных или о своевременности административной информации.

114. Таким образом, в дальнейшем можно будет изучать три аспекта качества:

↑ Качество источников, касающееся:

поставщика: доверие к держателю данных и его надежность (например, в плане пунктуальности предоставления данных), эффективность контактов и связи с поставщиком (например, являются ли они удовлетворительными/неудовлетворительными, периодическими/спорадическими);

релевантности: административная цель рассматриваемого атрибута (например, существует только юридическая обязанность или собственная заинтересованность в регистрации и хранении данных?);

конфиденциальности и безопасности: способ передачи данных, уровень дешифрования данных;

передачи: соглашение о передаче, юридическое обязательство, интервал периодической передачи данных (ежемесячно/ежеквартально/ежегодно/спорадически), дата передачи, расходы;

процедур: сбор и ведение данных.

↑ Качество метаданных, касающееся:

ясности: четкая структура и содержание данных и метаданных, хорошо организованные и точные метаданные, документирование изменений во времени;

обработки данных: описание процедур управления данными, проверки на непротиворечивость.

↑ Качество данных, охватывающее:

технические проверки: технические проверки входных данных, проверки формата данных (например, читаемости файла данных), соответствие данных определениям метаданных;

точность: степень точности, степень уверенности в достоверности записей данных (таких, как наличие неправдоподобных значений и непригодных записей);

полноту: определение регистра населения (с указанием уровней недостаточного или чрезмерного охвата), отсутствующие значения переменных;

временной аспект: процесс обновления, осуществляемый поставщиком, изменения в концепциях, определениях и охвате, независимо от того, установлена или нет дата пересмотра в целях обеспечения непрерывности и согласованности исторических рядов;

ясность: четкие структура и содержание данных;

целостность: степень, в которой источник данных может быть подвергнут интеграции или интегрирован в статистическую систему (на уровне единиц и по каждой переменной);

сопоставимость: определения данных, совместимые с функциями НСУ, уровень агрегирования, достаточный для статистических целей;

уникальные ключи: наличие уникального идентификатора на микроуровне, и, если существуют многочисленные источники, должна быть обеспечена возможность увязки (через единый идентификатор).

115. Система качества, как правило, может использоваться для изучения различных потенциальных источников и принятия решения о том, какие источники пригодны для использования в целях переписи. Инструменты, позволяющие систематическую оценку качества административных источников данных, были разработаны, например, в Нидерландах и Австрии¹⁸, о чем говорится во вставке б.1.

¹⁸ См., например, <http://www.cbs.nl/nl-NL/menu/methoden/onderzoek-methoden/discussionpapers/archief/2009/2009-42-x10-pub.htm> и http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/use_of_register/WP_16_Austria.pdf.

Вставка 6.1 Система качества в Нидерландах и Австрии

Нидерланды

Система качества может использоваться в качестве процедуры для систематического, объективного и стандартизированного определения качества источников данных. С этой целью Статистическое управление Нидерландов разработало систему качества, в которой различаются три различных ракурса качества, а именно: ракурс источника, ракурс метаданных и ракурс данных. В ракурсе источника основное внимание уделяется аспектам качества, необходимым для предоставления информации источником данных, в то время как в ракурсе метаданных основное внимание уделяется аспектам метаданных источника данных. В ракурсе данных изучаются технические аспекты качества данных и аспекты, связанные с их точностью. Система качества, разработанная в Нидерландах, использовалась в рамках многих различных проектов, включая национальную перепись населения 2011 года.

Австрия

Австрийская система оценки качества административных данных была разработана для регистровой переписи 2011 года. На каждом этапе обработки данных по каждому атрибуту рассчитывался показатель качества. Несмотря на то, что эта система была разработана для регистровой переписи, она рассчитана на общее применение. Благодаря модульной конструкции каждый этап системы может применяться индивидуально. Подход к оценке административных данных опирался на работу других НСУ (Daas, Ossen, Vis-Visschers, & Arends-Tóth, 2009) и основан на четырех гиперизмерениях, касающихся качества (документация, предварительная обработка, внешние источники и импутация), целью которых является измерение качества на трех этапах производства (исходные административные данные; объединенный набор данных, т.е. интеграция регистров; и окончательный набор данных, т.е. после редактирования и импутации данных).

На **уровне исходных данных** изучались три гиперизмерения: документация, предварительная обработка и внешние источники. **Документация** описывает процессы, связанные с качеством, а также документирование данных (метаданных) в административных органах. Степень доверия к держателю источника данных и его надежности отслеживалась с помощью вопросника, содержащего несколько открытых вопросов и вопросов со шкалой оценки. С помощью ответов на открытые вопросы собиралась информация, представляющая общий интерес, такая как своевременность предоставления данных. Вопросы со шкалой оценки измеряли, например:

- ↑ «историю данных» (например, хранятся или нет изменения с течением времени и доступна ли информация, относящаяся к контрольной дате переписи);
- ↑ «определения» (например, сопоставимы ли определения данных с определениями НСУ);
- ↑ «административная цель» (например, является ли признак релевантным для держателя источника данных и существует ли правовая основа для этого признака в административном источнике данных);
- ↑ «обработка данных» (например, как быстро изменения признака регистрируются в административном источнике данных, проверяются ли данные держателем источника данных, например путем запроса документов или удостоверений личности, и проводит ли держатель источника данных технические проверки и проверки согласованности между атрибутами).

Под **предварительной обработкой** понимается та часть записей данных, которая не может быть использована, например, без уникального идентификатора или без информации (отсутствие ответа по позиции), или когда значения выходят за пределы диапазона. **Внешний источник** сравнивает источник административных данных с другим источником, например Обследованием рабочей силы, путем сопоставления индивидуальных записей и расчета доли непротиворечивых наблюдений из расчета на переменную и административный источник данных. Путем объединения этих трех гиперизмерений был рассчитан показатель качества для каждой переменной в каждом административном источнике данных (например, по переменной «гражданство» из Центрального регистра населения, по переменной «гражданство» из Регистра занятости и т.д.).

116. Европейская статистическая система реализует комплексный проект по улучшению использования административных источников данных, известный как ESS.VIP ADMIN. Одно из направлений работы этого проекта состоит в разработке и продвижении показателей для оценки качества административных данных и статистических материалов, опирающихся на комбинацию источников, среди которых имеются и административные источники. Эти задачи охватываются проектом ESSnet по качеству статистики из множественных источников¹⁹, в котором участвуют восемь европейских НСУ. Основные цели заключаются в том, чтобы обобщить существующие знания по вопросам оценки и сообщения качества и провести их критический анализ; предоставить обновленные руководящие принципы оценки качества для целей статистического производства (входные данные, результаты и рамки для социальной статистики); разработать показатели для измерения качества выходных материалов, подготовленных на основе множественных источников, и методологию сообщения качества таких материалов; и разработать рекомендации в отношении обновления стандарта ЕСС и Руководства по отчетам о качестве ЕСС. В частности, в рамках проекта был проведен обзор текущей практики оценки качества административных источников, опробован ряд подходов и подготовлен рекомендуемый контрольный перечень для оценки качества входных данных (качество административных данных). Это, разумеется, имеет значение и для переписи, если планируется использовать административные источники. Дальнейшая работа в рамках проекта ADMIN будет включать в себя разработку руководящих принципов обеспечения качества в тех случаях, когда административные источники интегрируются в процесс статистического производства. Можно было бы обсудить конкретные примеры, касающиеся решения об использовании некоторых административных источников в целях переписи. Эту работу планируется завершить в 2019 году.

¹⁹ С текущей работой и промежуточными результатами ESSnet можно ознакомиться на портале CROS https://ec.europa.eu/eurostat/cros/content/essnet-quality-multisource-statistics-komuso_en.

7. Увязка и преобразование

117. Вполне вероятно, что будут наблюдаться значительные различия в качестве увязки между различными административными источниками данных с использованием и без использования уникальных ключей. Если во всех или большей части записей имеется уникальный идентификатор, увязка становится относительно простой, и уровень успешной увязки обычно высок (хотя качество переменных в статистическом регистре все еще предстоит измерить). Однако, если такой уникальный идентификатор отсутствует и не может быть построен, следует измерять качество увязки и оценивать его воздействие на результирующие материалы²⁰.

118. Например, из-за отсутствия уникальных идентификаторов в административных источниках данных Управление национальной статистики (УНС) Соединенного Королевства потратило ряд лет на разработку и совершенствование методов для преодоления этой проблемы и достижения прогресса в своей работе по изучению возможности перехода к переписи на основе административных данных. Используемый в настоящее время метод можно описать в два этапа:

- ↑ Детерминистический метод с использованием ключей сопоставления для увязки записей между административными источниками. Ключи сопоставления создаются путем объединения ключевых идентифицирующих переменных (или их частей), таких как имя, пол, дата рождения и почтовый индекс. Для каждого набора данных создается один и тот же набор ключей сопоставления. Если ключи сопоставления совпадают в каждом источнике, устанавливается связь.
- ↑ Вероятностный метод. Этот подход позволяет выявить связи между записями в двух наборах данных путем сравнения и количественной оценки относительного подобия записей (например, путем присвоения балльной оценки подобия). Основное отличие от детерминистической увязки состоит в том, что вероятностная увязка не требует, чтобы значения записей были идентичными между двумя записями («нестрогое сопоставление»)²¹.

119. На момент подготовки настоящих Руководящих принципов велась дальнейшая работа по совершенствованию этих методов, включая разработку статистической основы для решения проблемы множественных совпадений между тремя или более наборами данных²². Ссылки на доклады, представленные в этом и предыдущем пунктах, также содержат описание того, что было сделано к 2017 году для обеспечения качества при разработке этих методов увязки. Эти методы оказались многообещающими, однако для обеспечения высококачественной увязки, требуемой для получения надежных оценок численности населения, необходимо продолжить их совершенствование.

120. Значение термина «качество процесса» не вполне очевидно из его определения. Его элементы включают в себя:

- ↑ **Наилучшие методы** означают надежную методологию (включая надлежащие инструменты, процедуры и экспертные знания) и соответствующие статистические процедуры, применяемые на всех этапах – от сбора данных до их проверки на достоверность.
- ↑ **Затратоэффективность:** ресурсы используются эффективно.
- ↑ **Низкий уровень нагрузки на респондентов:** всякий раз, когда это возможно, следует использовать данные из административных источников (нагрузка на респондентов будет таким образом сведена к нулю). Однако, если сбор некоторых данных от респондентов является необходимым, то связанная с этим нагрузка должна быть пропорциональна потребностям пользователей и не должна быть чрезмерной для респондентов. НСУ контролирует нагрузку по предоставлению ответов и устанавливает цели по ее снижению с течением времени.

²⁰ Более подробную методологическую информацию см. в Руководстве по интеграции данных ГВУ-МОС ЕЭК ООН (<https://statswiki.unece.org/display/DI/Guide+to+Data+Integration+for+Official+Statistics>).

²¹ Более подробную информацию об этих методах можно найти по адресу <http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011-matching-anonymous-data--m9-.pdf>.

²² С более подробной информацией об этих методах можно ознакомиться по адресу <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/methodology/methodologyofstatisticalpopulationdatasetv20>.

8. Статистические регистры

121. Использование интегрированных данных из различных источников в целях переписи населения и жилищного фонда становится все более распространенным подходом. В странах с административными регистрами, использующими уникальные идентификаторы в отношении единичных записей, можно проводить различие между базовыми и дополнительными регистрами. Как отмечалось и было определено в разделе 5.3, ключевые регистры, позволяющие получить статистическое описание единиц в целях переписи (лиц, домохозяйств и жилищ), включают в себя регистры населения, регистры жилья (или жилищные или адресные регистры), регистры предприятий и регистры активности. Увязка информации из данных, имеющихся на уровне единичных записей, позволяет определить эти характеристики населения. Однако для целей переписи может потребоваться интеграция данных не только административных регистров, но и текущих исследований, касающихся конкретной категории населения и ее свойств (таких, как продольное исследование переписной выборки), результатов предыдущих переписей и информации, накопленной в рамках статистической основы выборки. Интегрированный набор данных из этих источников принимает форму статистического регистра (базы данных, которая может быть использована в дальнейшем процессе сбора и компиляции данных) для целей переписи.

122. Как отмечалось в разделе 5.1, УНС и ЦСУ Польши создали набор данных статистики народонаселения (НДСН) путем увязки административных источников данных. В частности, УНС произвела увязку четырех административных источников данных:

- ↑ пациенты, зарегистрированные у врача общей практики;
- ↑ лица, имеющие национальный страховой номер;
- ↑ студенты, зарегистрированные на курсе высшего образования;
- ↑ ученики государственных школ.

123. После увязки этих источников данных для формирования НДСН применяется ряд правил с целью принятия решения о том, какие данные относятся к обычно проживающему населению. Дополнительные данные о «активности» (или «признаках жизни») используются для определения местонахождения обычного проживающего лица, описываемого записью, в тех случаях, когда наблюдаются противоречия между различными административными источниками данных²³. Затем НДСН используется для расчета оценок численности населения. Для понимания качества этих демографических оценок, основанных на административных данных, проводятся сопоставления с оценками переписи и официальными демографическими оценками. Эти сопоставления демонстрируют обнадеживающие результаты²⁴.

²³ Более подробную информацию о методах построения НДСН УНС можно найти по адресу <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/methodology/methodologyofstatisticalpopulationdatasetv20>.

²⁴ См. <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/sizeofthepopulation>.

9. Качество на выходе

124. Материалы переписи предназначены для широкого круга пользователей, и, как отмечается в главе 4, для успешного перехода необходимо одобрение таких результатов как заинтересованными субъектами, так и широкой общественностью. Использование административных данных может дать возможность расширить диапазон и увеличить периодичность выпуска статистических материалов. Оно также может облегчить продольные исследования. Однако в результате этого могут быть внесены некоторые изменения как в определения переменных, так и в классификации выходных данных. Воздействие этих изменений на статистические материалы должно быть изучено и разъяснено пользователям.

125. Независимо от методологии сбора данных оценка качества выходных данных переписи всегда была важной и необходимой задачей. Существует несколько различных способов и методов оценки качества статистики, включая качество материалов переписи. Оценка качества переписи, проводимой с использованием новой методологии, имеет особо важное значение, поскольку она позволяет получить релевантную информацию о достоверности результатов новой переписи и о том, чем это качество может отличаться от результатов предыдущих переписей. Некоторые из этих различных подходов к оценке качества описаны в данной главе. Всегда полезно использовать несколько методов.

126. Качество результатов любой переписи может определяться путем оценки качества продукта, исследований охвата, в отчетах о качестве и комитетами по качеству. При рассмотрении аспектов качества и конфиденциальности вопрос заключается в том, каким образом можно измерить качество выходных данных. Например, при использовании выборочных обследований следует определить минимальную частоту попадания в группу, чтобы определить, являются ли полученные оценки достаточно точными для публикации. С другой стороны, в тех случаях, когда страна полагается только на регистры или сплошную регистрацию переписчиками, следует установить правила конфиденциальности для предотвращения раскрытия личной информации. В контексте переписи отчеты о качестве (в которых результаты могут сравниваться с другими выходными данными, например обследования рабочей силы) обычно показывают, что различия в случае демографических переменных относительно невелики, а различия в случае экономических переменных (таких, как текущий статус активности) могут быть относительно большими. Однако следует отметить, что, поскольку переписи чаще всего проводятся раз в десять лет, оценки качества переписи, публикуемые в отчетах о качестве, могут иногда устаревать ко времени проведения следующей переписи, и тогда они могут не иметь конкретного отношения к какой-либо новой методологии переписи.

127. Как отмечалось в предыдущей главе, УНС сравнивает свои демографические оценки, основанные на административных данных, с переписными оценками и официальными демографическими оценками и использует набор стандартов качества для лучшего понимания их качества²⁵. В 2017 году УНС разработало метод для независимой оценки качества наборов данных статистики народонаселения на исследовательском этапе этой работы (например, метод определения доверительного интервала для оценок НДСН). В долгосрочной перспективе УНС планирует проводить ежегодное обследование охвата населения, которое будет использоваться для измерения и корректировки погрешностей охвата в НДСН (по аналогии с его обследованием охвата переписи)²⁶. Это будет подкрепляться другими процессами обеспечения качества, например демографическим анализом и сопоставлениями с другими источниками данных.

9.1 Качество продукта

128. Качество официальной статистики можно описать с точки зрения качества продукта и качества процесса. На основе этих двух видов показателей на уровне Европейской статистической системы (ЕСС) и Статистической комиссии Организации Объединенных Наций был разработан широкий спектр показателей

²⁵ См. приложение А в <http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011-options-report-2--o2-.pdf>.

²⁶ Более подробная информация об этой работе представлена по адресу <http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011--producing-population-estimates-using-administrative-data-in-theory--m9-.pdf>.

качества, и некоторые из этих показателей должны использоваться для составления отчетов о качестве в рамках ЕСС. В нижеследующих пунктах содержится описание качества продукта в этом контексте.

129. Качество продукта включает в себя:

↑ **Релевантность** означает степень, в которой статистические данные отвечают текущим и потенциальным потребностям пользователей. Речь идет о том, все ли необходимые статистические данные на самом деле были сформированы и в какой степени используемые концепции (включая определения и классификации) отражают потребности пользователей. Тремя наиболее распространенными подходами к оценке релевантности являются:

- выяснение этого у пользователей непосредственно в ходе обследования потребностей пользователей (до переписи) или в ходе обследования удовлетворенности пользователей (после переписи);
- создание механизмов обратной связи с заинтересованными субъектами и клиентами;
- анализ использования данных с помощью исследования использования результатов переписи.

Степень соблюдения международных рекомендаций может служить определенным показателем релевантности на глобальном уровне.

↑ **Точность** означает близость оценок к неизвестным «истинным» значениям. Ее основными компонентами являются:

- оценка охвата;
- отсутствие ответов по единицам и позициям и методы, используемые для работы с неверными значениями (редактирование) и отсутствующими значениями (импутация);
- ошибки выборки.

Как правило, руководства по оценке качества содержат набор показателей для измерения точности.

↑ **Своевременность** означает период между появлением информации и описываемым ею событием или явлением. Общие показатели – это временные интервалы между окончанием учетного периода переписи, с одной стороны, и датами первого опубликования данных и публикацией окончательных результатов, с другой стороны. Обычно существует компромисс между точностью и своевременностью, в котором предпочтение отдается как можно скорейшему получению результатов в ущерб точности. Однако часто бывает так, что разные пользователи будут иметь разные взгляды на баланс между этими двумя требованиями.

↑ Под **пунктуальностью** понимается временной разрыв между датой опубликования результатов и намеченной датой (датой, к которой данные должны были быть предоставлены в соответствии с каким-то официальным графиком выпуска, или датой, установленной законодательством или предварительно согласованной с пользователями).

↑ **Доступность** и **ясность** означают условия и способы получения, использования и интерпретации (интерпретируемость) данных пользователями, определяемые:

- диапазоном доступных публикаций (или таблиц);
- уровнем доступа к базам данных;
- наличием соответствующих метаданных, включая информацию о качестве данных.

↑ **Сопоставимость** означает степень сопоставимости статистических данных между географическими районами и во времени (между переписями).

↑ **Согласованность** означает степень, в которой данные переписи могут надежно комбинироваться различными способами и для разных видов использования со статистической информацией из других источников в более широком контексте. Когда статистические данные берутся из различных источников, и в частности из статистических обследований, проводимых с использованием разных методологий, они зачастую не являются полностью идентичными и вызывают расхождения в результатах из-за различий в определениях, классификациях и методологических стандартах. Существует несколько областей, в которых оценка согласованности проводится на регулярной основе:

- между предварительной и окончательной статистикой;
- между ежегодной и краткосрочной статистикой;
- между статистикой из одной и той же социально-экономической области;
- между статистикой обследований и национальных счетов.

130. В австрийской системе, о которой говорится во вставке 6.1 выше, описывается методика расчета качества всех атрибутов, представляющих интерес для национальной регистровой переписи (см. вставку 9.1).

Вставка 9.1 Австрийская система обеспечения качества

На этапе обработки вся информация из регистров объединяется в **центральную базу данных**, которая охватывает все атрибуты, представляющие интерес для регистровой переписи. На этом уровне рассчитывается показатель качества для каждого атрибута по всем административным источникам данных. Если переменная получена только из одного административного источника данных, то качество этого атрибута на уровне входных данных будет таким же, что и в центральной базе данных. Если для выведения значения переменной (например, текущий статус активности) или определения наиболее вероятного значения (например, при наличии противоречивой информации о дате рождения в двух или более административных источниках данных) объединяются несколько административных источников данных, тогда рассчитывается показатель качества. Это производится с использованием теории Демпстера-Шафера для объединения показателей качества из различных источников данных. Кроме того, проводится сопоставление с внешним источником (например, обследованием рабочей силы). На последнем этапе обработки данных методом импутации рассчитываются значения, отсутствующие в центральной базе данных. Для оценки качества данных в **окончательном наборе данных** рассчитывается показатель качества по **импутации**.

131. О качестве любой переписи судят по качеству ее результатов. Общие требования к результатам переписи зафиксированы в руководящих принципах качества, которые являются общими для всех типов переписей. Следует учитывать три уровня качества:

- ↑ качество отдельных переписных переменных с точки зрения всех критериев качества;
- ↑ качество гиперкубов (многомерные переписные таблицы, как правило с четырьмя или более измерениями) оценивается с использованием показателей качества переписных переменных;
- ↑ качество переписи населения (с точки зрения охвата).

132. Каждый из них рассматривается в нижеследующих пунктах и разделе 9.2.

9.1.1 Качество отдельной переписной переменной

133. Качество отдельной переписной переменной зависит от:

- ↑ охвата переменной(ых) административных данных, используемой(ых) для расчета переписной переменной;
- ↑ точности переменной(ых) административных данных, используемой(ых) для расчета переписной переменной;
- ↑ адекватности программного обеспечения, используемого для расчета переписной переменной;
- ↑ технических или человеческих ошибок.

134. Если контекстуальные проверки переменной проводятся путем сопоставления с другими источниками, тогда измеряются следующие характеристики качества:

- ↑ количество отсутствующих значений;
- ↑ количество ошибок или резко отклоняющихся значений.

135. Как правило, поскольку в отношении резко отклоняющихся значений можно использовать подход, аналогичный подходу к отсутствующим значениям, то внимание следует сосредоточить на отсутствующих значениях. При определении стандартов качества целесообразно определить максимальную долю отсутствующих значений, которая позволяет обеспечить отличное, хорошее и удовлетворительное качество переписной переменной. Эти стандарты носят исключительно субъективный характер, поскольку не указывается соответствующий уровень каждого измерения качества.

9.1.2 Качество переписного гиперкуба

136. Качество переписного гиперкуба зависит от:

- ↑ качества всех соответствующих переписных переменных;
- ↑ адекватности программного обеспечения для генерирования кубов.

137. Общее правило заключается в том, что если гиперкуб содержит k переменных и эти переменные имеют соответственно недостающие значения m_1, m_2, \dots, m_k , то количество недостающих значений M в k -мерном гиперкубе определяется следующими неравенствами:

$$\max(m_1, m_2, \dots, m_k) \leq M \leq m_1 + m_2 + \dots + m_k$$

138. Это означает, что число отсутствующих значений в гиперкубе, как правило, больше по сравнению с любой отдельной переменной. А также, что чем выше размерность гиперкуба, тем больше, как правило, недостающих значений он содержит. Должны быть определены разумные показатели качества для гиперкубов. Одна из возможностей состоит в использовании следующего показателя:

$$M = \left\{ \sum m_{(i)} \right\} / k$$

139. Где $m_{(i)}$ для $i = 1, 2, \dots, k$ – показатели качества переписных переменных, включенных в гиперкуб с k измерениями, а M – показатель качества гиперкуба. Также желательно добавить дополнительное условие: показатель качества гиперкуба не может иметь высокое значение, если хотя бы один из $m_{(i)}$ имеет очень низкое значение.

140. Как отмечалось выше, при определении стандартов качества целесообразно определить максимальную долю отсутствующих значений, которая позволяет оценить качество переписного гиперкуба как отличное, хорошее или удовлетворительное. Уровни качества отдельных переменных и самого гиперкуба должны быть непротиворечивыми, т.е. если все переменные имеют отличное качество, то и гиперкуб также имеет отличное качество. И наоборот, гиперкуб, имеющий некоторые переменные низкого качества, не может иметь хорошего или отличного рейтинга качества.

141. Любой отчет о качестве переписи должен также описывать широкую методологию статистического контроля за раскрытием (СКР) и преобразований в целях защиты данных, применяемых НСУ для защиты конфиденциальности. Следует сообщать о последствиях применения таких приложений для качества данных (например, об уровне потери точности). Тем не менее страны должны опасаться предоставлять слишком много подробной информации о методологии СКР, поскольку это само по себе создает риск раскрытия информации.

9.2 Охват

142. Обеспечение полного охвата населения является одной из основных задач при проведении любой переписи. В случае традиционной переписи наиболее распространенной проблемой является недостаточный охват, особенно в настоящее время, когда люди, как правило, более мобильны и могут иметь более одного адреса проживания, что затрудняет их регистрацию. Другая проблема заключается в том, что некоторые люди очень высоко ценят неприкосновенность своей частной жизни и могут предпочесть не предоставлять свои данные через переписчиков. Кроме того, некоторые группы населения, такие как нелегальные или незарегистрированные мигранты, могут стремиться остаться вне поля зрения государственных органов. Нежелание светиться в официальной отчетности может также

быть причиной того, что некоторые люди не будут иметь записей в административных регистрах, что приведет к аналогичному неполному охвату регистровых и комбинированных переписей.

143. В случае регистровой переписи также возникает проблема чрезмерного охвата регистров, которую необходимо учитывать. Например, если люди официально не заявили о своей эмиграции, их данные могут храниться в административных регистрах без изменений и, следовательно, являться причиной чрезмерного охвата в рамках переписи. Если регистры являются недостаточно качественными в этом отношении, то одним из возможных способов избежать ошибок охвата является создание индекса резидентства на основе записей, хранящихся в нескольких регистрах, для определения так называемой бальной оценки «признаков жизни». Подход заключается в том, чтобы определить в отношении всех возможных жителей «признаки жизни» в виде двоичного показателя (со значением 0 или 1) для каждой записи в каждом регистре. Используя эти признаки жизни в качестве объясняющих переменных, можно построить модель прогнозирования размера неполного или избыточного охвата²⁷. Для более точного определения численности населения бальную оценку следует пересчитывать ежегодно.

144. Существует несколько вариантов проверки ошибок охвата в случае комбинированной переписи. Один из них заключается в использовании методологии, широко применяемой в отношении традиционных переписей (для компонента регистрации переписчиками), путем организации послепереписного обследования и статистической оценки как избыточного, так и неполного охвата. Подход «признаков жизни», заключающийся в сопоставлении записей из различных регистров, может быть затем использован для получения информации об уровне неполного и чрезмерного охвата регистрового компонента переписи.

145. Как отмечалось ранее, УНС планирует проводить ежегодное обследование охвата населения, результаты которого будут использоваться для измерения и корректировки ошибок охвата в НДСН (по аналогии с его традиционным обследованием охвата переписи). На момент подготовки настоящих Руководящих принципов (в 2017 году) УНС приступила к тестированию такого обследования с целью проведения полномасштабного пилотного обследования в 2020 году²⁸. Это будет подкрепляться другими процессами обеспечения качества, например демографическим анализом и сопоставлениями с другими источниками данных.

146. Система двойственной оценки (СДО) используется Центральным статистическим управлением Ирландии для изучения вопросов охвата демографических оценок, рассчитанных на основе административных данных. Численность населения оценивается путем применения методов двойного ввода к двум независимым спискам жителей. Первый список опирается на ежегодный регистр «признаков жизни» (ПЖ), составленный на основе одиннадцати административных источников данных и демографической статистики, который призван регистрировать активность населения всех возрастов. Более подробную информацию об этом методе можно найти в презентации «Составление демографических оценок на основе административных данных», представленной на пятом Семинаре по административным данным в 2016 году²⁹.

147. Исследования этого метода включают в себя:

- ↑ изучение основы регистра ПЖ;
 - включение/исключение различных источников данных;
 - включение/исключение лиц по уровню активности, например исключение, если количество отработанных недель составляет менее 20 и соответствующее лицо не фигурирует в других источниках данных;
- ↑ изучение потенциальных источников второго списка;
 - второй список, основанный на регистре (например, продление водительских прав в случае взрослых);

²⁷ См., например, Tiit, E.-M. (2012) Estimated undercoverage of census 2011. Quarterly Bulletin of Statistics Estonia, 4, 12, 110-119 and Tiit, E.-M., Meres, K., Vähi, M. (2012) Estimation of census population of census 2011. Quarterly Bulletin of Statistics Estonia, 3, 12, 79-108.

²⁸ Более подробная информация об этой работе представлена по адресу <http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011--producing-population-estimates-using-administrative-data-in-theory--m9-.pdf>.

²⁹ См. <http://www.cso.ie/en/csostatnews/eventsconferencesseminars/administrativedataseminars/5thadministrativedataseminar/>.

- второй список, основанный на обследовании (например, ежеквартальном обследовании домашних хозяйств, которое, вероятно, продолжится в будущем).

148. Эстонская оценка качества охвата переписи описывается в качестве примера во вставке 9.2.

Вставка 9.2 Оценка качества переписи с использованием административных баз данных в Эстонии

Проблема

Главной характеристикой качества переписи является **охват**, т.е. соотношение c/n численности населения c согласно переписи и фактической численности населения n . Если это соотношение больше 1, то перепись имеет чистый избыточный охват, что означает, что были зарегистрированы некоторые лица, не принадлежащие к фактическому населению, и их численность превышает число лиц, не прошедших перепись. Причина может заключаться в том, что некоторые люди были переписаны несколько раз. Если соотношение c/n меньше 1, то перепись страдает чистым неполным охватом, что означает, что часть фактического населения не была зарегистрирована и его численность превышает число тех, кто был ошибочно зарегистрирован.

В настоящее время **неполный охват** обусловлен несколькими причинами: люди очень мобильны, и может случиться так, что некоторых из них в момент переписи не будет дома, а некоторые настолько высоко ценят неприкосновенность своей частной жизни, что не захотят раскрывать свою личную информацию и не примут участия в переписи.

Если все проживающие в стране люди имеют идентификационные коды (ИН), то **избыточного охвата**, вызванного дублированием, можно легко избежать.

Традиционная методология

Оценка охвата переписи долгое время являлась серьезной задачей для статистиков. Одним из возможных решений является **сопоставление переписного населения с текущей демографической статистикой**. Этот метод является удобным в те периоды, когда население не подвержено каким-либо крупным изменениям в естественном или миграционном приросте, но требует высокого уровня соблюдения требований к регистрации актов гражданского состояния. Этот метод применялся в Эстонии несколько раз, и в 1934 и 1970 годах были достигнуты очень хорошие результаты.

Другой метод основан на **проверочном обследовании**. При использовании этой методологии независимое выборочное обследование проводится вскоре после завершения переписи с использованием некоторых вопросов переписи. Таким образом, число лиц $n(1)$, зарегистрированных в рамках как переписи, так и обследования, число лиц $n(2)$, зарегистрированных в рамках переписи, но не обследования, число лиц $n(3)$ зарегистрированных в рамках обследования, но не переписи, и число лиц $n(4)$, не зарегистрированных в рамках ни переписи, ни обследования, можно использовать для оценки доли лиц, не зарегистрированных в рамках переписи, хотя и входящих в состав населения. Хотя этот метод довольно популярен, он имеет несколько проблем – в нем используется предположение о том, что неохват какого-либо лица переписью или обследованием является случайным. Если существует несколько неслучайных причин недоучета в рамках переписи или обследования, тогда этот метод может не дать достаточно точного ответа.

Новая методология на основе существующих наборов административных данных

Если в стране имеется набор (система) надежных регистров или других наборов административных данных, то их можно использовать для проверки ошибок охвата переписи.

Проверка неполного охвата с использованием признаков жизни.

С этой целью Статистическое управление Эстонии изучает совокупность всех лиц, которые могут проживать в стране (быть резидентами). К этой совокупности (**суперсовокупности**) относятся следующие группы: все зарегистрированные в ходе переписи лица, все лица, проживавшие в

стране на момент последней переписи (или ранее) и впоследствии эмигрировавшие, и все лица, внесенные в списки некоторых административных регистров страны.

В отношении всех людей из суперсовокупности их **признаки жизни** можно найти следующим образом. Предположим, в стране существует K различных регистров. Если лицо j было активным хотя бы раз в течение последнего года (имеет запись) в регистре i , то у него имеется признак жизни $E(i,j)$ со значением 1; в противном случае оно имело бы $E(i,j) = 0$. Признаки жизни, определенные таким образом, являются бинарными переменными, характеризующими всех лиц из суперсовокупности.

Следующим шагом является создание **тестовых совокупностей**, состоящих из «**несомненных резидентов**» и «**несомненных нерезидентов**». Совокупность несомненных резидентов состоит из лиц, зарегистрированных в ходе переписи в качестве постоянных жителей Эстонии, которые также были постоянными жителями Эстонии по данным регистра населения. Несомненными нерезидентами являются лица, не зарегистрированные в качестве постоянных жителей Эстонии и фигурирующие в эстонском регистре населения в качестве нерезидентов. Для оценки вероятности постоянного проживания использовались тестовые совокупности и была создана логнормальная модель с использованием ПЖ в качестве объясняющих переменных. Модель использовалась в отношении всех лиц, принадлежащих к суперсовокупностям и не охваченных переписью. Некоторые лица, не охваченные переписью, имели высокую вероятность быть постоянными жителями (эти лица имели много ПЖ, свидетельствующих об их активности в Эстонии в течение последнего года). Не охваченные переписью лица, которые, согласно модели, имели высокую вероятность быть постоянными жителями страны, рассматривались в качестве недоучета и добавлялись к переписному населению при подсчете численности в будущем. Доля недоучета, согласно оценке, составила около 2%, а расчетная погрешность модели – менее 5% от соответствующей совокупности лиц.

Аналогичным образом, в переписном населении можно найти лиц, которые не принадлежат к фактическому населению и, таким образом, представляют собой избыточный учет.

9.3 Качество и конфиденциальность

149. Обеспечение качества данных и защита конфиденциальности персональной информации в некотором смысле противоречат друг другу, но и то, и другое являются важными аспектами результатов переписи. Важно понимать, что трудно, если не невозможно, опубликовать полностью точные результаты (особенно по малым районам), одновременно установив достаточные уровни контроля за раскрытием для обеспечения защиты конфиденциальности. Следует отметить, что другие публикации, в частности основанные на тех же административных данных, что и перепись, должны приниматься во внимание при установлении этих достаточных уровней.

9.4 Сопоставление результатов переписи с результатами обследований

150. Общепринятым методом оценки точности данных переписи является сопоставление полученных результатов с данными общенациональных обследований (и регистров в случае, если они использовались для других расчетов, помимо переписи). Этот подход является особенно полезным в случае регистровой переписи.

151. Как правило, чем больше размер выборки обследования, тем более надежным будет результат проверки. В этом контексте ОРС является хорошим выбором для оценки качества переписи. В то же время имеется ряд других вопросов, которые следует принимать во внимание; например, определения, используемые в двух источниках данных, и учетные даты могут быть различными. Кроме того, поскольку никакие обследования не могут охватить весь спектр переписных переменных, лишь только некоторые из них могут быть оценены таким образом. И последнее, но не менее важное замечание касается того, что при наличии различий не всегда будет возможно продемонстрировать, какие данные являются более точными.

152. Несмотря на все эти трудности, сопоставление результатов переписи с результатами обследований имеет большое значение и должно, по возможности, проводиться во всех случаях. Затем результаты должны быть тщательно проанализированы, чтобы объяснить причины обнаруженных расхождений.

9.5 Отчеты о качестве

153. Отчет о качестве – это документ, содержащий информацию обо всех этапах проверки качества: сопоставление и анализ результатов, включая объяснение любых расхождений между данными переписи и обследования. Если существуют объяснения тому, почему результаты регистровой переписи (или части комбинированной переписи, в рамках которой переменные выводятся из административных данных) отличаются от результатов обследований, то может возникнуть необходимость в дополнительных шагах:

- ↑ качество статистического регистра следует проверить с использованием альтернативных методов;
- ↑ качество обследования следует проверить путем сопоставления с другими обследованиями;
- ↑ следует проверить программное обеспечение (включая алгоритмы), используемое для создания переменных переписи на основе административных данных;
- ↑ следует проверить, не были ли эти расхождения вызваны применяемыми мерами обеспечения конфиденциальности;
- ↑ следует проверить возможность ошибок во всех процедурах.

154. Результаты всех таких проверок должны быть отражены в отчете о качестве, который затем может быть использован в качестве исходного материала для следующей переписи с оговоркой, сделанной в начале главы 9, о том, что оценки качества переписи, опубликованные в отчетах о качестве, иногда могут устаревать к моменту проведения следующей переписи.

9.6 Группы по обзору качества

155. Проверки качества регистровой или комбинированной переписи должны осуществляться на нескольких уровнях, которые могут включать в себя следующие:

- ↑ **внутренние комитеты по качеству**, сформированные НСУ и состоящие из членов группы по проведению переписи или других лиц, которые работали с группой по проведению переписи. Такой комитет будет проверять качество всех переменных и таблиц и может нести ответственность за подготовку отчета о качестве;
- ↑ **аудиторские проверки, основанные на руководящих принципах качества**, которые могут проводиться экспертами НСУ, не входящими в состав группы по проведению переписи;
- ↑ **внешние аудиторские проверки (для проверки конфиденциальности или экспертные оценки)**, проводимые независимыми экспертами, не входящими в состав НСУ, возможно, из академического сообщества или даже из другой страны.

156. Наконец, следует еще раз упомянуть проект ESS.VIP ADMIN, о котором говорится в главе 6. Цель этого проекта заключается в том, чтобы помочь статистикам более широко и эффективно использовать административные источники при подготовке официальной статистики. Это достигается путем решения наиболее типичных проблем, возникающих при использовании этих источников: ограниченный доступ к данным, недостаточное качество источников, методологические вопросы, связанные с обработкой данных и интеграцией нескольких источников. Он также направлен на обеспечение того, чтобы статистические данные, подготовленные с использованием административных данных, были сопоставимыми и обладали достаточным качеством благодаря предоставлению инструментов для оценки качества результатов.

10 Подходы и практические примеры разных стран

157. В настоящем разделе представлен общий обзор подходов, применяемых странами для преодоления некоторых трудностей при переходе к регистровой или комбинированной переписи. Некоторые практические примеры, описывающие опыт конкретных стран, приводятся в приложениях А-I.

10.1 Технические подходы

158. Хотя общий подход был описан в главе 5, каждая страна должна разработать свою собственную методику его реализации. НСУ полезно обмениваться опытом с другими странами, но часто бывает невозможно для конкретного НСУ просто скопировать практику другой страны, если оно учитывает, как это и должно быть, национальный контекст. Некоторые из технических подходов, использовавшихся в разных странах, кратко описываются в этой главе.

159. Как уже отмечалось, целый ряд стран уже перешли к регистровой переписи. Раннему переходу на эту методологию сбора данных лишь способствовало наличие уже существующих регистров населения. В качестве примера можно привести Данию и Финляндию. Другие страны осуществили этот переход позднее. В 2011 году Австрия провела регистровую перепись с использованием сочетания регистров населения и других административных источников. Эстония и Польша провели в 2011 году комбинированную перепись, включавшую самоисчисление в режиме онлайн. Другие страны, такие как Соединенное Королевство и Канада, работают над переходом к системе комбинированной переписи в долгосрочной перспективе, но все еще сталкиваются с рядом проблем, которые необходимо преодолеть.

160. Страны могут создать предохранительную систему для мониторинга качества регистра, методологии и технологического потенциала на переходном этапе. Например, в настоящее время в Англии и Уэльсе (на момент подготовки настоящих Руководящих принципов) ведется разработка системы анализа по принципу светофора для отслеживания хода работы, а в Эстонии установлены требования к качеству для держателей регистров.

161. Многие аспекты будут в значительной степени зависеть от внешних факторов, таких как правовая основа и одобрение заинтересованными субъектами, однако способность преобразовывать множественные административные источники данных в статистический регистр часто является основной внутренней проблемой для НСУ.

162. Там, где статистические регистры уже используются, число задействуемых административных источников и базовых регистров варьируется в зависимости от страны. Австрия, например, использовала восемь базовых регистров в целях переписи 2011 года и другие регистры – для оценки качества. Эстония объединила 22 источника для разработки своих статистических регистров, а Польша объединила целых 28 источников административных данных и три неправительственных источника (из примерно 300 потенциальных источников данных, которые первоначально были подвергнуты оценке) в статистический регистр под названием «Главный регистр».

163. Интеграция административных источников данных означает увязку информации о людях и жилищах для обеспечения всеобъемлющего охвата населения. Ключевое различие между странами заключается в том, имеются ли в наличии уникальные идентификаторы, которые позволяют проводить детерминированное сопоставление. Основная проблема в некоторых странах заключается в увязке записей данных из различных регистров без уникальных идентификаторов. В таких случаях можно попытаться провести вероятностное или нестрогое сопоставление по нескольким полям.

164. Как правило, при их использовании в целях переписи административные записи проходят процесс обезличивания для защиты персональных данных лиц, в ходе которого прямые идентификаторы, такие как

персональный идентификационный номер (ПИН) и косвенные идентификаторы, удаляются из регистра и заменяются суррогатными значениями. Примером уникального идентификатора является персональный номер Службы государственного Обеспечения (PPSN), используемый в Ирландии. ЦСУ использует защищенный идентификационный ключ, увязанный с PPSN сторонней организацией. Австрия, которая не имеет уникального персонального идентификатора, разработала способ обойти эту проблему, используя для различных административных подразделений отраслевые персональные идентификационные номера (bPINs). Эти bPIN могут быть увязаны с центральным регистром населения. В этом случае увязка осуществляется Австрийским органом по защите данных (см. приложение D). В дополнение к идентификации лиц применяется фильтр «признаков жизни» для выяснения того, является ли соответствующее лицо обычным резидентом или проживало в стране в какой-то конкретный момент времени.

165. Хотя статистика малых районов может быть рассчитана на основе географического района, в котором проживают лица (или домохозяйства), для расчета переменных, касающихся домохозяйств, должны быть идентифицированы жилища. В некоторых странах идентификация и определение местоположения жилищ могут осуществляться по уникальным идентификаторам зданий, таким как почтовый индекс в Ирландии и уникальный идентификационный номер объекта недвижимого имущества (UPRN) в Великобритании (Англия, Уэльс и Шотландия). В некоторых других странах были созданы новые регистры зданий и жилищ. Без уникального идентификатора зданий увязка может производиться путем автоматического сопоставления адресов. Для этих целей имеется специальное коммерческое программное обеспечение, хотя некоторые страны (например, Канада) разработали свою собственную внутреннюю систему. Сопоставление адресов – это сложный основанный на правилах процесс, который нуждается в частом обновлении. Адрес должен сопоставляться с адресами в других регистрах, затем с адресами лиц, а затем географически кодироваться по местоположению. В Эстонии, например, был введен «стандарт адреса», чтобы облегчить подобное сопоставление адресов. Следует разработать правила для разрешения проблем, связанных с любой противоречивой или двусмысленной информацией, такой как увязка лица с более чем одним адресом или информация о многоквартирных домах.

166. Необходимо разработать и тщательно протестировать систему контроля качества и стандарты, которые будут поддерживать процесс после того, как результаты традиционной переписи станут недоступны в качестве базы сравнения. Обследования для оценки качества и импутация или моделирование для получения надежных результатов являются дополнительными необходимыми шагами, которые обычно связаны с преобразованием административных данных в статистический регистр.

10.2 Опыт конкретных стран

167. Для иллюстрации того, каким образом может осуществляться переход и какие практические проблемы странам, возможно, придется преодолевать, в приложениях A-I более подробно приводятся некоторые примеры опыта конкретных стран. Эти примеры основаны на опыте стран, которые недавно отказались от традиционной переписи или планируют сделать это в будущем. Однако следует отметить, что эти практические примеры отражают позицию, которая была актуальна в каждой стране на момент подготовки настоящих Руководящих принципов (середина 2017 года). С тех пор, возможно, произошли изменения в методологии и процессах. Тем не менее другие страны могут извлечь пользу из этого опыта.

- ↑ Практический пример Ирландии (см. приложение A)
- ↑ Практический пример Эстонии (см. приложение B)
- ↑ Практический пример Польши (см. приложение C)
- ↑ Практический пример Австрии (см. приложение D)
- ↑ Практический пример Словении (см. приложение E)
- ↑ Практический пример Португалии (см. приложение F)
- ↑ Практический пример Англии и Уэльса (см. приложение G)
- ↑ Практический пример Италии (см. приложение H)
- ↑ Практический пример Германии (см. приложение I)

Приложение А – Практический пример Ирландии

Настоящая записка опирается на документ The Irish Statistical System and The Emerging Census Opportunity (Dunne, 2015), представленный на Конференции по новым методам и технологиям в области статистики (NTTS), состоявшейся в Брюсселе в 2015 году (Dunne, 2015).

Общая обстановка

Центральное статистическое управление (ЦСУ) Ирландии добилось значительного прогресса в использовании административных источников данных в статистических целях (MacFeely and Dunne, 2014). Эта работа подкрепляется развитием национальной инфраструктуры данных (НИС) в Ирландии. Вкратце, НИС представляет собой концептуальную основу, способствующую более эффективному использованию данных в государственном секторе Ирландии. Она поощряет и стимулирует использование единых официальных идентификаторов лиц, предприятий и объектов недвижимости во всех официальных системах (по скандинавской модели). Создание центрального регистра населения с обновленной информацией о том, где кто проживает, не предусмотрено.

Существует личный идентификационный номер, который обычно используется во всех административных системах, основанных на лицах (школы, социальное обеспечение, занятость и т.д.), который присваивается при рождении или при въезде человека в страну для проживания. Этот идентификационный номер хранится в главном файле или административном регистре. Этот регистр ведется и обновляется Департаментом социальной защиты (ДСП – правительственное ведомство, ведающее вопросами социального обеспечения); однако отсутствует требование о том, чтобы все организации государственного сектора предоставляли обновленную информацию ДСП.

Недавно был осуществлен проект по обобщению данных об активности каждого лица за год по ключевым системам государственного управления с простым показателем «да/нет» в Регистре активности лица (РАЛ). В РАЛ используется защищенный идентификационный ключ (Protected Identifier Key – PIK) для снижения риска нарушения конфиденциальности, сохраняющий при этом возможности увязки во времени и между источниками данных. Основными источниками административных данных являются данные о рождаемости, пособиях на детей, образовании (дошкольном, начальном, среднем, высшем и последующем), занятости, безработице, трудовых пенсиях и социальном обеспечении (включая государственную пенсию). РАЛ преследует двоякую цель:

- ↑ обеспечивает возможность продольного анализа когорт населения в различных административных системах;
- ↑ обеспечивает изучение структуры населения во времени;
- ↑ обеспечивает обобщенный мастер-ключ в отношении различных административных источников данных с целью изучения осуществимости различных потенциальных проектов.

Потенциал проведения продольного анализа конкретных когорт населения оказался весьма ценным для продвижения концепции «Объединенное правительство нуждается в объединенных данных» во всем ирландском государственном секторе. В строгом смысле этого слова РАЛ не является регистром, а просто набором данных статистики народонаселения (НДСН).

Существует ряд официальных идентификационных номеров, используемых в отношении предприятий в основных системах налогообложения, занятости и регистрации компаний. Эти номера взаимосвязаны и доступны в регистре предприятий ЦСУ для использования источников данных в статистических целях. Налоговые органы также осуществляют увязку этих номеров. В настоящее время планируется расширить использование официальных идентификационных номеров в отношении предприятий в государственном секторе.

На основе налоговых деклараций работодателей и работников был разработан увязанный статистический файл работодателей/работников. В этом файле отражены все виды оплачиваемой занятости в государстве, и он упрощает увязку компонентов НИС, касающихся предприятий и лиц.

Недавно в Ирландии была внедрена новая система почтовых индексов. Почтовый индекс однозначно идентифицирует каждый почтовый ящик в государстве. В настоящее время использование почтового индекса не является обязательным. Скорее всего, потребуется несколько лет для того, чтобы эта система почтовых индексов стала уникальным идентификатором для объектов недвижимости, имеющимся во всех официальных системах. В Ирландии более 30% адресных строк не являются уникальными и различаются только благодаря имени человека и знанию местного почтальона кто где живет для доставки почты. Система почтовых индексов и лежащий в ее основе регистр объектов недвижимости служат основой выборочной совокупности и генеральной адресной совокупности для социальных обследований и переписей населения 2011 и 2016 годов.

Очевидные пробелы в источниках данных для проведения переписи с использованием административных данных связаны с охватом населения и местом проживания населения. Помимо постоянного стремления к повышению качества данных, эти пробелы устраняются путем совершенствования процессов преобразования (второй этап подхода), в ходе которых административные данные преобразуются в статистический регистр с помощью таких методов, как моделирование и увязка.

Справочная информация о переписи

Ирландия традиционно характеризуется высоким уровнем мобильности населения и миграционных потоков, что обуславливает необходимость проведения переписи каждые пять лет (4,6 млн человек проживали в 1,6 млн домов в 2011 году). Как правило, Ирландия проводит традиционную перепись методом фактического учета, в ходе которой примерно 4 000 переписчиков вручают каждому домохозяйству переписной лист для заполнения главой домохозяйства в отношении каждого человека, присутствующего в момент переписи. Затем переписчики собирают переписные листы, которые затем обобщаются и сводятся воедино ЦСУ в централизованном порядке. Опыт 2016 года показал, что становится все труднее связаться с каждым домохозяйством.

Планы в отношении будущей переписи населения

Неизвестно, насколько осуществимо проведение переписи в Ирландии на основе административных источников данных и существующих обследований. Однако если Ирландия должна будет провести такую перепись населения, то ее можно было бы провести следующим образом:

Этап 1: создание обновленного главного адресного файла, в который будут также сведены характеристики зданий, полученные в ходе предыдущих переписей, и административные данные для оказания помощи в осуществлении жилищного компонента переписи. Также можно будет определять статус занятости (вакантное/занятое) жилищ на основе данных коммунальных предприятий.

Этап 2: выявление всех лиц, взаимодействующих с системами государственного управления в год проведения переписи, и обобщение этой информации в регистре с помощью личного идентификационного номера. Речь идет о регистре активности лиц (РАЛ), о котором говорилось выше. РАЛ будет содержать список всех лиц, которые являются активными в заданном году (в соответствии с подходом «признаков жизни»). Основные административные источники данных в ирландском контексте будут включать в себя: данные о рождениях; пособия на детей; база данных об учащихся начальной школы; источники данных о поступлении в высшие и последипломные учебные заведения и их окончании; источники данных о занятости (в том числе самозанятости) налоговых органов; лица, зарегистрированные по договорам аренды объектов недвижимости; лица, обращающиеся за медицинской помощью (или регистрирующиеся в течение года); социальное обеспечение и трудовые/государственные пенсии.

Этап 3: использование статистических методов для корректировки на недостаточный охват и присвоения поправочного коэффициента каждой записи. Статистические методы могут опираться на существующие

выборочные обследования или другой независимый источник. В настоящее время изучается один из таких методов, предусматривающий использование данных о продлении водительских прав в качестве второго независимого источника для выявления неполного охвата в регистре лиц с использованием методов двойного охвата.

Этап 4: присвоение каждому идентифицированному лицу адреса/почтового индекса из главного адресного файла. Использование надлежащего алгоритма дерева решений, который способен учитывать ситуации, когда в административных источниках данных выявляется более одного адреса человека. В этот алгоритм могут также включаться связи между членами домохозяйства, определенные в административных источниках данных (например, пособия на детей).

Этап 5: определение связей между членами домохозяйства с использованием связей, выявленных в административных источниках данных, и лиц, проживающих в одном доме.

Этап 6: оценка и включение атрибутов каждого лица в РАЛ с использованием существующих обследований, административных источников данных и соответствующих методологий.

Этап 7: все результаты переписи теперь компилируются на основе РАЛ, обновленного в ходе вышеуказанных этапов.

Перепись 2021 года с большей вероятностью будет использована для проверки некоторых аспектов проведения переписи на основе административных данных. Дополнительную информацию о методах, используемых и применяемых для составления демографических оценок из административных источников данных, см. в работе Zhang and Dunne, 2017. В ней также описывается расширение методологии DSE под названием усеченная DSE, которая может использоваться для поиска избыточных или ошибочных записей в системе DSE.

Справочная литература

Dunne, J (2015). The Irish Statistical System and The Emerging Census Opportunity. NTTS 2015. Statistical Journal of the IAOS, Vol 31, No. 3, pp. 391-400, 2015. В настоящее время с ним можно ознакомиться по адресу <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji915>.

MacFeely, S and Dunne, J (2014). 'Joining Up Public Service Information: The Rationale for A National Data Infrastructure'. Administration, Vol. 61, No. 4 (2014), pp. 93-107. https://unstats.un.org/unsd/trade/events/2014/india/background/Forum_Vol_61_4.pdf.

Zhang, LC and Dunne, J (2017) 'Trimmed Dual System Estimation' in Capture-Recapture Methods for the Social and Medical Sciences pp. 237 – 258, Chapman & Hall/CRC.

Приложение В – Практический пример Эстонии

Отказ от традиционной переписи

В 2000-е годы Статистическое управление Эстонии проделало весьма продуктивную работу по систематическому совершенствованию регистров, включая идентификацию данных во всех регистрах на основе личных идентификационных кодов и их увязку с эстонской системой X-Road³⁰ для облегчения обмена информацией. Однако первоначальный анализ показал, что эстонские регистры на тот момент не были готовы к успешному проведению переписи населения в 2011 году. Главные причины заключались в следующем:

- ↑ сопоставление с данными обследования рабочей силы за 2007–2015 годы показало, что по меньшей мере 20% адресов, указанных в Регистре населения, не являются фактическими местами проживания соответствующих лиц;
- ↑ информационная система образования содержит только данные о молодежи (дипломы об общем и высшем образовании за период с 2000 года; дипломы о профессионально-техническом образовании и другие свидетельства за более поздний период);
- ↑ ни один регистр не содержит информации о занятиях лиц;
- ↑ адреса регистрировались по-разному в различных регистрах, с различной степенью конкретности, что делало данные несовместимыми;
- ↑ регистры использовались только в течение короткого периода времени, а их качество и адекватность не были проверены;
- ↑ не были проанализированы согласованность определений, используемых в различных регистрах, и совместимость регистров с точки зрения информационной технологии.

Исходя из этого в 2009 году была утверждена комбинированная методология для переписи 2011 года с целью оптимизации использования регистров и варианта самоисчисления, а также максимального по возможности безбумажного проведения переписи:

- a) *Координация источников данных.* Ранее созданные источники данных (регистры) использовались в сочетании с интернет-опросом и очным опросом. В ходе переписи 2011 года регистры использовались тремя способами: в качестве инструмента подготовки переписи (подготовка рабочих списков и переписных таблиц), для предварительного заполнения переписных листов и дополнения результатов переписи в случае отсутствия данных. Информация об учебной деятельности подлежащих переписи постоянных жителей была взята из Информационной системы образования (EHIS), и соответствующий вопрос не был включен в переписные листы.
- b) *Комбинированная методология обследования.* В отличие от предыдущих переписей населения Эстонии, в 2011 году в дополнение к личному опросу использовались вопросники для самостоятельного заполнения. Это потребовало подготовки подробных инструкций по обучению переписчиков и предоставления исчерпывающих рекомендаций лицам, подлежащим переписи.
- c) *Комбинированная методология сбора данных.* Все предыдущие переписи населения Эстонии проводились с использованием бумажных вопросников или переписных листов. Ввод данных о зарегистрированных в рамках переписи лицах (ответы на вопросы переписи) осуществлялся с этих бумажных или картонных листов переписчиками с помощью специального карандаша для машинного считывания. В ходе переписи населения и жилищного фонда (ПНЖФ) 2011 года были внедрены две

³⁰ X-road – это система, которая облегчает взаимодействие граждан с государством с помощью электронных решений. С помощью X-road можно получить персональную информацию из административных регистров (с учетом ограничений по защите данных), но система также позволяет передавать данные из регистров в Статистическое управление Эстонии.

новые технологии, потребовавшие значительного обучения: самостоятельное заполнение вопросников через Интернет и ввод ответов непосредственно на портативные компьютеры в ходе личного опроса переписчиками. Вместе с тем в качестве резервного для чрезвычайных ситуаций был сохранен вариант использования бумажных вопросников. В особо исключительных обстоятельствах, особенно в тех случаях, когда доступ к домашним хозяйствам был крайне затруднен (например, на малых отдаленных островах), также планировалось проведение личного опроса по телефону. На практике эти чрезвычайные способы опроса были использованы в менее 1% случаев.

В целом, комбинированная перепись населения Эстонии прошла успешно. Две трети (67%) лиц, подлежащих переписи, использовали вариант самоисчисления через Интернет, а оставшаяся одна треть была опрошена лично (Statistics Estonia, 2014).

Подготовка к проведению регистровых переписей в Эстонии

К 2010 году – еще до проведения комбинированной переписи 2011 года – была начата подготовка к проведению в 2021 году регистровой переписи. С этой целью был использован опыт тех стран, которые уже провели регистровую перепись, и была начата реализация проекта (REGREL) по разработке перехода к регистровой методологии, первым этапом которого стал масштабный анализ (см. пункты 6–8 ниже), начавшийся осенью 2010 года и завершившийся в сентябре 2013 года (Puur et al., 2013).

Проект методологии REGREL (который на 80% финансируется Европейским социальным фондом) представляет собой партнерство между Статистическим управлением Эстонии, Эстонским институтом демографических исследований (в Таллиннском университете) и консалтинговой фирмой AS Ernst & Young Baltic. Анализ проводился несколькими десятками ученых и экспертов из Тартуского университета и Таллиннского университета, юристами и аналитиками Статистического управления Эстонии. Весьма важную роль сыграли представители баз данных и регистров, которые приняли активное участие в этом процессе (Puur et al., 2013).

Анализ состоял из двух частей:

- ↑ мета-анализ обязательных признаков ПНЖФ;
- ↑ подробный анализ признаков, которые требуют анализа качества данных (согласно результатам мета-анализа).

В дополнение к этому анализу группа также провела другие работы по подготовке к регистровым переписям:

- ↑ юридический анализ;
- ↑ подготовка методических указаний по созданию глоссария переписи;
- ↑ анализ международного опыта и практики.

Всего проектная группа проанализировала данные, хранящиеся в около 20 регистрах. Одним из наиболее значимых результатов данного проекта по методологии стала сеть основных регистров и баз данных для REGREL (содержащих данные по обязательным признакам ЕС, предусмотренным Регламентом ЕС № 763/2008).

Результаты проекта по методологии REGREL показали, что для подготовки к регистровым переписям еще многое предстоит сделать. Статистическое управление Эстонии будет осуществлять и координировать эту работу. Однако значительная часть этой работы осуществляется заинтересованными субъектами за пределами Статистического управления Эстонии, отвечающими за ведение регистров, в целях устранения недостатков, преодоления проблем и узких мест в эстонской системе регистров.

Параллельно с процессом реализации проекта REGREL еще одной важной частью подготовительной работы являлась разработка программных решений для переноса необходимых данных из регистров, расчета признаков переписи и их сохранения в базе данных Статистического управления Эстонии. Кроме того, были разработаны необходимые предложения по внесению изменений в законы с целью получения доступа к административной информации, необходимой для расчета переписных переменных. Еще одна важная задача заключалась в подготовке и внедрении алгоритмов для расчета переписных переменных на основе административных данных.

Регулярное сотрудничество с держателями административных регистров имело важнейшее значение; проводилась проверка качества содержащихся в них данных и, при необходимости, формулировались рекомендации по повышению качества.

Пилотная перепись 2014 года

Основная цель пилотной переписи REGREL заключалась в тестировании производственной системы с использованием обязательных признаков переписи ЕС, а именно: место обычного жительства; пол; возраст; юридическое брачное состояние; страна рождения; место рождения; гражданство; связи между членами домохозяйства и уровень образования. После проведения пилотной переписи ее результаты были проанализированы, а возникшие проблемы учтены в планах на последующие годы.

План действий на период 2016–2020 годов

Основная подготовительная работа к проведению регистровой переписи в течение этого периода касалась решения следующих задач:

- ↑ получение данных из регистров (контракты, описание набора данных, проверки качества данных и процедура получения);
- ↑ формирование признаков переписи, программирование необходимых правил;
- ↑ тестирование статистической системы в целом;
- ↑ тестирование системы статистических регистров и наполнение ее данными 2015–2018 годов;
- ↑ анализ потребностей и ожиданий потенциальных пользователей.

Статистическое управление Эстонии разработало комплекс правовых и организационных мер для улучшения качества, своевременности и охвата административных регистров, необходимых для переписи населения. Этот комплекс мер был представлен соответствующим ведомствам. В настоящее время регистры не содержат данных, необходимых для расчета всех требуемых признаков переписи по всему населению, а также неясно, достаточно ли регулярно обновляются все данные регистров.

Пробная перепись 2016 года

Мероприятия пробной переписи охватывали период со 2 января по 8 декабря 2016 года и предусматривали расчет:

- ↑ общей численности обычно проживающего населения Эстонии;
- ↑ общей численности традиционных жилищ, независимо от статуса занятости, и занятых нетрадиционных жилищ в Эстонии.

Цель заключалась в проведении на практике регистровой переписи населения и жилищного фонда с соблюдением рекомендаций и требований к качеству Евростата, насколько это возможно. Для этого данные должны были быть получены из 24 общенациональных баз данных, после чего они должны были быть обработаны и проанализированы.

На основе переписных признаков, рассчитанных с помощью регистровых данных с использованием специально разработанных алгоритмов, планируется создать серию гиперкубов (часть гиперкубов, требуемых Евростатом в соответствии с Регламентом ЕС).

Последующая проверка качества будет состоять из трех частей:

- ↑ качество всех переписных признаков будет проверяться с точки зрения пяти ключевых критериев качества;
- ↑ качество всех гиперкубов (и одновременно маргинальных кубов) будет проверяться с учетом качества признаков, включенных в кубы;
- ↑ сопоставление структур домохозяйств и семей при использовании двух различных определений домохозяйств: концепции совместного ведения хозяйства (используемой чаще в случае традиционных переписей) и жилищной концепции (принятой для регистровых переписей).

Будет составлен отчет о качестве, в котором будут определены все показатели и проблемы, связанные с качеством. В тех случаях, когда в связи с данными из регистров возникнут вопросы качества, соответствующие держатели данных будут проинформированы о них, и с ними будут обсуждаться возможные решения.

В ходе следующей пробной переписи, запланированной на 2019 год, подготовка к регистровой переписи будет сосредоточена на наборах данных (решение проблем, выявленных на более ранних этапах) и потоках данных. Кроме того, будет более подробно проанализировано влияние мер по обеспечению конфиденциальности на качество данных.

Резюме

В Эстонии ведется подготовка к проведению регистровой переписи в 2021 году, в рамках которой данные будут получены из различных баз данных при одновременном принятии мер по защите данных и обеспечению статистической безопасности. Разрабатываемые для баз данных требования будут достаточными для обеспечения совместимости государственных информационных систем, если:

- ↑ все релевантные данные будут предоставляться вместе с метаданными, включая классификационные коды;
- ↑ ввод и обновление данных будут выполняться через систему X-Road;
- ↑ данные будут предоставляться в формате XML, описание данных предоставляться разработчиком сервиса X-Way в виде XSD, а обновления содержать время предоставления.

Первичные данные в базах данных должны соответствовать требованиям к качеству, чтобы гарантировать достижение целей переписи. Качество будет определяться следующими факторами:

- a) охват регистров должен составлять не менее 97% населения и 95% в отношении отдельных признаков;
- b) около 95% данных должны быть увязаны с классификациями, зарегистрированными в RIHA (Управление информационных систем Эстонии);
- c) всем постоянным жителям и иностранным гражданам должен быть присвоен уникальный идентификационный код.

Справочная литература

Puur A, Sakkeus L and Aben S (2013). Development of a register-based population and housing census (REGREL) methodology. Project Final Report. Ernst & Young Baltic AS, Estonian Institute of Demography, TU. Tallinn (имеется только на эстонском языке).

Statistics Estonia (2014). Annual Report 2013. Скачать можно по адресу <http://www.stat.ee/annual-report>.

Приложение С – Практический пример Польши

Вводная информация о польской переписи населения и жилищного фонда 2011 года

Национальная перепись населения и жилищного фонда (НСП-2011), проведенная в Польше в 2011 году Центральным статистическим управлением (ЦСУ), была разработана и осуществлена с применением смешанной модели, которая предусматривает использование данных административных регистров (охватывающих базовые демографические показатели по всему населению) и данных, полученных непосредственно от респондентов в ходе двух обследований (одно охватывает 20% общенациональной выборочной совокупности с использованием подробного переписного листа, а другое – 80% с использованием короткого переписного листа) и собранных с помощью электронных вопросников. В результате использование бумажных вопросников было полностью прекращено.

Законодательство, регулирующее проведение переписи 2011 года (Закон об общенациональной переписи 2011 года), предусматривало максимально широкое использование для целей переписи систем государственного управления, что означало, что информация из таких административных источников должна была использоваться для подготовки и обновления регистра адресов и жилья (с последующим формированием базовой совокупности адресов и жилья для выборочного обследования), а также в качестве источника самих переписных данных. Данные, отсутствующие в информационной системе государственного управления или не обладающие достаточным качеством для использования в целях переписи, собирались непосредственно от респондентов. Этот метод был признан более безопасным и эффективным с учетом тогдашнего уровня развития административных источников, их качества и степени продвинутой методологической работы по вопросам оценки и импутации отсутствующих данных.

Использование административных источников

Необходимость использования данных из административных систем в польской статистике была обусловлена:

- ↑ экономическими причинами – в частности, требованиями повышения эффективности, минимизации затрат на подготовку статистических данных и уменьшения нагрузки на респондентов;
- ↑ риском увеличения коэффициента непредставления ответов в ходе статистических обследований, включая переписи;
- ↑ активным развитием ИТ-систем государственного управления с использованием передовых технологий.

Проведение переписи на основе административных и неадминистративных систем данных принесло множество выгод, включая:

- ↑ эффективное использование административных и неадминистративных источников;
- ↑ снижение расходов;
- ↑ уменьшение нагрузки на респондентов;
- ↑ улучшение безопасности данных;
- ↑ гарантию гармонизации с обследованиями благодаря использованию единых идентификаторов;
- ↑ перспективу предоставления данных переписи в будущем на ежегодной основе;
- ↑ наличие данных по любому уровню территориального дезагрегирования;

- ↑ улучшенную возможность выявления ошибок двойного ввода (двойного счета);
- ↑ создание базы микроданных, поддерживающей косвенную оценку – моделирование на уровне единиц;
- ↑ повышение качества оценки по малым районам;
- ↑ повышение согласованности и достоверности статистических данных.

Решение об использовании данных из административных регистров в целях переписи потребовало углубленного обзора широкого спектра информации, которая будет доступна в этих источниках. Был проведен анализ всех источников и переменных, которые могут быть использованы в целях переписи. Для облегчения этого были получены метаданные по примерно 300 административным регистрам, из которых 30 были сочтены наиболее полезными. В каждом из этих регистров были открыты отдельные записи, и все переменные из этих источников были подвергнуты анализу полезности. Переменные оценивались на предмет их соответствия с точки зрения определений и классификации с использованием словарей, существующих в польской и европейской статистике. Были определены соответствующие весовые коэффициенты как для переменных, так и для административных регистров, из которых эти переменные были взяты, с учетом их полезности и качества. Оценка качества и полезности переменных из различных регистров послужила основой для разработки правил интеграции данных, а также для оценки и импутации в созданной оперативной базе микроданных. Результатом этой работы стали бесценные знания, касающиеся полезности каждого регистра и потенциала интеграции тех различных регистров, которые имелись в распоряжении статистической службы.

В рамках данного мероприятия ЦСУ использовало 28 источников центральных и местных органов власти, а также внешних органов государственного управления, такие как регистры управляющих компаний, жилищных кооперативов, энергораспределительных компаний и телекоммуникационных операторов. Администраторам этих баз данных были направлены запросы об использовании их информации в целях переписи.

Для того чтобы администраторы могли дистанционно передавать данные из своих различных систем (включая, например, более 2 500 местных органов власти), ЦСУ создало электронную платформу для сбора и обработки данных вместе с сетевым приложением для прямой передачи данных по защищенному каналу связи.

Данные записей уровня единиц, полученные из регистров, были преобразованы в статистические регистры, при этом одновременно осуществлялся процесс очистки, удаления дубликатов и стандартизации данных. Процесс осуществлялся в среде DQS SAS. Одновременно собирались метаданные о качестве входных данных, полученных из регистров, применявшихся процедурах очистки и окончательном качестве, полученном после применения процедур DQS. Данные из административных источников, преобразованные в статистический регистр, затем использовались для расчета главного файла – набора переменных, полученных из регистров, которые были включены в переписные листы для проверки (подтверждения или обновления) респондентами.

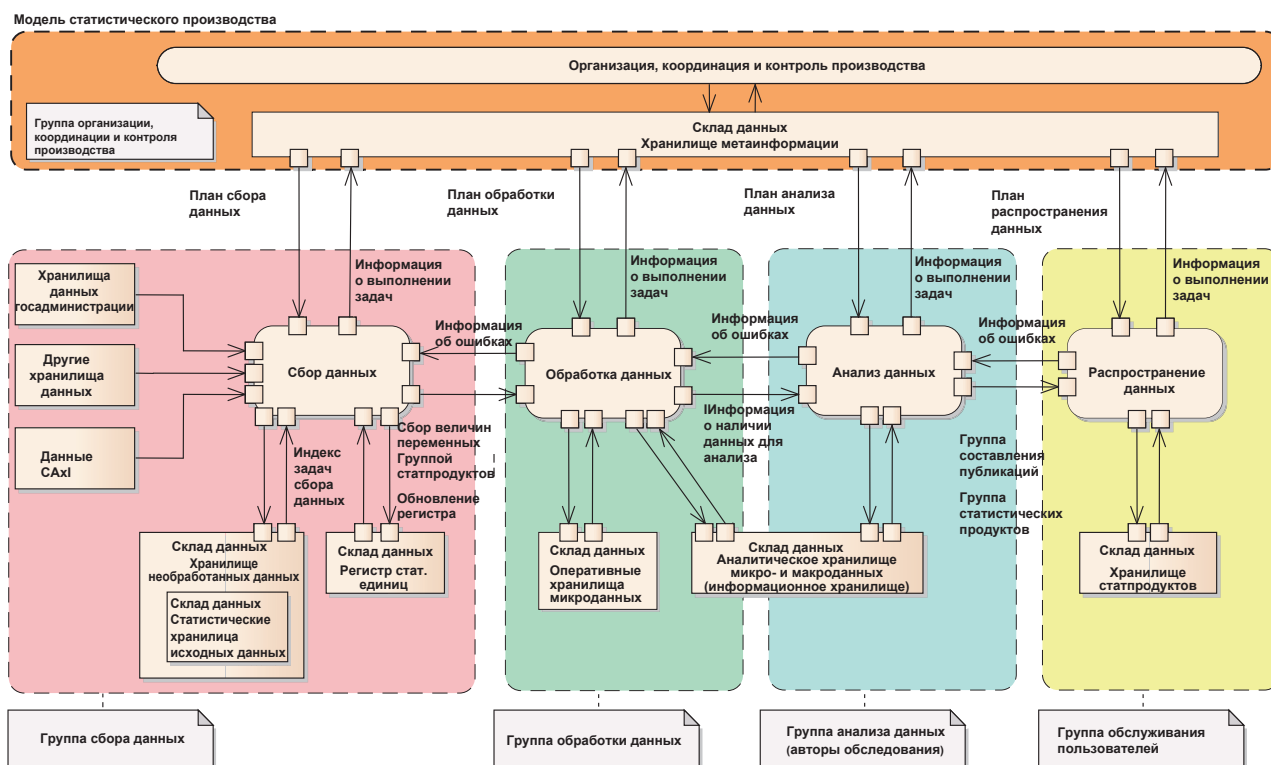
Оценка качества данных из административных источников

Качество данных систем государственного управления было трудно измерить из-за сложности и многогранности таких вопросов, как отсутствие какой-либо возможности использовать единый синтетический показатель. Таким образом, оценка качества была основана на множественных показателях.

Оценка качества включала в себя три элемента: анализ источников исходных данных; систем, наборов данных после преобразования (т.е. после их адаптации для использования в целях переписи) и статистических продуктов; и результирующих данных. Что касается наборов из административных источников данных, то была проведена оценка качества наборов исходных данных, предоставленных администраторами, и наборов, полученных после преобразования (т.е. после их адаптации для использования в целях переписи). Измерение качества проводилось на всех этапах, во всех процессах разработки административных данных и в процессе интеграции данных из различных источников. Приведенная ниже схема служит графическим отображением процессов переписи, включая оценку качества.

Для сбора методологических, технических и оперативных метаданных было создано «хранилище метаданных». Оно позволяло осуществлять мониторинг и контроль качества всех процессов переписи, начиная со сбора данных и заканчивая распространением результатов, и обеспечивать их качество.

Диаграмма С.1 Переписные процессы и оценки качества



Другие методы сбора данных, использовавшиеся в ходе переписи

Польша стала одной из первых стран в мире, которая разработала совершенно инновационную методику, предусматривающую одновременное использование для сбора переписных данных нескольких самых современных методов. Наряду с использованием государственных административных регистров применялись три метода сбора данных от респондентов, которые фигурируют под общим названием САХI:

- ↑ САИ/САWИ (компьютеризованный интернет-опрос/компьютеризованный веб-опрос) – самостоятельно заполняемый онлайн-вопросник, который обеспечивает проверку данных о респонденте, полученных из административных источников, в установленные сроки и, в случае необходимости, корректировку и предоставление информации, отсутствующей в регистрах;
- ↑ САТI (компьютеризованный опрос по телефону) – компьютеризованный телефонный опрос, проводимый переписчиком с целью дополнения данных, которые являются неполными или отсутствуют в результатах выборочного обследования;
- ↑ САРI (компьютеризованный личный опрос) – опрос, проводимый переписчиком по месту жительства, при котором данные регистрируются на переносном устройстве.

Все три канала опирались исключительно на адаптивный электронный вопросник, что обеспечило высокое качество данных на этапе сбора. Данный электронный вопросник был разработан специально для использования в рамках каждого из трех режимов сбора данных САХI. Соответствующее приложение вопросника (доступное на терминале мобильной связи или в интернет-браузере) проверяло точность заполнения анкеты, в том числе с помощью логического и счетного контроля.

Архитектура переписи: ИТ-система переписи

Для обеспечения оптимального применения передовых ИТ- и телекоммуникационных технологий в ходе переписи необходимо было создать соответствующую архитектуру переписи (см. ниже). Для целей разработки и проведения переписи ЦСУ внедрило ИТ-систему переписи (ИСП), состоящую из более чем 10 компонентов, поставленных различными подрядчиками, которые обеспечивали ИТ-поддержку всех операций в рамках переписи. ИСП включала в себя различные технологии (начиная с приложений, установленных на мобильных терминалах, и заканчивая системами управления опросами по телефону и оказания помощи в их проведении, и заканчивая специализированными базами данных, хранилищами данных и инструментами анализа и представления отчетности).

Диаграмма С.2 ИТ-система переписи



ИСП использовала различные решения, обеспечивающие высокий уровень безопасности обрабатываемых переписных данных. Были внедрены соответствующие меры и процедуры контроля, обязывающие всех участников соблюдать статистическую конфиденциальность и гарантировать защиту персональных данных.

ЦСУ осуществило ряд этапов обработки данных переписи, включая обработку данных административных регистров. Они включали в себя:

- ↑ подготовку правил нормализации, контроля, редактирования и импутации в отношении наборов данных из административных источников;
- ↑ подготовку правил синхронизации данных из административных систем – гармонизация учетных периодов – таблиц перехода из состояния, в котором данные были получены из административных систем, в желаемое состояние;
- ↑ подготовку правил дополнения отсутствующих данных опроса переписчиками – импутация и калибровка;
- ↑ подготовку правил и методов точной и строгой увязки данных из различных административных систем;
- ↑ определение значений переменных, включенных в программу переписи (правила источника данных);
- ↑ подготовку правил расчета значений переписных переменных;
- ↑ подготовку правил создания производных объектов – создание новых объектов (домохозяйств, семей);

- ↑ подготовку модели/метода расчета данных с использованием данных из административных систем и статистических обследований;
- ↑ подготовку правил обезличивания данных.

В соответствии с Законом об общенациональной переписи 2011 года в ЦСУ была разработана, подготовлена и реализована Оперативная база микроданных (ОБМ). Эта система включала в себя аппаратно-программную инфраструктуру (компьютерное оборудование, системное программное обеспечение, инструментальное программное обеспечение) и прикладное программное обеспечение (компьютерные программы). Эта база данных позволила включить переписные данные, передаваемые в электронной форме по четырем информационным каналам лицами, которые должны делать это согласно Закону (т.е. лицами, ответственными за ведение административных регистров, переписчиками, интервьюерами по телефону или респондентами), и облегчила дальнейшую обработку данных (проверку, редактирование, импутацию и обезличивание). На следующем этапе обезличенные данные были переданы в Аналитическую базу микроданных (АБМ).

В метаинформационной подсистеме хранилища были обобщены необходимые метаданные, описывающие данные и процессы переписи, в том числе необходимые для подготовки отчетов о качестве. Задача этой метаинформационной подсистемы заключалась в обеспечении согласованного определения статистических единиц для ОБМ и АБМ. Система также использовалась для хранения оперативных метаданных систем ОБМ и АБМ. Эта подсистема представляет собой Центральное хранилище метаданных (ЦХМ).

Функция АБМ заключается в хранении обезличенных данных переписи в их окончательном виде. В этой базе данных проводится статистический анализ для получения результатов, необходимых для публикации, – итогов переписи. АБМ обеспечивает всем получателям статистической информации удобный доступ к данным в виде агрегированных показателей. Система АБМ представляет собой платформу для анализа и представления данных, которая в настоящее время позволяет осуществлять статистическую подготовку выходных данных. Результаты анализа в форме отчетов, таблиц, карт и других графических материалов доступны как внутренним, так и внешним пользователям.

Технология ГИС

Впервые в Польше ГИС (географические информационные системы) использовались как при проведении переписи, так и при формировании ее итогов.

Использование различных справочных материалов и регистров, содержащих пространственную информацию, позволило ЦСУ создать пространственные данные по статистическим адресным точкам и единицам статистического деления страны. Цифровые карты являются незаменимым справочным источником, используемым переписчиками для того, чтобы они могли ориентироваться и проверять местоположение жилищ при обходе; руководителями гмин (эквивалент малой территориальной единицы уровня ЛАЕ 2) – для мониторинга хода переписи на своей территории; руководителями воеводств (эквивалент региона уровня 2 КТЕС) и главными инспекторами – для мониторинга хода переписи на региональном или национальном уровне. Карты использовались для мониторинга хода переписи в определенном районе или работы конкретного переписчика; местоположение по запросу или ежедневный маршрут могли быть визуализированы на карте.

Геостатистический портал

Геостатистический портал служит инструментом интерактивного картографического представления и публикации данных, собранных в ходе переписи. Он служит для хранения и представления данных, а также для обмена информацией между широким кругом пользователей.

Интерфейс Геостатистического портала позволяет его пользователям быстро и легко получить доступ к опубликованной статистической информации переписи. Данные представляются с использованием таких методов картографического представления, как картограммы и карты-диаграммы. Пользователи могут также определить собственные параметры визуализации тематических явлений для заданной картограммы. В дополнение к использованию готовых видов пространственного анализа на Геостатистическом портале пользователи могут формировать запросы микроданных, нарисовав полигон вручную на карте и/или

используя инструменты эскиза, которые включают линейный/дистанционный анализ и буферизацию объектов. Запросы микроданных могут формироваться по отдельным переменным переписи.

Резюме

Перепись 2011 года в Польше стала инновационным проектом не только на национальном, но и на глобальном уровне, исходя из следующих критериев:

- ↑ впервые в Европе был осуществлен в общенациональном масштабе одновременный сбор данных по четырем различным каналам (административные регистры, самоисчисление через Интернет, прямые опросы, проводившиеся переписчиками с использованием электронных вопросников на переносных устройствах, и опросы по телефону, проводившиеся статистическими счетчиками);
- ↑ бумажные вопросники были полностью устранены и заменены решениями ИКТ;
- ↑ данные, полученные из 28 административных регистров и трех неадминистративных источников, были эффективно интегрированы;
- ↑ технология ГИС использовалась в подготовительной работе для мониторинга хода работы переписчиков и обеспечивала возможность компилировать и представлять результаты переписи на основе многомерного пространственного анализа;
- ↑ ИТ-система переписи состояла из ряда решений, обеспечивших высокий уровень безопасности и конфиденциальности обрабатываемых данных;
- ↑ были разработаны современные технологии обработки статистических данных, которые также окажут значительное влияние на методологию будущих статистических обследований;
- ↑ была создана комплексная телеинформационная структура, значительно повысившая автоматизацию обработки статистических данных.

Всесторонний анализ всего хода переписи позволил ЦСУ сделать несколько выводов, а извлеченные уроки позволяют оценить возможности дальнейшего совершенствования переписей в будущем. Новая технология, применявшаяся в ходе переписи 2011 года, доказала, что она также может быть применена в других анкетных обследованиях. Она является более дешевой, использует современные механизмы контроля, повышая качество собираемых данных, и снижает нагрузку на респондентов.

Запланированные сроки были соблюдены, ни один из сроков, установленных для этапа сбора данных (оговоренных в Законе о переписи), не был продлен, и бюджет переписи не был превышен. Подробный график проведения переписи, включающий более 250 отдельных мероприятий, регулярно обновлялся. Рамочный график и подробные графики выполнения включенных в него подзадач (таких, как подготовка и процедуры контроля работы переписчиков и ИТ-поддержка систем переписи) хранились в отдельных файлах. График включал в себя в общей сложности несколько тысяч таких задач.

Следует признать, что эффективность проведения переписи была обусловлена как методологической, так и детальной организационной и материально-технической подготовкой.

Впереди еще один раунд переписей населения. Благодаря накопленному опыту Польша применительно к раунду 2020 года рассматривает возможность использования еще более новых технологий с целью сделать следующую перепись еще более эффективной и инновационной. В межпереписной период продолжается работа по совершенствованию методов проведения переписей с использованием в качестве отправной точки опыта, приобретенного в 2011 году. Вместе с тем по-прежнему требуется предпринять значительные усилия для разработки новой стратегии переписи, с тем чтобы гарантировать внедрение прогрессивных решений. Будут продолжены попытки, направленные на:

- ↑ снижение расходов;
- ↑ использование административных источников более эффективным образом;
- ↑ снижение нагрузки на респондентов при сборе данных;
- ↑ повышение безопасности и конфиденциальности передаваемых данных;
- ↑ повышение согласованности и достоверности статистических итогов.

Приложение D – Практический пример Австрии

В рамках переписи населения и жилищного фонда 2011 года впервые в истории переписей Австрии информация о лицах, зданиях и жилищах была сформирована исключительно на основе административных и статистических регистров. Десятью годами ранее перепись проводилась традиционным методом с использованием вопросников и переписчиков. Переход от традиционной переписи к регистровой был осуществлен за относительно короткое время. В настоящем приложении описываются общие условия и правовые рамки данного перехода, а также дается краткое описание регистровой модели переписи.

Решение правительства

В 2000 году правительство Австрии объявило о своих планах перехода на новую методологию проведения переписи после переписи 2001 года. Перепись на основе вопросников воспринималась как устаревшая; нагрузка на респондентов была слишком высокой, результаты представлялись слишком поздно, а затраты были слишком высоки. Кроме того, значительная часть информации, которая запрашивалась у респондентов, уже имела в административных регистрах. В долгосрочной перспективе регистровые переписи рассматривались в качестве гораздо более дешевой альтернативы традиционному подходу.

В конце 1990-х годов Статистическое управление Австрии провело оценку возможностей замены традиционных переписей, следующую из которых планировалось провести в 2001 году, использованием данных из административных регистров. Однако из-за отсутствия централизованного регистра населения (на тот момент существовало более 2 300 отдельных муниципальных регистров населения), отсутствия важных базовых регистров (таких, как регистры уровня образования и жилья) и уникальных идентификаторов лиц или адресов по итогам этой оценки был сделан вывод о том, что качество регистровой переписи в 2001 году будет очень низким. Увязка персональных данных из разных источников могла быть осуществлена только с использованием имени и характеристик значительного числа признаков, которые в любом случае были недоступны.

Правительство последовало рекомендации Статистического управления Австрии о проведении традиционной переписи населения в 2001 году и ее выводу о том, что для успешной замены традиционной переписи регистровой необходимо будет осуществить ряд мер. Поэтому решение правительства, принятое в 2000 году, сопровождалось объявлением о том, что Министерство внутренних дел создаст центральный регистр населения, что будут установлены правовые и технические требования в отношении обезличенной увязки административных регистров и что качество административных регистров требуется повысить.

Создание необходимых условий

Центральный регистр населения (ЦРН)

В Австрии регистрация населения является обязательной, но до 2002 года каждый муниципалитет имел свой собственный регистр. Центральный регистр населения (ЦРН) начал функционировать 1 марта 2002 года. Первоначальная оценка численности населения была рассчитана на основе 2 300 или более муниципальных регистров в рамках переписи 2001 года³¹.

³¹ В рамках переписи 2001 года муниципалитеты должны были загрузить данные из своих систем регистрации населения в центральную базу данных (GSG **G**emeinde**S**oftware**G**roßzählung 2001), которая была предоставлена Статистическим управлением Австрии. Первоначальные данные для Центрального регистра населения были взяты из этой базы данных. Центральная база данных также поддерживала перепись в муниципалитетах (переписное районирование, распределение переписных участков между счетчиками, отчетность и т.д.). Она была предварительно заполнена адресами зданий из адресного регистра Статистического управления Австрии. Лица из местных регистров населения были распределены по этим адресам, и предстояло ввести результаты переписи в отношении места жительства.

ЦРН содержит такие переменные, как пол, возраст, страна гражданства, место и страна рождения, тип проживания, адрес основного места жительства и других мест проживания в Австрии, страна предыдущего места жительства и страна назначения в случае эмиграции из Австрии. В 2006 году в качестве новой характеристики была включена переменная «юридическое брачное состояние». Однако ЦРН не содержит информации о связях между членами семьи, которую, например, можно найти в регистрах населения стран Северной Европы.

Регистр уровня образования, Регистр учащихся и студентов

Одним из первых мероприятий по обеспечению базовых условий для проведения регистровой переписи в Австрии стало создание статистического регистра уровня образования. Вместе с созданием этого нового регистра была обновлена база статистики образования в целом (на основе Закона 2002 года о документировании образования). Начиная с 2003 года учебные заведения должны представлять индивидуальные данные об учащихся и студентах раз в год (учетная дата 30 сентября), содержащие информацию о текущем охвате школьным образованием (тип учебного заведения, уровень обучения и т.д.) и об успеваемости по состоянию на конец предыдущего учебного года, включая данные об окончании учебного заведения.

В соответствии с Законом 2002 года регистр уровня образования был создан с использованием статистической информации о самом высоком достигнутом уровне образования и других переменных, таких как дата рождения, пол и адресный код места жительства, из результатов переписи населения 2001 года. Регистр регулярно обновляется с использованием информации о выпускниках школ, университетов и программ профессионально-технического обучения (ученичество). Регистр безработных и органы, ответственные за признание зарубежных дипломов о среднем и высшем образовании, служат дополнительными источниками информации о самом высоком достигнутом уровне образования.

Регистр зданий и жилищ (РЗЖ)

Регистр зданий и жилищ был создан в соответствии с Законом о Регистре зданий и жилищ 2004 года; держателем регистра является Статистическое управление Австрии. Он начал функционировать в ноябре 2004 года и выполняет роль административного регистра для муниципалитетов, которые обязаны обновлять информацию в случае перестройки зданий или жилищ или возведения новых зданий. Регистр содержит информацию об адресах и характеристиках зданий и жилищ и других жилищных единиц, а также о строительной деятельности (разрешения на строительство, завершение строительством зданий и жилищ).

Базовые данные этого регистра были получены из результатов переписи населения 2001 года и статистики строительных работ с целью восполнения пробелов за период 2001–2004 годов. Первоначальные данные также были получены из цифровой кадастровой карты и базы данных земельного кадастра Федерального управления по метрологии и топографической съемке.

Адреса, здания и жилища имеют уникальные идентификационные номера, которые также используются в Центральном регистре населения (ЦРН). Местные власти не могут регистрировать лиц в ЦРН, если адрес отсутствует в РЗЖ.

Отраслевой персональный идентификационный номер для официальной статистики (bPIN OC)

Хотя номер социального страхования используется во многих административных регистрах Австрии, его не удалось использовать для увязки данных из различных источников в ходе регистровой переписи населения, главным образом по причинам, связанным с защитой персональных данных. Номер социального страхования содержит дату рождения лица и не гарантирует анонимность.

В 2004 году Законом об электронном правительстве был введен так называемый отраслевой персональный идентификационный номер (bPIN) для обмена данными между государственными органами в рамках электронного правительства. Каждая «отрасль», такая как «здравоохранение», «социальное обеспечение», «налоги» или «официальная статистика», использует свой собственный ПИН. bPIN-коды формируются на основе исходного ПИН-кода, который присвоен каждому лицу в Центральном регистре населения. Органом, ответственным за эту процедуру, является «Stammzahlenregisterbehörde» Австрийского органа по защите данных (ОЗД).

При проведении регистровой переписи использование ПИН для целей официальной статистики bPIN (bPIN OC) осуществляется по следующим этапам. Прежде чем данные из регистра будут переданы Статистическому управлению Австрии, держатель регистра предоставляет идентификационные данные лиц ОЗД. На первом этапе эти данные сопоставляются с данными ЦРН с помощью этих идентификационных данных. В случае совпадения bPIN OC и bPIN держателя регистра можно сформировать на основе исходного ПИН-кода с использованием несколько сложного алгоритма. Оба bPINs шифруются, и их размер составляет 172 цифры. После возврата кодов держателю регистра данные, извлеченные для Статистического управления Австрии, дополняются зашифрованными bPIN и, наконец, представляются ему. Только Статистическое управление Австрии имеет ключ для расшифровки bPIN OC, который затем имеет размер в 28 цифр. Зашифрованный bPIN держателя регистра позволяет идентифицировать соответствующую запись в случае запроса Статистического управления Австрии. Благодаря использованию дешифрованных bPIN OC данные из различных административных регистров могут быть увязаны на индивидуальном уровне.

Переписное законодательство

Существовавший на тот момент закон о переписи не обеспечивал достаточную правовую базу для проведения регистровой переписи в 2011 году, и, таким образом, необходимо было разработать новый закон о переписи на основе подробных концепций и требований к такой переписи, предоставленных Статистическим управлением Австрии в консультации с экспертами. Он был подготовлен в 2004 году и одобрен межведомственной рабочей группой в июне 2005 года. В марте 2006 года вступило в силу соответствующее законодательство о регистровой переписи.

Основные принципы регистровой переписи

Охват переписи

Регистровая перепись охватывает только те основные признаки, которые содержатся в Рекомендациях КЕС и предусмотрены Регламентом 763/2008 о переписи ЕС. Признаки, по которым имеется информация в регистрах, но не входившие в перечень признаков предыдущих переписей (такие, как «доход»), не включаются. Кроме того, поскольку данные по некоторым переменным не имеются ни в одном из регистров (например, используемый язык или вид транспорта), их невозможно собрать исключительно с помощью регистрового подхода. Вместе с тем следует отметить, что Закон о переписи также регулирует перепись местных структур занятости, которая является частью переписи в Австрии с 1981 года.

Критический момент переписи был установлен на 31 октября 2011 года. Установление критического момента переписи на конец/начало года не является хорошим вариантом для Австрии, поскольку в некоторых муниципалитетах численность населения может быть искажена из-за сезонной занятости в отрасли зимнего туризма.

Увязка регистров

Личный идентификационный номер

Как отмечалось ранее, увязка регистровых данных осуществляется с использованием bPIN OC. На практике присвоение bPIN OC является порой невозможным, если идентификационные данные являются неточными. Тем не менее держатель регистра обязан предоставить эти записи в Статистическое управление Австрии без ПИН-кода и без имени. Эти записи данных подлежат статистической увязке с использованием таких характеристик, как дата рождения, пол и почтовый индекс.

Идентификаторы адресов

Регистр зданий и жилищ связан с центральным регистром населения с помощью уникальных кодов адресов, номеров зданий и жилищ. То же самое относится и к регистру предприятий.

Избыточность

Разнообразные регистры, которые служат основой переписи, могут содержать расходящиеся значения по определенной характеристике одного и того же лица. Поэтому считается ненадежным доверять данным, взятым только из одного регистра. Для обеспечения приемлемого качества, применяется принцип избыточности: информация о поле, дате рождения, гражданстве, стране рождения и юридическом брачном состоянии собирается из максимально возможного числа регистров. Речь идет о так называемых «множественных атрибутах». Каждой переменной назначается базовый регистр и регистры «сравнения», которые используются для подтверждения значения в базовом регистре. Если значения различаются, то для определения наилучшего значения переменной применяется алгоритм.

Такие признаки, как «текущий статус активности», должны быть получены из различных регистров, в частности содержащих информацию об экономической активности и посещении учебных заведений.

Регистры

Новый Закон о регистрах определяет восемь базовых регистров. Базовый регистр также выполняет функцию регистра, который используется для обеспечения качества переменной, если эта переменная назначена другому базовому регистру. Следовательно, эти восемь регистров служат одновременно базовыми регистрами и регистрами сравнения для различных переменных. Базовыми регистрами (показаны на диаграмме D.1) и их держателями являются:

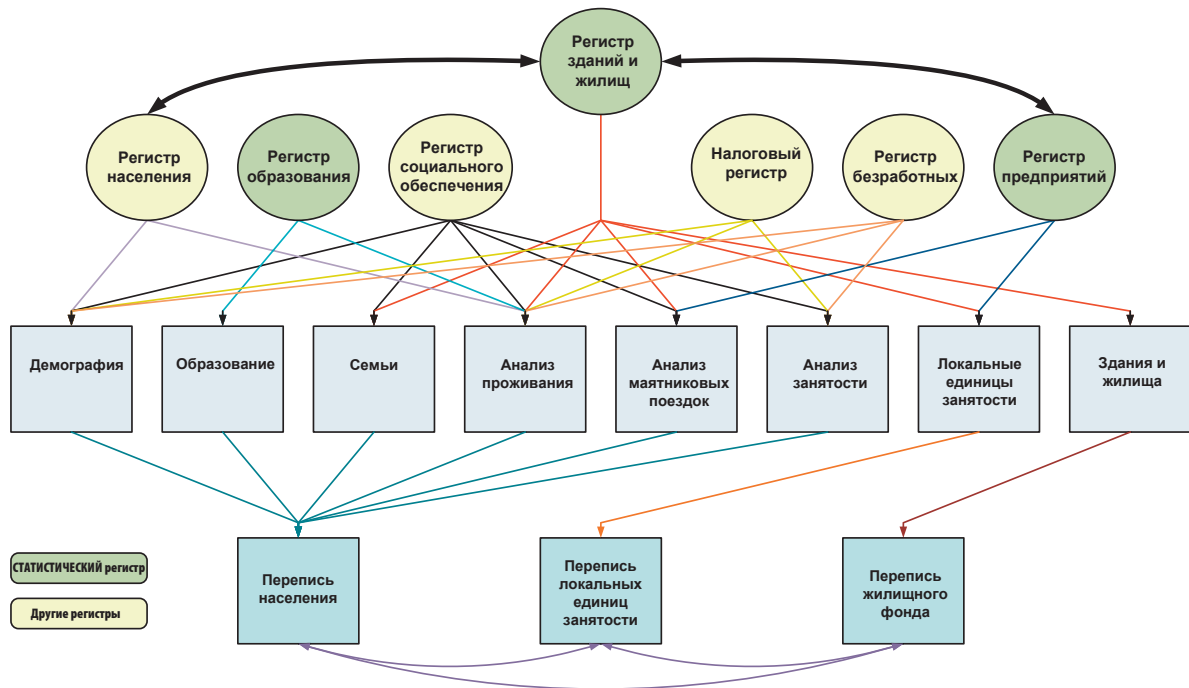
- ↑ Центральный регистр населения (ЦРН), Министерство внутренних дел
- ↑ Центральный регистр социального обеспечения (ЦРСО), Ассоциация учреждений социального обеспечения
- ↑ Налоговый регистр (НР), Министерство финансов
- ↑ Регистр безработных (РБ), Государственная служба занятости Австрии
- ↑ Регистр уровня образования (РУО), Статистическое управление Австрии
- ↑ Регистр учащихся и студентов (РУС), Статистическое управление Австрии
- ↑ Регистр зданий и жилищ (РЗЖ), Статистическое управление Австрии
- ↑ Регистр предприятий и местных подразделений (РП), Статистическое управление Австрии

Эти регистры содержат информацию по всем переменным, необходимым для регистрационной переписи. Дополнительными регистрами, которые используются для обеспечения качества результатов («регистры сравнения»), являются:

- ↑ Регистр семейных пособий (РСП)
- ↑ Центральный регистр иностранцев (ЦРИ; включая просителей убежища)
- ↑ Регистры государственных служащих федерального государства и земель (РГС)
- ↑ Регистр владельцев автомобилей (РВА)
- ↑ Регистр получателей социальной помощи (РПСП)
- ↑ Регистр призывников (РП)
- ↑ Регистр альтернативной гражданской службы (РАГС)

Регистры, служащие только для сравнения, содержат, главным образом, базовые демографические данные, а также дополнительную информацию о занятости государственных служащих (например, работают ли они полный или неполный рабочий день, месте работы и отрасли экономической деятельности) или информацию о военной или альтернативной гражданской службе в качестве дополнения к ЦРСО и НР.

Диаграмма D.1 Регистровая модель переписи, базовые регистры* и принцип избыточности



* Регистр образования представляет собой одновременно регистр уровня образования и регистр учащихся и студентов.

Анализ места жительства – избыточный охват в Регистре населения

Закон о регистровой переписи предусматривает, что каждое лицо, имеющее основное место жительства, зарегистрированное в ЦРН, подлежит переписи, если:

- ↑ он или она прожили в Австрии по меньшей мере три месяца, охватывающие критический момент переписи; и
- ↑ проживание подтверждается процедурой, называемой «анализом проживания».

Анализ проживания проводится для того, чтобы избежать избыточного охвата, которым, как ожидается, будет страдать ЦРН³². В Австрии доказательства регистрации требуются для различных видов деятельности, поэтому недостаточный охват в ЦРН, как ожидается, будет иметь гораздо меньшее значение.

На первом этапе все регистровые данные подвергаются увязке с использованием bPIN OC. Затем записи без bPIN OC статистически сопоставляются с записями ЦРН. В конце этого процесса неизбежно появятся записи (лица), не совпадающие с данными ЦРН и другими административными данными. Наличие записи в административном регистре, отличном от ЦРН, означает, что лицо имеет признаки жизни в учетный день (например, является занятым или безработным, получателем социального пособия, учится в университете) и, таким образом, подтверждает ЦРН.

Лица, которые зарегистрированы в ЦРН, но больше не имеют нигде удовлетворительных признаков жизни. Их проживание должно быть подтверждено путем предоставления ответа на официальное письмо³³, в котором задается вопрос о месте основного жительства в учетный день. В 2011 году около 96 000 человек был направлен соответствующий запрос, и они должны были вернуть заполненную и подписанную форму. Если респондент подтверждает факт проживания в Австрии, он должен быть учтен, и процедура анализа проживания завершается. Дело обстоит иным образом, если факт проживания не был подтвержден (например, когда лицо заявляет, что он или она живет за границей); письмо не может быть доставлено, потому что лицо

³² Эмигрировавшие лица часто не уведомляют местный регистрационный орган, или смерть может быть зарегистрирована с задержкой во времени.

³³ Только в отношении таких лиц Закон о регистровой переписи позволяет Управлению по защите данных проводить повторную идентификацию.

неизвестно; письмо доставлено, но ответа не поступило). Вместо того, чтобы просто не учитывать всех этих лиц, Статистическое управление Австрии обязано информировать муниципалитеты о результатах запросов. Местное самоуправление имеет право опровергнуть это, но должно доказать, что данное лицо по-прежнему имеет основное место жительства в муниципалитете, и представить доказательства этого Статистическому управлению Австрии.

По состоянию на 31 октября 2011 года ЦРН охватывал население численностью около 8 466 000 человек (не считая лиц, зарегистрированных менее трех месяцев назад), в то время как численность населения, согласно данным регистровой переписи, составила 8 401 940 человек. Около 62 800 человек были исключены из показателя численности населения с учетом результатов направления запросов лицам с незначительными признаками жизни.

Анализ проживания включал в себя и другие процедуры, такие как проверка случаев смерти с использованием данных из центрального регистра социального обеспечения (около 3 800 человек были установлены умершими, но все еще зарегистрированными), устранение двойного учета и устранение задержек с отменой регистрации. В определенной степени учитывается также недосчет по причине задержек с регистрацией, например в случае новорожденных детей или переезда на новый адрес, с использованием информации из данных ЦРН, полученных через шесть месяцев после учетного дня. Баланс ретроспективных поправок ЦРН составил около 3 100 человек. Таким образом, в то время как чрезмерный охват в ЦРН составил около 0,8%, недосчет из-за задержек с регистрацией был гораздо меньше.

Пробная перепись 2006 года

Один из основных элементов подготовки регистровой переписи 2011 года заключался в проведении пробной переписи (31 октября 2006 года) для оценки источников административных данных, которые было намечено использовать в 2011 году, анализа проживания и процедур обработки данных. Она содержала все элементы регистровой переписи, запланированной на 2011 год (сплошная регистрация и весь спектр признаков). Основное отличие заключалось в том, что подсчет населения не должен был иметь никаких правовых последствий для муниципалитетов (в плане финансовых трансфертов федерального центра муниципалитетам).

Для оценки качества результатов было проведено выборочное обследование. Выборка включала 25 000 человек (или 0,3% от общей численности населения) в 100 отобранных районах, и данные собирались переписчиками с использованием бумажных вопросников.

Результаты пробной переписи (Statistik Austria, 2009) показали, что информация из регистров в основном имела весьма высокое качество и что в ожидаемых проблемных областях, таких как статистика домохозяйств и семей, результаты в целом были удовлетворительными. Был сделан вывод о том, что концепция регистровой переписи доказала свою пригодность. Затем Статистическому управлению Австрии было поручено продолжить подготовку первой регистровой переписи 2011 года.

Ежегодная оценка численности населения по состоянию на 31 октября и регистровая статистика рынка труда

С учетом качества данных, полученных в ходе пробной переписи 2006 года, правительство приняло решение изменить порядок расчета финансовых трансфертов федерального центра муниципалитетам. В Австрии ключевым фактором расчета суммы налоговых поступлений, подлежащих перечислению, является численность населения муниципалитета, определяемая по результатам десятилетней переписи населения. Однако отныне было решено, что распределение средств должно пересчитываться ежегодно на основе ежегодной оценки численности населения по состоянию на 31 октября, начиная с 2008 года. Новый закон предусматривал, что для расчета этих ежегодных оценок численности населения должны применяться процедуры, предусмотренные для регистровой переписи. Единственным исключением является то, что анализ проживания в период между годами проведения переписи не предусматривает направления запросов лицам без признаков жизни, как это делалось в ходе переписи 2011 года (и пробной переписи 2006 года). Поэтому для сокращения избыточного учета используется модель оценки, основанная на результатах анализа проживания 2006 года (до 2010 года), а с 2012 года – на результатах анализа 2011 года.

На основе ежегодно предоставляемых данных административных регистров Статистическое управление Австрии приняло решение о подготовке ежегодной статистики переписного типа под названием «регистрация статистика рынка труда». В первый (2008) год были опубликованы данные лишь по небольшому набору признаков. В 2009 году были добавлены такие признаки, как наивысший достигнутый уровень образования и место работы. В 2011 году была проведена регистрационная перепись, включая анализ проживания, описанный в разделе 3.5. Начиная с 2012 года ежегодная регистрационная статистика охватывает все признаки, за исключением жилищных.

Оценка качества

Методика сбора данных из административных регистров потребовала совершенно новой концепции оценки качества. С этой целью была разработана система качества³⁴, которая позволяет проводить независимую оценку качества административных регистров, качества результатов и качества процессов на уровне индивидуальных переменных. Результаты оценки качества были опубликованы в отношении регистрационной переписи 2011 года и доступны в отношении ежегодной регистрационной статистики рынка труда, что дополнительно дает возможность сопоставлять результаты во времени.

Ожидаемые усовершенствования в рамках переписи 2021 года

Регистрационная перепись 2021 года будет проводиться на основе Закона о переписи 2006 года, без внесения существенных изменений в методологию и процедуры обработки. С ноября 2014 года данные о регистрации рождений, смертей, браков, зарегистрированных партнерств и разводов заносятся в центральную базу данных, которая ведется Министерством внутренних дел. Одновременно с этим был создан центральный регистр граждан. Статистическому управлению Австрии разрешен доступ к информации обоих этих административных регистров. Хотя качество данных по демографическим переменным уже является весьма хорошим, ожидается, что оно еще больше повысится, особенно по признакам «юридическое брачное состояние» и «связи между членами семьи».

В настоящее время Статистическое управление Австрии тестирует использование информации о занятиях лиц из налоговых файлов и регистра социального обеспечения. Для эффективного использования этой информации потребуется автоматическое кодирование.

Резюме

В Австрии необходимые условия для перехода от традиционной к регистрационной переписи касались:

- ↑ наличия комплексной системы регистров, созданной для административных потребностей (налоги, социальное обеспечение, регистрация места жительства и т.д.), готовности к созданию новых (статистических) регистров;
- ↑ правовой основы;
- ↑ возможности увязки данных из различных источников на индивидуальном уровне при одновременном обеспечении защиты данных;
- ↑ согласия правительства и заинтересованных сторон, общественного одобрения;
- ↑ сотрудничества между Статистическим управлением и другими органами и держателями регистров;
- ↑ получения углубленных знаний об административных источниках данных (путем постоянной работы с данными и оценки их качества).

Использование административных и статистических регистров стало неотъемлемой частью современного процесса формирования статистики в Статистическом управлении Австрии.

³⁴ См. справочные материалы в главе 5.

Справочная литература

Statistik Austria (2009). Bericht über die Probezählung 2006: Ergebnisse und Evaluierung. Wien. Доступны только на немецком языке. Скачать можно по адресу http://www.statistik.at/wcm/idc/idcplg?IdcService=GET_PDF_FILE&dDocName=036181.

Приложение Е – Практический пример Словении

История и эволюция переписи населения Словении с 1971 по 2002 год

Использование административных источников в целях формирования статистики и проведения переписи имеет в Словении давнюю традицию. В соответствии с долгосрочной стратегией по внедрению статистической модели скандинавских стран такие разработки были начаты уже в начале 1970-х годов. В то время Статистическое управление Республики Словения (СУРС) являлось одновременно инициатором и разработчиком предложения в отношении законодательства о национальных инфраструктурных регистрах. Кроме того, внедрение уникального персонального идентификатора (ПИН) в административные базы данных и статистические обследования имело решающее значение для разработки переписной статистики на основе регистров. В отсутствие аналогичных инициатив в других государственных органах в 1980-х годах в тесном сотрудничестве с соответствующими ведомствами СУРС самостоятельно в качестве разработчика создало четыре базовых регистра:

- ↑ Центральный регистр населения (ЦРН)
- ↑ Регистр территориальных единиц (адреса)
- ↑ Статистический регистр занятости
- ↑ Регистр предприятий

Поскольку административная функция регистра должна отличаться от статистической функции, три из них (за исключением Статистического регистра занятости) были переданы соответствующим министерствам после принятия в 1995 году Закона о национальной статистике. Закон о национальной статистике также предусматривал использование всеми государственными органами единых классификаций, и в тех случаях, когда это возможно и целесообразно, согласование концепций своих административных данных и переменных со статистическими концепциями и определениями.

На деле, первая (но безуспешная) попытка сделать ЦРН основой для статистики народонаселения и переписей была предпринята еще до переписи 1971 года. Однако вторая попытка, предпринятая в 1980 году, стала реальной отправной точкой для составления регистровой статистики, поскольку ПИН (по-прежнему используемый в том же формате) был присвоен перед началом переписи 1981 года всем постоянным жителям Словении, а также регистрировался в качестве переменной переписчиками, что привело к 80-процентному охвату ПИН в итоговой базе данных переписи 1981 года. Впервые в дополнение к ПИН в ходе переписи 1981 года для рационализации процесса сбора данных и повышения их качества были использованы некоторые административные данные (об уровне образования, занятии и отрасли), предоставленные работодателями, но только на бумажных формулярах. Данные переписи 1981 года также использовались в качестве основы для регулярного ежедневного обновления данных, основанных на статистических демографических обследованиях и некоторых административных данных, что привело к распространению данных о численности населения непосредственно из ЦРН с 1986 по 1994 год с использованием определения постоянного места жительства.

Следующую (1991 года) перепись населения, проведенную всего за три месяца до обретения Словенией независимости, можно охарактеризовать как первую переходную перепись. Впервые для сбора данных на местах были использованы предварительно заполненные переписные листы с использованием данных ЦРН (ПИН, имя, фамилия, адрес) и Регистра территориальных единиц (территориальные коды). Кроме того, на стадии обработки использовались данные Статистического регистра занятости. Это стало первым процессом реальной интеграции данных, использовавшимся в статистическом производстве в Словении. В то же время в ходе переписи 1991 года в последний раз проводилась классическая статистическая обработка данных (главным образом редактирование вручную). Оценка численности населения на основе данных ЦРН была на 1,8% выше, чем по итогам переписи 1991 года; охват ПИН в итоговой базе данных переписи составил 99%.

Организация сбора данных переписчиками и система статистической обработки в рамках переписи 2002 года стали важными шагами на пути к принятию полностью регистрового подхода к переписи 2011 года. Действительно, перепись 2002 года была единственной комбинированной переписью Словении, поскольку помимо полного очного опроса переписчиками с использованием бумажных вопросников данные по некоторым признакам переписи (место/страна рождения, последняя миграция, гражданство, брачное состояние, род занятий, отрасль, место работы) были взяты полностью из регистров и не собирались переписчиками, тогда как по некоторым другим признакам (пол, дата рождения, статус активности) переписные данные собирались исключительно переписчиками, если они отсутствовали в подготовленной до переписи базе данных, представлявшей собой набор данных, сформированный на основе различных административных и других статистических источников, в том числе данных переписи 1991 года.

Основными нововведениями переписи 2002 года стали следующие:

- ↑ построение двух баз данных, сформированных на основе девяти различных административных и статистических источников;
- ↑ унифицированные идентификаторы и штрих-коды, предварительно напечатанные на переписных листах;
- ↑ оптический архив изображений всех переписных листов;
- ↑ одновременная проверка и автоматическое кодирование на основе изображений;
- ↑ редактирование статистических данных в онлайн-режиме (проверка достоверности) с использованием изображений переписных листов.

Доперписная база данных была создана за шесть месяцев до контрольной даты (31 марта) и использовалась для предварительного заполнения переписных листов конкретными данными (например, коды переписных участков, адрес, имя, пол и ПИН), а также для планирования ряда переписных работ и мероприятий по обработке данных. Итоговая база данных, использовавшаяся для обработки данных, была создана через пять месяцев после контрольной даты. Охват ПИН был полным. Расхождение между данными переписи 2002 года и данными ЦРН было несколько меньшим, чем 11 лет тому назад (1,6%).

Создание необходимых условий

Общие предварительные условия

Решение перейти к полностью регистровой переписи было принято в 2007 году руководством СУРС при трех предварительных условиях, которые в конечном итоге были им выполнены:

- ↑ законодательство, обеспечивающее свободный доступ к административным источникам данных и увязку данных из различных источников;
- ↑ наличие соответствующих административных или статистических источников с уникальными идентификаторами для увязки данных о лицах, домохозяйствах и жилищах;
- ↑ наличие соответствующих переменных в источниках, охватывающих большую часть потребностей национальных пользователей и соответствующих Регламенту ЕС (тогдашнему проекту) о переписях населения и жилищного фонда, который был впоследствии принят в 2008 году.

Переписное законодательство

Не было необходимости в том, чтобы закон конкретно предписывал проведение регистровой переписи, поскольку правовая основа для этого уже существовала. Получение и интеграция данных разрешены статьями 32 и 33 Закона о национальной статистике (Официальный вестник Республики Словения, № 45/95 и 9/2001). Решение Словении о проведении регистровой переписи в качестве метода сбора и обработки данных было принято в соответствии со Среднесрочной программой статистических обследований на 2008–2012 годы (Официальный журнал РС, № 119/2007) и годовой программой статистических обследований на любой конкретный год, в который будет проводиться полная регистровая перепись.

Новая дата переписи

Согласованность результатов переписи с другой статистикой является важным шагом вперед по сравнению с предыдущими результатами очной переписи. Новая контрольная дата переписи (1 января) вместо 31 марта была выбрана по следующим причинам:

- ↑ многие административные источники используют календарный год;
- ↑ для облегчения сопоставимости данных переписи с данными ежегодных демографических обследований;
- ↑ более высокая согласованность административных источников в конце календарного года.

Новые разработки после переписи 2002 года

С точки зрения содержания, решение о проведении регистровой переписи было обоснованным, поскольку единственный отсутствующий регистр – регистр жилищ (регистр недвижимости) – был создан в 2007 году с использованием как специальной переписи недвижимости, проведенной Управлением топографической съемки Словении, так и имеющихся источников (геодезический кадастр, регистр приусадебных участков). В соответствии с Законом о национальной статистике создание административных регистров на основе статистических данных не допускается; поэтому данные о жилищах переписи 2002 года не могли быть использованы для этой цели. В то же время было определено число жилищ в многоквартирных домах и включены соответствующие адреса лиц в ЦРН.

Номер квартиры в настоящее время также является частью записи официального адреса в ЦРН, и он стал последним недостающим звеном, увязывающим лиц с их жилищами. Кроме того, в рамках компьютеризации административных внутренних дел МВД создало электронный регистр домашних хозяйств, который ранее велся вручную в виде картотек. Регистр домохозяйств является отличительной особенностью Словении, поскольку другие страны, проводящие регистровую перепись, не имеют такого высококачественного источника данных о структуре домохозяйств. Важнейшим преимуществом регистра домохозяйств является возможность внедрения концепции совместного ведения хозяйства и возможность разработки данных о связи с основным лицом домохозяйства, необходимых для определения состава семьи.

Доперписная оценка качества исходных данных

Первым шагом после оценки методологических решений, основанных на имеющихся административных и статистических источниках данных и одобренных соответствующим органом СУРС, стало проведение пробной переписи с целью анализа и оценки качества исходных данных с точки зрения охвата, релевантности, надежности, своевременности, доступности и сопоставимости. На этом раннем этапе были выявлены три важных препятствия достижению приемлемого качества:

- ↑ расхождения в данных о составе домохозяйств (незначительная проблема, решенная на стадии обработки);
- ↑ чрезмерный недостаточный охват номеров жилищ (особенно в многоквартирных домах) в ЦРН;
- ↑ общее низкое качество данных о жилье в качестве основной проблемы, которая не была надлежащим образом решена к контрольной дате переписи 2011 года.

Повышение качества данных о номерах жилищ в ЦРН

Тот факт, что полнота обновления данных о номерах жилищ в ЦРН была намного ниже ожидаемой, создал проблему, поскольку эта переменная имеет решающее значение для увязки жилищ с лицами и домохозяйствами. Эти данные отсутствовали в ходе пробной переписи в отношении приблизительно 400 000 человек (более половины всех лиц, проживающих в многоквартирных домах). Для исправления положения Министерство внутренних дел и СУРС провели в тесном сотрудничестве два основных мероприятия. Во-первых, были разработаны методологические решения для автоматизированного определения отсутствующих номеров жилищ путем увязки данных о владельцах жилья с зарегистрированным местом жительства собственников и их домохозяйств (при этом предполагалось, что большинство собственников проживают в принадлежащих им жилищах). Затем основному лицу домохозяйства, проживающего в многоквартирном доме, номер квартиры которого был неизвестен, направлялось официальное письмо с просьбой сообщить эту информацию; было разослано около 49 000 писем, и доля ответивших составила 75%.

Основные принципы регистровой переписи

Увязка данных о лицах, домохозяйствах и жилищах

Увязка данных о лицах, домохозяйствах и жилищах с использованием уникальных идентификаторов является одной из наиболее важных задач при подготовке многомерных данных переписи с использованием данных, собранных переписчиками, или регистровых данных. В случае регистровой переписи прямая увязка всех источников данных о лицах с использованием ПИН (как показано на диаграмме Е.1) является основной статистической операцией. ПИН преобразуется в статистический идентификатор (СИН) для защиты конфиденциальности и частной жизни до статистической обработки данных переписи, и при этом используется идентификатор домохозяйства из регистра домашних хозяйств. Идентификатор домохозяйства представляет собой порядковый номер домохозяйства, составляющий от 1 до NNNN по одному и тому же адресу. Номер жилища является идентификатором, увязывающим лиц и жилища, он также увязан с адресом (через порядковый номер).

Диаграмма Е.1 Идентификаторы, используемые при проведении регистровой переписи в Словении

Регистр/База данных	ПИН	Адрес	ИН жилища	ИН домохозяйства	ИН предприятия
Центральный регистр населения	X	X	X		
Регистр недвижимости	X	X	X		X
Регистр домохозяйств	X	X	X	X	
Статистический регистр занятости	X				X
Другие демографические источники	X				

Качество базовых идентификаторов

ПИН является наиболее важным идентификатором с (как было отмечено) полным охватом в ЦРН, но может отсутствовать в некоторых других административных или статистических источниках, используемых для целей переписи. Основной проблемой с точки зрения качества идентификатора домохозяйства является тот факт, что идентификатор домохозяйства (а также связь с основным лицом домохозяйства, которая также считается ключевым идентификатором в словенской регистровой системе) имеется в наличии только в отношении постоянного места жительства. Несмотря на усилия по улучшению охвата номеров жилищ лиц, проживающих в многоквартирных домах, в ЦРН, перед первым этапом интеграции данных многие идентификаторы жилищ все еще отсутствовали. Распределение входных и выходных данных по ключевым идентификаторам в рамках всего статистического процесса регистровых переписей 2011 и 2015 годов иллюстрирует таблица Е.1.

Основное различие между двумя переписями заключалось в качестве входных данных. В 2015 году по сравнению с четырьмя годами ранее отсутствовало гораздо меньше номеров жилищ (30 000 по сравнению с 89 000 в 2011 году). Еще более высокое качество было отмечено в регистре недвижимости в связи с принятием (впоследствии отмененного) Закона о массовой оценке недвижимости, который предусматривал, что размер налога на жилье будет, в частности, определяться площадью пола и годом строительства, с разными ставками для жилых и незанятых жилищных единиц. Соответственно, доля правильных записей, касающихся ИН жилищ, значительно возросла – с 75% в 2011 году до 94% в 2015 году.

Административные и статистические источники

Основу системы регистровой переписи формируют три административных регистра:

- ↑ Центральный регистр населения (ЦРН), который ведет Министерство внутренних дел;
- ↑ Регистр домохозяйств (РД), являющийся частью Центрального регистра населения;
- ↑ Регистр недвижимости (РН), который ведет Управление топографической съемки и картографии Словении.

Таблица Е.1 Показатели качества ключевых идентификаторов регистровой переписи Словении

Идентификатор	Число записей	Без изменений	Импутация	Исправление	Число записей	Без изменений	Импутация	Исправление
	2011				2015			
ИН жилищ ¹⁾	724 479	75,3	12,3	12,4	712 989	94,0	4,6	1,4
ИН домохозяйств ²⁾	2 016 423	94,9	2,1	3,0	2 024 604	93,9	1,5	4,6
Отношение к основному лицу ²⁾	2 016 423	91,6	4,2	4,2	2 024 604	91,6	1,5	6,9

Источник: СУРС, Статистическое управление Республики Словения.

¹⁾ Многоквартирные дома. ²⁾ Частные домохозяйства.

Большинство данных для переписи населения и жилищного фонда (как это предусмотрено Регламентом ЕС 763/2008) были получены из одного из этих источников, например:

- ↑ обычное место жительства, пол, возраст, юридическое брачное состояние, страна гражданства, место обычного жительства за год до проведения переписи – из ЦРН;
- ↑ связи между членами домохозяйства – из РД;
- ↑ все жилищные признаки, включая статус пользования жильем, – из РН.

Данные по другим персональным переменным были получены путем объединения входных данных из нескольких источников. Основным методологическим принципом составления статистики в таких случаях является иерархия источников. Это означает, что в случае наличия данных из нескольких источников по каждому лицу (определяемому ПИН) предпочтение отдается источнику с более высоким приоритетом (соответствующим более низкому значению в таблице Е.2 и таблице Е.3), присвоенным после завершения оценки качества всех источников. Для получения данных об уровне образования использовались девять источников. Данные об уровне образования в настоящее время обновляются на ежегодной основе с использованием источников данных, перечисленных в таблице Е.2 (за исключением переписи 2002 года).

Данные по экономическим характеристикам (текущий статус активности, занятие, отрасль, статус в занятости, нахождение места работы), которые формируются из восьми источников, перечисленных в таблице Е.3, как правило, относятся к контрольной дате переписи.

Данные по миграционным характеристикам (страна/место рождения, проживание за границей, прежнее место обычного проживания) были получены исключительно из статистических обследований, основанных на данных ЦРН:

- ↑ ежегодное статистическое обследование миграции (данные за 2002–2010 годы);
- ↑ ежегодное статистическое обследование рождаемости (данные за 2002–2010 годы);
- ↑ ежеквартальное статистическое обследование населения по состоянию на 1 января 2010 года; и
- ↑ перепись населения 2002 года.

Таблица Е.2 Источники данных об уровне образования, регистровая перепись 2011 года, Словения

Приоритет	Держатель источника	Содержание источника	Период	Доля (%) ³⁵
1	СУРС	Выпускники высших учебных заведений	1989–2010	11,1
2	Национальный экзаменационный центр	Выпускники общеобразовательных школ и профессионально-технических училищ	2002–2010	9,1
3	Палаты	Профессионально-техническое образование второй ступени	2002–2010	0,2
4	СУРС	Число студентов высших учебных заведений – в начале учебного года	2002/03–2010/11	2,6
5	Национальный экзаменационный центр	Национальные экзамены по окончании начальной школы	2006–2010	4,6
6	СУРС	Получатели стипендий	2006–2010	0,5
7	СУРС	Данные Статистического регистра занятости об уровне полученного образования	1986–2010	56,0
8	Служба занятости Словении	Зарегистрированные безработные	1.1.2011	0,8
9	СУРС	Перепись населения 2002 года – самый высокий уровень образования	31.3.2002	13,6

Источник: СУРС, Статистическое управление Республики Словения.

Таблица Е.3 Источники данных по экономическим характеристикам, регистровая перепись 2011 года, Словения

Приоритет	Держатель источника	Содержание источника	Период	Доля (%) ³⁵
1	СУРС–Статистический регистр занятости	Лица, работающие по найму Самозанятые лица и фермеры	На прошлой неделе	45,7
2	Служба занятости Словении	Зарегистрированные безработные	1.1.2011	5,9
3	СУРС	Студенты очной и заочной форм обучения в профессионально-технических и высших учебных заведениях	Учебный год	4,4
4	СУРС	Получатели стипендий в заведениях верхнего уровня среднего и высшего образования	1.1.2011	1,0
5	Институт пенсионного страхования и страхования по инвалидности	Получатели пенсии по старости, пенсии по инвалидности, пенсии по потере кормильца и государственных пенсий	1.1.2011	29,1
6	Институт медицинского страхования	Члены семьи застрахованных лиц и других неактивных лиц с медицинским страхованием	1.1.2011	10,6
7	Министерство труда, семьи и социальных дел	Получатели социальной и иной помощи и пособий	2010	0,8
8	Налоговое управление	Плательщики подоходного налога	2010	0,6

Источник: СУРС, Статистическое управление Республики Словения.

Статистический процесс

Наличие данных определило график четырехэтапной подготовки и распространения результатов регистровой переписи населения 2011 года:

- ↑ интеграция входных данных о населении, домохозяйствах и жилищах (первый выпуск некоторых окончательных данных переписи населения в конце апреля 2011 года);
- ↑ обработка данных о домохозяйствах и семьях (первый выпуск в июне 2011 года);

³⁵ Население в возрасте 15 лет и старше.

- ↑ данные по всем другим признакам переписи населения (экономические и образовательные характеристики, миграция, рождаемость) и предварительные данные о занятых жилищах были опубликованы в конце 2011 года;
- ↑ данные о занятых и незанятых жилищах обрабатывались последними, поскольку данные по характеристикам жилья из РН по состоянию на 1 января 2012 года были вновь обновлены (первый выпуск состоялся 21 июня 2012 года).

Данные, которые уже были опубликованы, впоследствии не пересматриваются и не обновляются ни на одном из последующих этапов процесса, в связи с чем для обеспечения отслеживаемости любых изменений готовятся специальные таблицы метаданных. Цель заключалась в том, чтобы сохранить последний статус индивидуальной записи в итоговой базе данных переписи. Были также достигнуты две другие цели: отслеживаемость и повторяемость. Другими словами, все изменения в данных, внесенные в ходе статистического процесса, регистрировались прозрачно и четко.

Был создан специальный веб-сайт для распространения данных регистровой переписи 2011 года (<http://www.stat.si/popis2011/eng/Default.aspx?lang=eng>), который содержал основные методологические пояснения и информацию. Кроме того, данные также доступны в режиме онлайн на портале данных SI-Stat Data Portal (<http://pxweb.stat.si/pxweb/Database/Demographics/Demographics.asp>) в разделе «Население и уровень жизни». Был опубликован только один печатный отчет, озаглавленный «Люди, семьи, жилища» (Статистическое управление Республики Словения, 2013).

Качество итоговых материалов регистровой переписи

Общее качество статистических данных в значительной степени зависит от качества лежащих в их основе административных данных, поскольку регистровая перепись становится не более чем статистической операцией по преобразованию административных данных в статистические материалы. Необходимыми условиями для обеспечения качества итоговых материалов с этой точки зрения являются:

- ↑ хорошо отлаженное и последовательное использование административных данных в статистическом процессе;
- ↑ очень тесное сотрудничество с держателем административного источника; и
- ↑ обратная связь по статистической оценке административного источника.

Регулярный мониторинг качества

Избыточный охват является наиболее распространенной проблемой любой статистической системы, основанной на регистрах. Для оценки качества охвата населения переписью в Словении используются два основных метода:

- ↑ Коэффициенты импутации данных об уровне образования и статусе участия в составе рабочей силы могут служить индикатором двойного учета, поскольку данные доступны только в ЦРН. Качество исходных данных ЦРН улучшилось, о чем свидетельствует тот факт, что коэффициент импутации данных о статусе участия в составе рабочей силы в 2011 году составил 1,50, в то время как в 2016 году – лишь 1,14;
- ↑ Статус проживания респондентов, отобранных для опроса в рамках социальных выборочных обследований, мог быть следующим:
 - проживает по включенному в выборку адресу (правильность);
 - умер, но не зарегистрирован (административные мертвые души) – всего несколько случаев;
 - проживает в другом месте в Словении (расхождения между фактическим и зарегистрированным местом жительства являются наиболее значимыми для территориального распределения данных о населении внутри Словении) – в диапазоне от 3% до 6%;
 - проживает за рубежом – избыточный учет составил порядка 1–2%;
 - отсутствие ответа о статусе проживания (отсутствие ответа) – менее 1%.

Специальное обследование охвата

В 2016 году было проведено специальное исследование, посвященное изучению вероятности незарегистрированной эмиграции. В выборку были включены лица, статус участия в составе рабочей силы которых пришлось определять методом импутации в ходе переписи 2015 года (15 500 человек); вторая выборочная группа состояла из лиц, указанных в административных данных (ЦРН) в качестве нерезидентов, однако данные о статусе участия которых в составе рабочей силы Словении можно было найти, по крайней мере, в одном из девяти источников (2 700 человек).

Были использованы два метода: метод почтовой рассылки вопросников для самоисчисления по всей выборке и очный опрос (1 915 человек) по избранным случаям непредоставления ответов по почте (71%). Окончательным результатом обследования стала оценка избыточной регистрации в размере 0,5% на агрегированном уровне (не более 10 000 человек). По сравнению с данными последней очной переписи 2002 года, когда почти 1% населения был учтен дважды (чрезмерный охват) и чуть менее 2% населения вообще не были учтены (недосчет), результаты этого обследования выглядят обнадеживающими.

Заключение

Основные демографические (переписного типа) данные в Словении формируются ежеквартально с использованием определения обычно проживающего населения. Статистическое определение населения полностью согласуется со всеми действующими Регламентами ЕС, определяющими обычное место жительства. Основные входные административные данные передаются ежеквартально из ЦРН, примерно через три месяца после учетной даты. Кроме того, ежегодно формируются данные о социально-экономических характеристиках с использованием статистических переписных процессов. Полная регистровая перепись в качестве регулярной статистической операции проводилась дважды (в 2015 и 2018 годах) в промежутке между обязательными годами, определенными в Регламенте ЕС (2011 и 2021 годы).

Регистровый подход Словении позволяет достичь ключевых целей, изложенных в документе *Вызовы для будущих переписей населения и жилищного фонда*, подготовленном Статистическим управлением Канады, Статкомитетом СНГ и секретариатом ЕЭК ООН для шестидесятой пленарной сессии Конференции европейских статистиков, состоявшейся в Париже в июне 2012 года, а именно:

- ↑ растущая обеспокоенность по поводу расходов: отсутствие дополнительного или специального бюджета, поскольку регистровая перепись в Словении проводится в настоящее время в форме регулярного статистического обследования в рамках ежегодной программы статистических обследований.
- ↑ Повышение качества данных: контролируемый методологический подход используется во всех этапах процесса, и не возникало никаких проблем, связанных с недосчетом или отсутствием ответов по отдельным признакам или трудностей с вводом и редактированием, присущим переписи методом обхода.
- ↑ Нагрузка на респондентов и снижение уровня участия в переписи: эта проблема более не существует в Словении.
- ↑ Неприкосновенность частной жизни: гораздо меньшее число людей в настоящее время имеют доступ к информации в ходе переписи, по сравнению с тысячами лиц, имевшими доступ к персональным данным при регистрации переписчиками.

Справочная литература

Statistical Office of the Republic of Slovenia (2013). *People, Families, Dwellings*. Population Census 2011. Издано и опубликовано Статистическим управлением Республики Словения, Любляна. Скачать можно по адресу <http://www.stat.si/StatWeb/File/DocSysFile/3712/people.pdf>.

Приложение F – Практический пример Португалии

В настоящем приложении кратко описывается работа, проделанная Статистическим управлением Португалии (СУП) в период 2014–2016 годов по созданию Статистического набора данных о населении (СНДН), целью которой является отражение постоянного населения страны и его описание с помощью набора демографических и социально-экономических переменных. В приложении описываются административные источники данных, а также методологический подход, показатели качества для оценки его пригодности и сопоставимости с итогами переписи 2011 года, ежегодными оценками численности населения (ОЧН) за 2015 год и пробной переписи (ПП) 2016 года.

В отличие от других стран, которые уже перешли к регистровой или комбинированной переписи, Португалия не имеет ни центрального регистра населения, ни уникального персонального идентификационного номера (ПИН). Кроме того, в стране отсутствует правовая база, позволяющая доступ к полному имени и адресу лиц в регистрах. Несмотря на ограничения этого процесса, результаты, достигнутые за этот короткий период времени, обнадеживают и прокладывают путь к новым мерам, которые могут привести к изменению парадигмы и разработке среднесрочных и долгосрочных стратегий, которые будут осуществляться в 2021 году и в последующий период.

Переписи в Португалии

Португалия проводит переписи с 1864 года, каждые десять лет начиная с 1890 года. На протяжении всей последовательности переписей СУП вносило изменения в ее процесс, чтобы сделать его более эффективным. В 2011 году был весьма успешно внедрен электронный вопросник, поскольку коэффициент представления ответов составил 50%.

Как и в некоторых других странах, в Португалии в настоящее время проводится исследование целесообразности с целью анализа различных методологических вариантов проведения переписей. В настоящее время проводится оценка потенциала использования регистров в целях повышения эффективности переписных операций и обеспечения возможности более частого выпуска и обновления статистических данных.

Построение статистического набора данных о населении в Португалии

Основная цель переписи заключается подсчете и описании постоянного населения, в частности путем публикации информации по малым географическим районам. Таким образом, чтобы понять, позволяет ли административная информация, имеющаяся в Португалии, провести качественную перепись постоянного населения, были осуществлены два мероприятия по созданию португальского СНДН. Эта база данных была создана на основе ряда регистров, предоставленных различными административными источниками данных. В качестве базисного года для первого мероприятия был выбран 2011 год, с тем чтобы использовать перепись 2011 года в качестве базы для сопоставления результатов. В качестве базисного года для второго мероприятия был выбран 2015 год, и результаты сопоставлялись с результатами ОЧН за тот же год.

Отправной точкой для СНДН послужил файл регистра актов гражданского состояния (РАГС). Этот регистр содержит демографические характеристики всех граждан Португалии. Однако РАГС не является центральным регистром населения и увязывает граждан только с их юридическим или зарегистрированным адресом в Португалии (что не обязательно соответствует переписной концепции обычного места жительства). РАГС завышает численность постоянного населения страны на 10% (более 1,1 млн человек) по сравнению с численностью населения, зарегистрированной в ходе переписи 2011 года, и не учитывает большинство иммигрантов, проживающих в Португалии и регистрируемых отдельно в Иммиграционном регистре (ИР).

СНДН был построен по методике «признаков жизни», определяемых присутствием человека в более чем одном регистре. В весьма упрощенном виде, лицо считается резидентом страны, если оно зарегистрировано в РАГС или ИР, и, кроме того, является «активным» по меньшей мере в одном из других регистров (свидетельствующем, например, о том, что лица учатся, работают, пользуются медицинскими услугами, платят налоги и т.д.). Применение алгоритмов, опирающихся на присутствие лица в различных регистрах, позволило определить и обособить тех, кто фактически проживает в Португалии, от тех, кто, не проживая в Португалии, сохраняет свой юридический адрес на территории Португалии.

На втором этапе лицам, включенным в СНДН, были присвоены соответствующие демографические и социально-экономические административные переменные.

Источники и переменные административных данных

В дополнение к Регистру актов гражданского состояния и Иммиграционному регистру для построения СНДН были использованы восемь следующих административных наборов данных:

- ↑ Регистр социальной защиты государственных служащих,
- ↑ Регистр государственных и трудовых пенсий,
- ↑ Регистр образования,
- ↑ Регистр занятых в частном секторе,
- ↑ Регистр безработных,
- ↑ Регистр социального обеспечения,
- ↑ Налоговый регистр и
- ↑ Национальный регистр пациентов системы медицинского обслуживания.

Административные переменные, имеющиеся в этих различных источниках, позволяют получить данные по 16 признакам переписи, включая 13, требуемых на географическом уровне, предписанном Регламентом ЕС, – пол, возраст, место обычного жительства, место жительства за год до проведения переписи, брачное состояние, гражданство, страна/место рождения, трудовой статус, род занятий, отрасль экономической деятельности, статус в занятости, место работы и уровень образования, – и три других на национальном уровне – количество отработанных часов, число работников на предприятии и посещение школы.

Охват каждой из этих переменных зависит от наличия индивидуальных записей в соответствующих исходных файлах: только в случае семи из шестнадцати переменных (пол, возраст, место обычного проживания, место жительства за год до переписи, брачное состояние, гражданство, страна/место рождения) обеспечивается полный охват информации для СНДН. Охват носит лишь частичный характер в случае социально-экономических переменных. Кроме того, по ряду важных переменных, касающихся характеристик лиц и домохозяйств (в частности, классификации домохозяйств), в настоящее время в регистрах искомая информация носит либо частичный характер, либо вообще отсутствует.

Кроме того, существуют некоторые связанные с жильем переменные, которые, хотя и имеются в регистрах, не могут быть использованы по причине конкретных правовых ограничений, указанных ниже. Исходя из этого Статистическое управление Португалии создало Статистический набор данных о жилищах (СНДЖ), который был сформирован на основе результатов переписи 2011 года и обновлен с использованием данных из ряда других источников, включая Национальную систему показателей городских работ (НСПГР)³⁶ и других выборочных обследований домохозяйств и лиц, проводимых СУП. Но из-за отсутствия разрешенного законом доступа к полному адресу (только название населенного пункта и почтовый индекс) он не может быть увязан с СНДН.

Качество регистров оценивалось в соответствии с рядом общих критериев, включая такие показатели, как уровень охвата каждой переменной, точность и своевременность информации, а также более точная оценка на уровне микроданных, путем сопоставления информации, собранной в ходе переписи 2011 года, с информацией, содержащейся в регистрах, начиная с базисных дат 2011 года.

³⁶ НСПГР опирается на административные данные 308 португальских муниципалитетов, касающихся разрешений на строительство и завершённых строительных работ.

Ограничения в процессе построения португальского СНДН: конфиденциальность и защита частной жизни

Доступ к регистрам предусмотрен Законом о национальной статистической системе (Закон № 22/2008 от 13 мая 2008 года). Национальная комиссия по защите данных подготовила ряд рекомендаций по обеспечению конфиденциальности данных каждого лица. Индивидуальные записи обезличиваются в источнике с помощью алгоритма, который шифрует цифровые идентификаторы, блокирует доступ к полному имени лица (только первые три буквы имени и три последние буквы фамилии), а также к полному адресу лица (только место жительства и почтовый индекс).

В Португалии уникальный ПИН отсутствует. Вместо этого существуют четыре цифровых идентификатора: номер РАГС (NIC) или номер ИР (в случае иммигрантов), номер налогового регистра (NIF), номер социального страхования (NISS) и номер Национальной службы здравоохранения. Регистры содержат один, два или три из этих цифровых идентификаторов, в зависимости от административного источника, но они не всегда охватывают все записи; так, например, РАГС содержит только NIC; налоговый регистр содержит только NIF; Регистр образования содержит NIC и NISS, но в первом случае это касается 90% записей, а во втором – менее 70%.

В отсутствие уникального ПИН, который, несомненно, повысил бы точность увязки и коэффициенты соответствия, пришлось создавать ключи увязки, главным образом с помощью детерминированных методов, основанных на личных характеристиках.

Результаты и оценка качества

В 2015 году численность населения, согласно оценке на основе СНДН, составила 10 434 161 человек, с отклонением 0,9% (около 93 тысяч человек) по сравнению с ОЧН за тот же год (10 341 330 человек). Возрастная структура населения и распределение по полу согласно СНДН также соответствовали показателям ОЧН 2015 года.

По сравнению с первым набором, относящимся к 2011 году и оценившим численность населения в размере 8,6 млн человек, эта новая версия СНДН весьма точно оценила численность населения на национальном уровне. Диаграмма F.1 иллюстрирует динамику результатов с внедрением новых административных источников. Улучшение качества СНДН 2015 года по сравнению с изданием 2011 года обусловлено двумя факторами: использование данных налогового регистра и усовершенствования в процессе увязки. Налоговый регистр с 9 миллионами записей был полезен для определения ПЖ, особенно тех групп населения, которые не ведут никакого бизнеса, не получают никаких социальных пособий, не учатся, но зарегистрированы в качестве членов домохозяйства в Налоговом регистре.

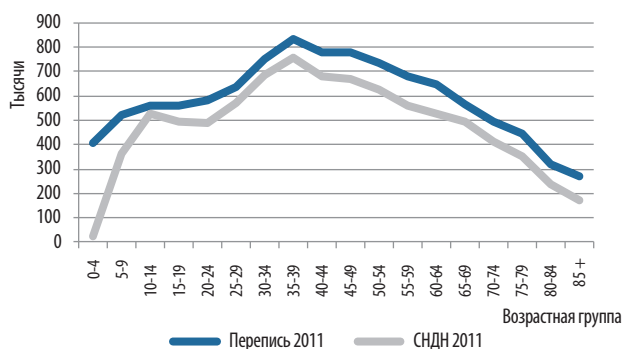
Кроме того, на региональном уровне (КТЕС 2) расхождения в подсчете населения согласовывались с ОЧН 2015 года, в диапазоне от 2% до 0,7%. Однако при рассмотрении численности населения на уровне «муниципалитет», диаграмма F.2 показывает, что существовали незначительные различия между СНДН 2015 года и ОЧН за тот же год в большинстве районов. Однако в 77% муниципалитетов страны расхождения между численностью населения, по оценкам СНДН и ОЧН, составили менее 5%.

Оценка расхождений по малым районам была проведена с помощью контрольного обследования (КП 2016 года), в пяти общинах (на уровне ЛАЕ 2), по одной в каждом из пяти муниципалитетов (на уровне ЛАЕ 1). Размер выборки составил 45 000 жилищ – 1% жилищных единиц, взятых из совокупности СНДЖ. Около 70 000 человек приняли участие в обследовании. Сопоставление результатов КП 2016 года и СНДН 2015 года привело к следующим выводам:

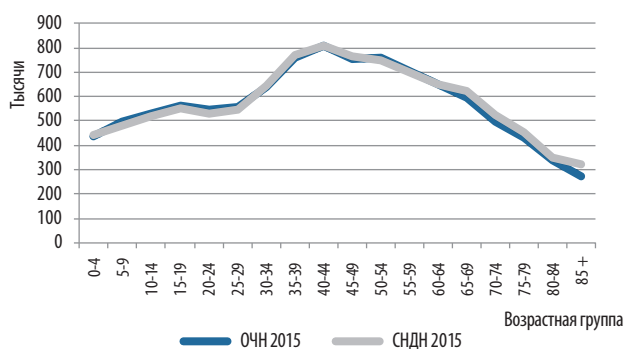
- ↑ Перепись населения на основе регистров по-прежнему страдает определенными ограничениями. СНДН 2015 года завысил численность постоянного населения во всех общинах КП 2016 года, с расхождениями от –5,7% до –21,8%.
- ↑ Хотя численность населения на уровне общин страдала некоторыми расхождениями, структура и характеристики населения в общинах КП, согласно СНДН 2015 года, полностью соответствовали структуре и характеристикам населения общин, зарегистрированным в рамках КП 2016 года.

Диаграмма F.1 Постоянное население в разбивке по возрастным группам, Португалия, 2011 и 2015 годы

Расхождение 2011 (СНДН в сравнении с переписью) –18%



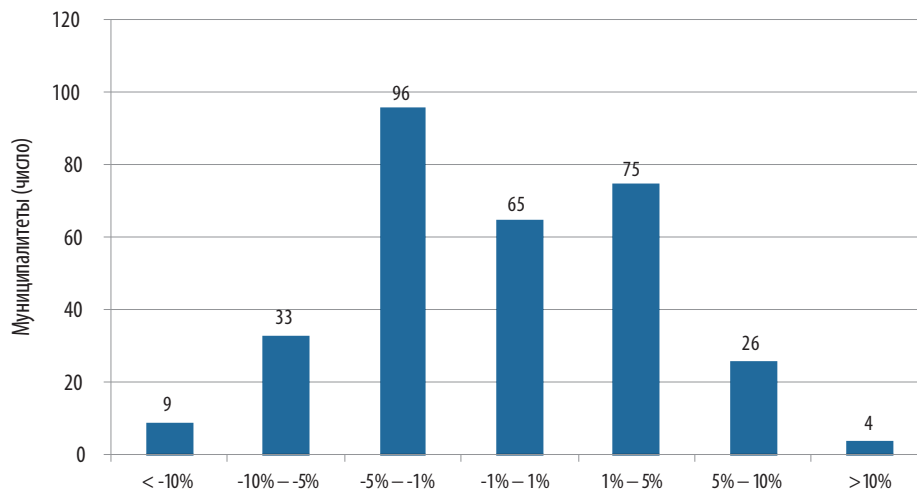
Расхождение 2015 (СНДН в сравнении с переписью) +0,9%



Источники данных:
 Гражданский регистр
 Иммиграционный регистр
 Регистр социального обеспечения
 Регистр социальной защиты государственных служащих
 Регистр государственных и трудовых пенсий
 Регистр занятых в частном секторе
 Регистр безработных
 Регистр образования

Дополнительные источники данных:
 Налоговый регистр
 Регистр пациентов

Диаграмма F.2 Количество муниципалитетов в разбивке по величине расхождений между СНДН и ОЧН 2015 года



- ↑ Совпадение на уровне микроданных между КП 2016 года и СНДН 2015 года (77%) было не вполне удовлетворительным из-за ограничений увязки, касающихся доступности полного имени и адреса этих лиц.
- ↑ Административные данные приблизительно соответствовали высокому качеству информации, собранной в общинах КП 2015 года, с очень высоким или, по крайней мере, приемлемо высоким уровнем совпадения по отдельным переменным (гражданство – 99,3%); место рождения – 96,8%; брачное состояние – 95,1%; посещение школы и уровень образования – 94,1%; и трудовой статус – 79,4%).

Выводы и перспективы

Построение СНДН позволило впервые в Португалии качественную и количественную оценку потенциала использования административных данных для целей переписи. Полученные результаты являются обнадеживающими, но недостаточно удовлетворительными для проведения полностью регистровой переписи в 2021 году. Главным образом это обусловлено тем, что:

- ↑ подсчет постоянного населения на основе регистров имеет ограничения в отношении малых районов (это может объясняться необновлением адресов в регистрах и неоптимизированными методами увязки данных); и
- ↑ имеющиеся в настоящее время административные данные не являются достаточными для получения ответов по всем переменным, обеспечиваемых переписью. Данные по ключевым признакам, таким как характеристики жилья, домохозяйств и семей и уровень образования, по-прежнему не могут быть получены из имеющихся регистров.

Однако, с учетом того что только одна методология, опирающаяся исключительно на административные данные, позволит получать более часто информацию переписного типа, текущие исследования по использованию регистров должны быть продолжены. Первым шагом в реализации этой стратегии является совершенствование методологии построения СНДН.

Эти текущие исследования опираются на успешное институциональное сотрудничество и благоприятные условия в системе государственного управления, являющиеся результатом проводимых в последние годы процессов модернизации страны, которые привели к открытию доступа к новым регистрам.

В настоящее время продолжается работа по переходу к регистровой переписи. Создание более благоприятной правовой базы (доступ к полному имени и адресу) является ключом к преодолению проблем, связанных с увязкой данных, в частности для обеспечения увязки между СНДН и СНДЖ, а также повышения точности оценок численности населения по малым районам.

Приложение G – Практический пример Англии и Уэльса

Справочная информация

В мае 2010 года Управление национальной статистики (УНС) Соединенного Королевства приступило к реализации «Программы на период после 2011 года» для изучения будущих возможностей формирования статистики народонаселения в Англии и Уэльсе, с тем чтобы информировать правительство и парламент о вариантах проведения следующей переписи. В частности, в рамках этой Программы основное внимание уделялось возможности замены переписи статистикой, формируемой на основе административных данных, уже имеющихся у правительства, дополняемых обследованиями домохозяйств.

На основе исследований и собранной фактологии тогдашний Государственный статистик рекомендовал в марте 2014 года (ONS, 2014):

- ↑ *«провести онлайн-перепись всех домохозяйств и институциональных заведений в Англии и Уэльсе в 2021 году... [с]... уделением особого внимания оказанию поддержки тем, кто не может пройти перепись в режиме онлайн; и*
- ↑ *шире использовать административные данные и обследования в целях улучшения статистических данных переписи 2021 года и совершенствования межпереписной статистики».*

Государственный статистик далее отметил, что:

«...[Это] может обеспечить будущему правительству и парламенту возможность добиться дальнейшего прогресса в переходе от традиционной десятилетней переписи к ежегодному формированию демографической статистики с использованием административных данных и ежегодных обследований».

Этот подход был одобрен в официальном ответе правительства на рекомендацию, вынесенную в июле 2014 года (Cabinet Office, 2014), в котором подчеркивается стремление:

«к тому, чтобы переписи после 2021 года проводились с использованием других источников данных и обеспечивали более оперативное представление статистической информации... [при условии] достаточного подтверждения предполагаемой практической осуществимости такого подхода».

В результате УНС стремится воспроизвести информацию, собираемую в рамках переписи, с использованием административных данных, уже имеющихся у правительства, дополняемых обследованиями. Цель заключается в том, чтобы иметь возможность сравнить результаты, опирающиеся на административные данные и целевые обследования, с данными переписи 2021 года, чтобы продемонстрировать, что данная альтернатива способна обеспечить формирование высококачественной информации при меньших затратах и делать это на более регулярной основе. (Аналогичные исследования параллельно проводятся Службой национального учета Шотландии и Агентством статистики и исследований Северной Ирландии, которые соответственно отвечают за проведение переписей в Шотландии и Северной Ирландии.)

Тем не менее такое изменение подхода является сложной задачей для Англии и Уэльса (как, по сути, и для всего Соединенного Королевства) по причине отсутствия регистра населения и уникального идентификатора в административных источниках. С учетом важности подготовки точных статистических данных прямой переход к такой системе без сопоставления результатов новых методов с переписью 2021 года был бы сопряжен с высоким риском. Это соответствует практике других стран, которые предпочли более постепенный подход.

Эта работа направлена на реализацию амбициозных целей правительства, о которых говорилось выше. Это также соответствует стратегии УНС (ONS, 2013a) – быть лидером в деле интеграции и использования данных из различных источников, более широко используя административные данные во всех отраслях статистики.

Что представляет собой перепись на основе административных данных?

УНС стремится к формированию информации, которая собирается в рамках десятилетней переписи населения (о лицах, домохозяйствах и жилищных единицах), с помощью переписи на основе административных данных. Это потребует комбинирования:

- ↑ административных данных уровня единиц, имеющихся в государственной администрации;
- ↑ обследования охвата населения;
- ↑ обследования характеристик населения.

Административные данные уровня единиц, имеющиеся в государственной администрации

УНС потребуется доступ к широкому спектру данных правительственных ведомств. Ему потребуется доступ к записям уровня единиц, с тем чтобы обеспечить взаимную увязку этих источников. Для обеспечения высококачественной увязки необходимы имя, адрес, дата рождения и пол, как это более подробно описано в документе *Matching Anonymous Data* (ONS, 2013b), поскольку комбинации этих переменных могут использоваться для установления связей с высокой степенью достоверности. Увязка множественных источников будет способствовать повышению качества и охвата результатов, которые могут быть получены, и будет поддерживать формирование перекрестных итоговых таблиц, таких как занятость в разбивке по квалификации на подробных географических уровнях в рамках территории местного самоуправления (ТМС).

Обследование охвата населения (ОБОН)

УНС предполагает необходимость проведения обследования охвата, аналогичного традиционному обследованию охвата переписи, для измерения и корректировки неполного или чрезмерного охвата в административных данных, а также в целях создания условий для формирования высококачественных статистических данных о численности населения. Как указано в информационном документе *Producing population estimates using administrative data: in theory* (ONS, 2013c) «Программы на период после 2011 года» УНС, ежегодное ОБОН может охватывать приблизительно 350 000 домохозяйств (1%). Кроме того, потребуется провести дополнительную работу для уточнения деталей как обследования, так и методов последующего расчета оценок с использованием ОБОН и административных данных.

Обследование для расчета оценок характеристик населения и жилищного фонда

В настоящее время УНС предполагает необходимость проведения обследования для сбора информации по переписным характеристикам. Оно может выполнять две функции:

- ↑ Что касается характеристик, имеющихся в административных данных, то обследование должно будет обеспечить измерение и корректировку недостаточного или чрезмерного охвата административных данных, по аналогии с ОБОН.
- ↑ Это обследование может также использоваться для получения прямых оценок по признакам, которые не могут быть получены из имеющихся административных данных (например, количество часов неоплачиваемого ухода). По таким признакам можно рассчитать оценки только на уровне ТМС (где численность населения колеблется в пределах от 2 200 до 1 074 000 человек).

Точная программа (включая размер выборки) этого обследования будет зависеть от степени доступа УНС к административным данным и понимания их статистического качества.

Другие источники данных

Другие данные (такие, как большие данные, данные частного сектора или коммерческие данные) также могут быть изучены на предмет формирования некоторых видов данных, традиционно получаемых в рамках переписи, таких как информация о потоках маятниковой миграции с использованием данных мобильной связи или информация о владельцах жилья с использованием веб-сайтов по вопросам жилья.

Эти различные источники данных, возможно, потребуется увязать друг с другом и использовать в сочетании с целым рядом методов и приемов моделирования для получения требуемых пользователям видов материалов, а также использовать для обеспечения качества результатов. Такой подход может также предоставить

пользователям возможность получать необходимые им материалы на более частой основе, чем проводимая раз в десять лет перепись.

Адаптация общей системы к подходу Англии и Уэльса

УНС также разработало и внедрило общую систему, описанную в главе 5 настоящих Руководящих принципов, с тем чтобы учесть свои конкретные проблемы, а именно:

Что необходимо сделать для перехода к переписи на основе административных данных?

Для проведения переписи на основе административных данных в Англии и Уэльсе (и, по сути, во всем Соединенном Королевстве) предстоит решить четыре основные задачи:

- ↑ получить доступ к широкому спектру данных, необходимых для формирования результатов, которые в настоящее время предоставляются в рамках десятилетней переписи населения;
- ↑ обеспечить точную увязку большого количества независимо собранных данных при одновременном сохранении конфиденциальности и безопасности данных;
- ↑ разработать методы преобразования увязанных данных в материалы, отвечающие потребностям пользователей;
- ↑ обеспечить приемлемость увязки и использования таких данных для основных заинтересованных сторон, например, благодаря повышению эффективности расходования средств и предоставлению гарантий того, что данные будут храниться в безопасной среде с сохранением конфиденциальности на протяжении всего периода применения данного подхода.

Для решения этих задач необходимо будет создать ряд условий.

Диаграмма G.1 Общая система проведения регистровых или комбинированных переписей Соединенного Королевства



Оперативный доступ к существующим и новым источникам данных

Для максимального расширения охвата и повышения качества статистических данных, которые могут быть получены в результате переписи на основе административных данных, УНС должно иметь оперативный доступ к новым и существующим источникам данных всех правительственных структур. Это также необходимо будет распространить на другие источники имеющихся данных, которые могли бы принести дополнительную пользу. Перед внесением изменений в административные данные, которые могут повлиять на качество и стабильность результатов переписи на основе административных данных с течением времени, необходимо будет также консультироваться с УНС. Закон о цифровой экономике, который вступил в силу в апреле 2017 года, предлагает решение по крайней мере в отношении некоторых из этих требований.

Способность производить эффективно и точно увязку данных

Все страны, отказавшиеся от проведения традиционной переписи раз в пять или десять лет, приняли на вооружение комбинированную или регистровую методологию проведения переписи, в основе которой лежит регистр населения и, как правило, система идентификационных номеров. Обычно это означает, что административные данные могут быть увязаны с регистром(ами) через единый идентификационный номер, что позволяет производить высокоточную увязку. Эти регистры также призваны обеспечить полный охват населения, что не всегда обеспечивают административные данные.

Соединенное Королевство не имеет ни идентификационных номеров, ни регистра населения. Вместо этого, как описано выше, перепись на основе административных данных предполагает увязку с помощью других средств многочисленных административных источников данных и обследований для формирования статистических данных по целому ряду признаков, которые в настоящее время формируются в рамках переписи. Это, конечно, непростая задача.

В Соединенном Королевстве лица не имеют единого идентификационного номера, который бы присутствовал в имеющихся у правительства данных, что усложняет эту увязку. Например, в данных о налогах и льготах соответственно Королевской налогово-таможенной службы (КНТС) и Министерства труда и пенсий (МТП) используется номер социального страхования, в то время как в данных Регистра врачей общей практики – номер Национальной службы здравоохранения (НСЗ), а в данных школьной переписи – уникальный идентификационный номер учащегося. Поэтому УНС необходимы методы, позволяющие увязку этих независимых источников данных для получения высококачественных статистических данных при одновременном сохранении конфиденциальности персональной информации и безопасности данных, как уже отмечалось выше.

Методы формирования статистических материалов достаточного качества для удовлетворения приоритетных информационных потребностей пользователей

Доступ к данным и их увязка – это только часть пазла. УНС необходимо разработать методы, которые могут преобразовывать увязанные административные данные и данные обследований в статистические материалы для удовлетворения приоритетных информационных потребностей пользователей. Это означает предоставление статистических данных по признакам, которые требуются пользователям, на должном уровне детализации (например, по малым районам) и надлежащего качества. В рамках проведенного в 2013 году опроса общественности пользователи указали, что любая такая статистическая методология должна обеспечивать предоставление:

- ↑ надежных оценок численности населения и количества домохозяйств;
- ↑ оценок характеристик населения в определенный момент времени, позволяющих сравнивать схожие районы друг с другом;
- ↑ информации со степенью детализации, необходимой пользователям для измерения изменений во времени (например, позволяющей определять изменения в уровне безработицы по этническому признаку на уровне малых районов за десятилетие).

Другой ключевой областью является разработка подробной программы требуемых обследований и методов моделирования на основе обследований и административных данных.

Приемлемость для заинтересованных сторон (пользователей, поставщиков, общественности и парламента)

Для успешного перехода к переписи на основе административных данных в следующем десятилетии пользователи данных, поставщики данных, общественность и парламент должны быть убеждены в том, что этот подход отвечает их потребностям и ограничивает расходы. Приемлемость для этих четырех ключевых заинтересованных сторон будет зависеть от обеспечения:

- ↑ удовлетворения основных информационных потребностей пользователей;
- ↑ хранения, обработки и уязки данных с соблюдением требований неприкосновенности частной жизни, конфиденциальности и безопасности; и
- ↑ значительного снижения расходов.

Эффективность затрат

Что касается конкретно третьего требования, то перепись на основе административных данных должна будет продемонстрировать, что она обеспечивает более эффективное использование средств по сравнению с десятилетней переписью. Это означает либо демонстрацию того, что она способна обеспечить выгоды, которые пользователи получают от десятилетней переписи населения, при меньших затратах, либо того, что экономия средств является достаточной для оправдания меньших выгод. Например, перепись на основе административных данных может оказаться не в состоянии обеспечить разработку всех материалов десятилетней переписи, но она может обеспечить дополнительные выгоды, такие как более своевременно и часто формируемые данные и новые материалы, которые в настоящее время не предоставляются в рамках десятилетней переписи. Это – ключевой компромисс, который необходимо будет принимать во внимание.

Как УНС узнает, возможна ли перепись на основе административных данных?

Для того чтобы правительство приняло решение о будущем переписей после 2021 года, УНС должно предоставить доказательства того, является или нет перепись на основе административных данных жизнеспособным подходом к проведению переписи. Для этого УНС планирует сделать следующее:

- ↑ Достичь прогресса в получении доступа к новым административным источникам данных, отдавая приоритет тем источникам данных, которые относятся к ключевым признакам, которые в настоящее время разрабатываются в ходе десятилетней переписи, или позволяют анализировать их. Что касается новых источников данных, то сопоставления на уровне записей можно провести с использованием данных переписи 2011 года, которая обеспечивает хорошую базу для сравнения статистического качества административных данных. Например, данные сопоставления могут помочь в определении того, не страдает ли административный источник проблемами охвата или обновления информации об адресах. Сопоставления с другими источниками данных могут также быть полезны для понимания статистического качества. УНС будет ежегодно публиковать обновленную информацию о прогрессе в получении данных.
- ↑ Публиковать на ежегодной основе результаты исследований по переписи на основе административных данных. Ежегодные результаты исследований будут демонстрировать тип и качество материалов, которые могут быть получены в рамках переписи на основе административных данных. На сегодняшний день УНС опубликовало результаты исследований по следующим вопросам:
 - численность населения (ONS, 2017a);
 - домохозяйства и семьи (ONS, 2017b);
 - характеристики населения (ONS, 2017c).

Круг вопросов будет расширяться в будущих публикациях в зависимости от наличия данных и их статистического качества. Одна из основных целей этих публикаций заключается в том, чтобы дать пользователям возможность высказать свое мнение о данных и используемых методах для того, чтобы помочь сфокусировать будущие разработки. УНС публикует краткое резюме отзывов, полученных по каждому набору результатов, демонстрируя пользователям, как учитываются их мнения и замечания.

- ↑ Проводить ежегодную оценку способности УНС осуществить переход к переписи на основе административных данных. Это в конечном итоге завершится сопоставлением результатов комбинированного использования административных данных и обследований с результатами переписи 2021 года для сравнительной оценки этого подхода. Итогом этого станет вынесение в 2023 году рекомендации о возможности перехода УНС к переписи на основе административных данных.
- ↑ Получить оценку методов и результатов исследований от внешней экспертной группы. Эти обзоры планируется провести в 2017, 2020 и 2022 годах.

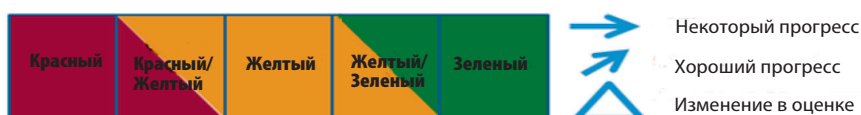
В июне 2017 года УНС опубликовало вторую оценку своего прогресса на пути перехода к переписи на основе административных данных после 2021 года (ONS, 2017d). Диаграмма G.2 иллюстрирует результаты этой оценки по состоянию на середину 2017 года. Полная оценка подкрепляется фактическими данными и описанием того, что будет сделано в будущем для улучшения оценки.

Диаграмма G.2 Текущая (2017 год) и ожидаемая в будущем (к 2023 году) оценка высокого уровня

Критерии оценки		Оценка 2016	Прогресс с 2016 года	Оценка 2017	Ожидаемый прогресс к 2018 году	Ожидаемая оценка к 2023 году
Доступ к данным						
Способность обеспечить увязку						
Способность обеспечить удовлетворение информационных потребностей пользователей	Оценки численности населения					
	Оценки семей и домохозяйств					
	Характеристики лиц и домохозяйств					
	Характеристики жилья					
Приемлемость для заинтересованных сторон						
Эффективность затрат						

Примечания к диаграмме

Значение КЖЗ



Справочная литература

Cabinet Office (2014). Government's response to the National Statistician's recommendation. Letter from the Rt Hon Francis Maude, MP, Minister for the Cabinet Office to Sir Andrew Dilnot, Chairman of the UK Statistics Authority. Скачать можно по адресу <https://www.statisticsauthority.gov.uk/archive/reports---correspondence/correspondence/letter-from-rt-hon-francis-maude-mp-to-sir-andrew-dilnot---180714.pdf>.

ONS (2013a). ONS Strategy, 2013-2023. Скачать можно по адресу https://www.ons.gov.uk/ons/dcp14298_323384.xml.

ONS (2013b). Beyond 2011: Matching Anonymous Data. Beyond 2011 Information Paper. Скачать можно по адресу <http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011-matching-anonymous-data--m9-.pdf>.

ONS (2013c). Beyond 2011: Producing population estimates using Administrative Data: In Theory. Beyond 2011 Information Paper. Скачать можно по адресу <http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011--producing-population-estimates-using-administrative-data--in-theory--m8-.pdf>.

ONS (2014). The Census and Future Provision of Population Statistics in England and Wales: Recommendation from the National Statistician and Chief Executive of the UK Statistics Authority. Скачать можно по адресу <http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/beyond-2011-report-on-autumn-2013-consultation--and-recommendations/national-statisticians-recommendation.pdf>.

ONS (2017a). Size of Population: Statistical Dataset Population (SDP) estimates for England and Wales, for selected years. ONS Administrative Data Census Research Outputs. Скачать можно по адресу <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/sizeofthepopulation>.

ONS (2017b). Households and Families. ONS Administrative Data Census Research Outputs. Скачать можно по адресу <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/householdsandfamilies>.

ONS (2017c). Population characteristics. ONS Administrative Data Census Research Outputs. Скачать можно по адресу <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/populationcharacteristics>.

ONS (2017d). Annual assessment of ONS's progress towards an Administrative Data Census post-2021. ONS Administrative Data Census annual assessments. Скачать можно по адресу <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusannualassessments>.

Приложение Н – Практический пример Италии

Справочная информация

Начиная с 2018 года Италия отказалась от традиционной десятилетней переписи методом обхода переписчиками в пользу «комбинированного» подхода, предусматривающего интеграцию административных данных и выборочных обследований. В 2012 году положение о так называемой «непрерывной» переписи населения и жилищного фонда (на итальянском языке *Censimento permanente della popolazione e delle abitazioni*) уже было введено в итальянское законодательство (статья 3 Законодательного декрета № 179/2012, который впоследствии с поправками был преобразован в Закон 221/2012). Целью проведения «непрерывной» переписи является ежегодное формирование данных, заменяющее собой перепись на основе десятилетнего цикла с использованием Интегрированной системы регистров (ИСП). Эта работа будет осуществляться в рамках программы модернизации ИСТАТ (Национальный институт статистики Италии), направленной на интеграцию административных данных и проведение вспомогательных статистических обследований, в соответствии с новой организационной, технологической и методологической моделью, ориентированной на полное задействование потенциала уже имеющихся данных. Новая стратегия переписи позволяет значительно сократить расходы на ее проведение, нагрузку на респондентов и организационную нагрузку на муниципалитеты (которые отвечают за сбор данных на местах).

В соответствии с итальянским законодательством в конце каждой переписи муниципальные служащие должны анализировать расхождения между итогами переписи и муниципальными регистрами населения, и регистры населения должны обновляться и пересматриваться на основе результатов переписи. Как правило, сообщалось о значительных расхождениях между численностью обычно проживающего населения, зарегистрированного в ходе переписи, и числом лиц, зарегистрированным в местных регистрах населения.

В ходе переписи 2011 года, несмотря на новую методологию (перепись проводилась на основе регистров населения) и технические решения (рассылка вопросника вместо доставки счетчиком, многорежимный сбор данных, включая веб-сбор), направленные на повышение качества и расширение охвата переписи, с регистрацией некоторых подгрупп населения по-прежнему возникли большие трудности. На основе проведенного в 2011 году после переписи населения обследования, в котором приняли участие более 320 000 человек, ИСТАТ сделал вывод о том, что переписью не было охвачено около 650 000 постоянных жителей Италии, из которых около 80% имели иностранное гражданство (Istat, 2016). Эти результаты, наряду со все менее стабильными расходами на проведение традиционной переписи, заставили ИСТАТ принять совершенно иную стратегию формирования переписных данных, основанную на комбинированном использовании множественных источников данных и ежегодных обследований (Crescenzi and Sindoni, 2015).

Основой новой переписи населения служит регистр населения. На деле, с 2011 года ИСТАТ занимается сбором данных о лицах и домохозяйствах из муниципальных регистров населения на ежегодной основе. Эти данные были использованы для рассылки переписных листов 2011 года домохозяйствам. Муниципальные регистры населения также служат основой выборки для социальных обследований (таких, как обследование рабочей силы, обследования ОДУЖ-ЕС и обследования бюджетов домохозяйств), а переменные, такие как место жительства, возраст, пол и гражданство, используются для стратификации и расчета отсутствующих ответов. Кроме того, итальянская государственная администрация работает над созданием централизованного регистра населения, так называемого «Национального регистра постоянного населения» (*Anagrafe Nazionale della Popolazione Residente*). В создаваемом централизованном регистре каждому лицу будет присвоен уникальный идентификационный номер, который позволит увязку регистра с другими административными источниками, что станет важным шагом на пути к проведению регистровой переписи.

Кроме того, путем интеграции сведений актов гражданского состояния (рождения, смерти, внутренняя и международная миграция) с микроданными переписи населения ИСТАТ осуществляет расчет и ведение ежегодной статистической совокупности, так называемой «ANagrafe Virtuale Statistica» (ANVIS). Эта

статистическая совокупность обеспечивает более высокий уровень качества данных, чем микроданные регистров населения, обеспечивая более прочную основу для проведения регистровой переписи. Данная совокупность в сочетании с другими соответствующими административными источниками (включая муниципальные регистры населения) служит основой Регистра населения, подлежащей проверке и корректировке на местах в рамках «непрерывной» переписи.

Разумеется, помимо разбивки населения по полу, гражданству, возрасту и месту рождения, перепись должна давать набор гиперкубов, требуемых в настоящее время Евростатом по социально-экономическим переменным (статус в занятости, уровень образования, миграционный статус и т.д.). С этой целью и для осуществления изменений, необходимых для интеграции европейских социальных обследований (EUROSTAT, 2013), ИСТАТ осуществляет интеграцию «непрерывной» переписи с Комплексной системой социальных обследований.

Интегрированная система переписи и социальных обследований

Интегрированная система переписи и социальных обследований (ИСПСО) представляет собой комплексный статистический процесс использования и интеграции информации, получаемой из регистров и собираемой в рамках обследований социально-экономических переменных. Она разработана в виде двухступенчатой модели, базирующейся на основной выборке (ОВ) и наборе сбалансированных и скоординированных выборочных обследований (ОПС, ОДУЖ и т.д.). Она предназначена для ведения регистра населения (РН) в целях увеличения объема предоставляемой статистической информации и улучшения охвата и качества.

В этом процессе РН служит базовым элементом системы формирования социальной статистики, в котором содержится отдельная строка по каждой целевой единице, т.е. *обычно проживающему лицу* (независимо от того, проживает ли оно в частном или институциональном домохозяйстве). В ИСП базовую информацию (взятую из РН) распространяют на все социальные переменные, включая статус в занятости и санитарные условия, благодаря совместному использованию административных источников и социальных обследований.

Для формирования переписных данных переменные, включенные в ИСП, классифицируются как *полностью заменяемые, частично заменяемые или не подлежащие замене*. К первой категории относят те переменные, по которым административные источники представляют соответствующую косвенную информацию и которые в конце статистического процесса, включающего редактирование и импутацию частично отсутствующих данных, считаются полными (поскольку они имеются по всем единицам РН), а также точными (т.е. имеющими хороший уровень охвата и качества). Например, пол и возраст являются переменными, сведения о которых имеются в отношении всех единиц РН, в связи с чем они считаются полностью заменяемыми.

Административные источники также предоставляют соответствующую *косвенную* информацию и по *частично заменяемым* переменным, которые, однако, считаются полными и точными только в отношении определенной подгруппы целевой совокупности. Что же касается остающейся подгруппы целевой совокупности, то эти переменные либо неизвестны, либо не могут считаться точными по причине неприменимости к ним синтетической модели импутации. В качестве примера можно привести переменную «занятость», которая является полностью заменяемой только в отношении подгруппы «регулярная занятость».

Наконец, по *не подлежащим замене* переменным соответствующая косвенная информация из административных регистров не доступна напрямую. Целевые параметры для этих переменных можно оценить с помощью выборочных обследований и использования вспомогательной информации из РН (т.е. набора переменных, содержащихся в регистре, которые, как считается, обладают предсказуемостью в отношении изучаемых переменных, не подлежащих замене). По определению, набор оценок должен будет удовлетворять требованиям:

- a) *надежности*, которая обеспечивается приблизительно несмещенной оценкой или основанным на модели методом, используемая модель которого является в некотором смысле правдоподобной. В обоих случаях коэффициенты вариаций должны быть ниже выбранного порогового значения; и
- b) *непротиворечивости* в том смысле, что данные, полученные путем комбинирования оценок различными способами, должны давать одинаковые результаты.

Основной функцией ИСПСО является восполнение пробелов в социально-экономической информации СИР для оценки тех целевых параметров, которые упоминались выше как *частично заменяемые* и *не подлежащие замене* переменные. С этой целью формируется ОВ для использования вместе (*объединения*) эффективным

образом всей общей информации (по целевым и вспомогательным переменным), собранной в рамках различных выборочных обследований, входящих в состав системы. Кроме того, стратегия оценки ОВ использует всю полную вспомогательную информацию РН. Эта стратегия позволяет получать более эффективные прямые оценки, чем оценки, полученные путем принятия отдельной стратегии оценки для каждого обследования. В этом контексте согласование единых переменных – *основных структурных* переменных (которые являются целевыми переменными для всех обследований) и *согласованных* переменных (которые являются целевыми переменными для более чем одного, хотя и не для всех обследований) – и унификация процессов статистического производства имеют крайне важное значение.

Первый этап формирования ОВ опирается на два различных ежегодных выборочных обследования: У (обследование участков) и С (списочное обследование). Что касается основных целей «непрерывной» переписи, то число жителей каждого муниципалитета рассчитывается на основе РН, проверяется и корректируется на местах с помощью выборочного обследования (это является основной целью так называемого обследования У). Данные, необходимые для заполнения гиперкубов переписи, оцениваются путем совместного использования информации, уже имеющейся в регистрах, и данных, собранных на местах (это является основной целью так называемого обследования С, проводимого, при необходимости, путем регулярного выборочного обследования домохозяйств).

Точнее говоря, выборочное обследование У, основанное на выборке переписных участков (ПУ) и/или адресов, отобранных из интегрированного адресного файла (ИАФ), предназначено для оценки показателей недостаточного охвата (НО) и чрезмерного охвата (ЧО) РН как на национальном, так и на местном уровне по различным профилям подгрупп населения, таким как различные комбинации признаков пола, возраста и гражданства. Эти показатели будут применяться к РН для получения взвешенных оценок численности населения с поправкой на ошибки охвата. Оценка численности населения производится с помощью расширенной системы двойной оценки (РСДО) с учетом как неполного, так и избыточного охвата.

Выборочное обследование С, основанное на списке домохозяйств, предназначено для тематической интеграции (ТИ), т.е. оценки гиперкубов, которые невозможно получить с использованием заменяемой информации из регистров. Возможность использования данных, собираемых с помощью обследования С, также для оценки пространственной вариации показателей недостаточного охвата (НО) РН находится в настоящее время в стадии рассмотрения. С другой стороны, выборочное обследование У предназначено также для достижения целевого показателя тематической интеграции (ТИ).

В более общем плане, первый этап разработки ОВ (т.е. два обследования: У и С) ориентирован на следующие цели:

- a1) получение выборочной информации по *частично заменяемым и не подлежащим замене* переменным;
- b1) установление первого контакта с включенными в выборку домохозяйствами, подвыборка которых будет повторно опрошена в следующем году на втором этапе (второй этап касается других обследований, таких как ОРС, ОДУЖ, ОБДХ). Первые опросы будут проводиться таким образом, чтобы уменьшить потенциальное непредставление ответов на втором этапе;
- c1) получение обновленной контактной информации (номера телефонов и адреса электронной почты), отсутствующей в рамках выборки, с тем чтобы на втором этапе можно было использовать менее дорогостоящие методы опроса (такие, как CAWI или CATI).

Из выборки первого этапа для второго этапа будет сформирован набор выборок домохозяйств с отрицательной координатой с целью:

- a2) предоставления информации по согласованным и специфическим социально-экономическим переменным, наблюдаемым в настоящее время в рамках обследования рабочей силы (ОРСТ), ОДУЖ-ЕС и обследования бюджетов домохозяйств (ОБДХ);
- b2) подтверждения общих структурных переменных, по которым уже была получена информация в ходе опроса первого этапа, и обеспечения непротиворечивости переменных, которые являются общими для вышеупомянутых социальных обследований. В настоящее время эти обследования опираются на стратифицированную двухступенчатую выборку (муниципалитеты–домохозяйства). Каждое обследование планируется и осуществляется отдельно на основе обследований различных планов выборки. По этим причинам могут наблюдаться значительные расхождения между оценками, касающимися одних и тех же переменных, которые наблюдались в рамках различных социальных обследований, даже если

определения и формулировки соответствующих вопросов являются одинаковыми. В этом случае, для того чтобы иметь возможность в будущем объединять ту же информацию, полученную в ходе различных обследований, стратегический вопрос будет заключаться в повышении согласованности между социальными обследованиями. По сути, одна из основных целей системы, описанной выше, заключается в том, чтобы уменьшить потенциальные системные расхождения между обследованиями путем согласования плана выборки.

Кроме того, выборка первого этапа (ОВ) была сбалансирована с использованием переменных РН для выявления тех территориальных или структурных подгрупп населения, которые создают проблемы охвата или подвержены структурным и характерным изменениям в течение краткосрочного периода.

«Непрерывная» перепись проводится в течение четырехлетнего цикла. Годовая выборка включает в себя около 950 000 домохозяйств в случае обследования С и 450 000 домохозяйств в случае обследования У, охватывая в общей сложности 2 852 муниципалитета (из 7 992). Саморепрезентативные (СР) муниципалитеты участвуют в переписи каждый год, в то время как несаморепрезентативные (НСР) муниципалитеты были разделены на 4 группы, каждая из которых участвует в переписи раз в 4 года. В каждом муниципалитете оба обследования проводятся параллельно в отдельных частях муниципальной территории, в то время как в муниципалитетах с населением менее 1 000 жителей на сплошной основе проводится только обследование У³⁷.

Общая схема ИСПСО, основанной на двухступенчатой модели ОВ, изображена на диаграмме Н.1. Административные записи в основном поддерживают формирование основы выборки переписи населения (ОВПН), которая интегрирует РН с другими источниками (главным образом с Налоговым регистром). Выборка обследования С формируется на основе ОВПН, в результате чего окончательными единицами выборки (ОЕВ) являются домохозяйства, входящие в ОВПН.

Обследование У основано на выборке, в которой ОЕВ являются ПУ и/или адресами интегрированной адресной основы выборки (ИАОВ). ИАОВ получают путем интеграции адресов, входящих в ОВПН, с адресами, касающимися новых зданий.

Основное различие между обследованием С, в котором домохозяйства используются в качестве ОЕВ, и обследованием У, в котором в качестве ОЕВ используются адреса, заключается в том, что последние являются «слепыми» в отношении информации и единиц, входящих в ОВПН. Таким образом, будут полностью выполнены необходимые условия, лежащие в основе РДО.

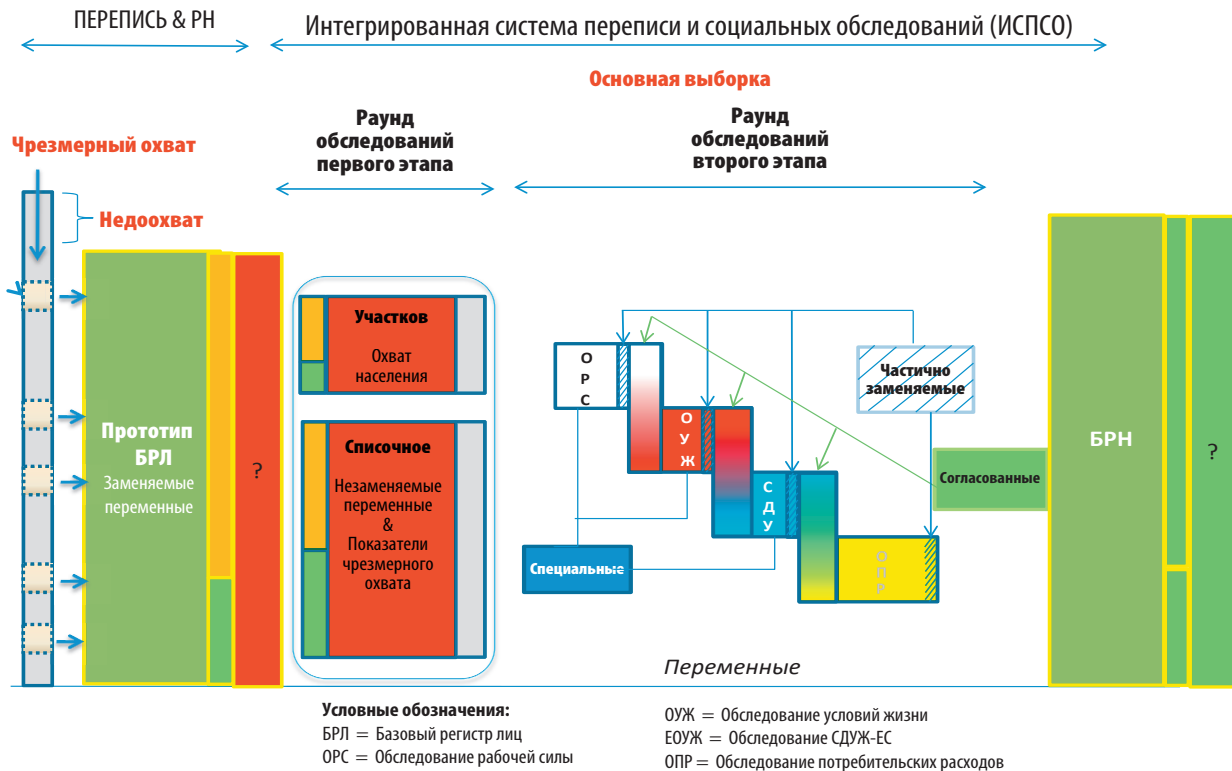
Что касается аналогичных международных инициатив, то при изучении модульного подхода к разработке плана комплексных социальных обследований Евростат предложил проекты, аналогичные итальянской модели общего плана основной выборки для социальных обследований. Кроме того, Австралийское бюро статистики (АБС) изучило возможность разработки комплексной системы обследований, очень похожей на описанную здесь модель. Однако данная система, получившая название Обследование населения Австралии, не призвана заменить собой перепись.

Кроме того, итальянская модель с двумя компонентами, поддерживающими регистровую перепись, аналогична методологии, разработанной Управлением национальной статистики (УНС) Соединенного Королевства для предлагаемой им переписи на основе административных данных для возможного внедрения в 2023 году после переписи 2021 года (ONS, 2017). В 2021 году УНС проведет традиционную перепись и параллельную перепись на основе создания интегрированного регистра населения с использованием ряда административных источников данных и результатов двух обследований, аналогичных по своим характеристикам с компонентами С и У стратегии, принятой в Италии. Следует отметить, что ежегодно в период с 2015 по 2023 год УНС будет производить расчеты для оценки отклонения от будущей модели (более подробную информацию см. в приложении G).

Израильская скользящая интегрированная перепись представляет собой еще одну модель, которая также демонстрирует сходство с тем, что планируется делать в Италии. Центральное бюро статистики использует интегрированный регистр, который корректируется с помощью весов, рассчитанных с применением ИСПСО (Pfeffermann, 2015).

³⁷ Муниципалитетами СР являются муниципалитеты с населением > 17 800 человек плюс муниципалитеты, которые не подлежат ротации в выборке ОРС. Четырехгодичные группы муниципалитетов НСР формируются таким образом, чтобы по каждой провинции в них ежегодно были представлены муниципалитеты каждого класса размера.

Диаграмма Н.1 Общая схема ИСПСО



Для определения окончательного плана обследования ИСПСО ИСТАТ провел пилотное обследование в 2017 году. Основные цели обследования заключались в проверке качества работы ИАФ, оценке показателей участия на первом и втором этапах и уточнении оперативных аспектов. Исходя из его результатов в начале октября 2018 года начался этап переписи населения первого (2018) года первого цикла (2018–2021 годы).

Справочная литература

Crescenzi, F and Sindoni, G (2015). The combined use of multiple data sources in the population census. Документ, представленный на совещании Группы экспертов по переписям населения и жилищного фонда ЕЭК ООН. Женева, 30 сентября – 2 октября 2017 года, http://www.unecce.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2015/mtg1/CES_GE.41_2015_7-Istat_rev.pdf.

Di Bella, G and Ambroselli, S (2014). Towards a more efficient system of administrative data management and quality evaluation to support statistics production in ISTAT. Paper presented at the European Conference on Quality in Official Statistics Q2014. Vienna, 2-5 June 2014.

EUROSTAT (2013). D12. (2013). Roadmap for the integration of European social surveys, http://ec.europa.eu/eurostat/cros/sites/crosportal/files/D12_Roadmap.pdf.

Istat (2016). Atti del 15° Censimento generale della popolazione e delle abitazioni. 6 La valutazione della qualità, http://www.istat.it/it/files/2016/07/AttiPOP_Fascicolo-6-web.pdf.

ONS (2017). Annual assessment of ONS's progress towards an Administrative Data Census post-2021. <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusannualassessments>.

Pfeffermann, D (2015). Methodological Issues and Challenges in the Production of Official Statistics, *Journal of Survey Statistics and Methodology*, 3: 425–483.

Приложение I – Практический пример Германии

Введение

Внедрение комбинированной модели переписи населения в Германии было непростым по нескольким причинам. Во-первых, по сравнению с другими странами, применяющими подходы, основанные на регистрах, в стране имеется меньше подходящих регистров. Во-вторых, строгие правила защиты данных, которые были приняты в контексте последней традиционной немецкой переписи 1987 года, затрудняют нахождение решений для увязки регистров из различных областей. В силу этих правил до сих пор в стране не были введены ни личные идентификационные номера, ни идентификационные номера жилищ, что делает любую увязку регистров обременительным мероприятием.

В то же время традиционная перепись, основанная на сборе данных переписчиками, не пользуется популярностью среди заинтересованных сторон по причине огромных затрат. Это является еще более важным фактором, поскольку ввиду федеральной структуры официальной статистики в Германии расходы должны покрываться совместно федеральным правительством и правительствами 16 регионов (земель), что в прошлом вызывало длительные споры.

С 1983 года традиционный сбор данных в рамках переписи также пользовался лишь ограниченной популярностью среди респондентов. Последняя традиционная перепись населения в Германии была в конечном итоге проведена в 1987 году, но и то с продолжительной задержкой. С учетом этого особого контекста в Германии была разработана комбинированная модель переписи, прошедшая тестирование в рамках крупномасштабной проверки в 2001 году и, наконец, внедренная в 2011 году.

Правовая и институциональная основа

Разработка комбинированной модели немецкой переписи населения может быть понята только на основе контекста последней традиционной переписи, которая была отменена Конституционным судом Германии всего за несколько недель до ее проведения в 1983 году (краткую историю переписей населения Германии до 1980-х годов см. в работе Scholz and Kreyenfeld, 2016). Она была проведена в модифицированной форме в 1987 году. После сопровождавшегося острой полемикой обсуждения вопросов переписи в период 1983–1987 годов федеральное правительство не испытывало желания вновь поводить традиционную перепись. Таким образом, вместо проведения сплошной переписи в рамках раунда переписей 2001 года была проведена крупномасштабная/пробная перепись для оценки пригодности регистрового подхода, в рамках которой данные, полученные из регистров, комбинировались с рядом наборов первичных данных.

Новая модель должна была соответствовать решению Конституционного суда Германии, вынесенному в отношении запланированной на 1983 год переписи населения. С тех пор это решение оказало большое влияние на регулирование защиты данных в Германии. В нем было указано, что право на информационное самоопределение непосредственно вытекает из основного права на личную свободу, гарантированного статьей 2 Конституции. Поэтому любой сбор данных, требуемый от общественности, считается конституционным только в том случае, если он имеет правовое обоснование, которое должно быть конкретным и ясным, а также соизмеримым с соответствующими общественными интересами. Хотя данные для административных целей могут собираться только в конкретных, хорошо обоснованных и соразмерных целях, сбор официальной статистики, учитывая ее особую роль, допускается в отношении определенного объема информации, который может быть использован для различных целей. Следовательно, данные, собранные для целей статистики, должны использоваться только в статистических целях и ни при каких обстоятельствах не могут быть переданы другим государственным органам («Rückspielverbot»). В более общем плане, любая увязка регистров, позволяющая составить указатель лиц, была запрещена как неконституционная. Аналогичным образом Конституционный суд подчеркнул, что универсальный идентификационный номер,

который облегчает увязку данных, хранящихся в различных регистрах, также был признан неконституционным (Bundesverfassungsgericht, 1983).

Комбинированная модель переписи населения и ее внедрение в 2011 году

Идея, которая лежит в основе комбинированной модели переписи населения в Германии, заключается в использовании данных в области демографии и занятости из имеющихся административных регистров (например, регистров населения, которые ведутся муниципалитетами, и регистра статистики занятости Федерального агентства занятости). В сочетании с полной переписью зданий и жилищ (поскольку по таким единицам отсутствовала достаточная регистровая информация) и дополнительным выборочным обследованием (по касающимся лиц переменным, отсутствующим в регистрах), предстояло создать набор данных переписного типа. Пробная перепись 2001 года показала, что для исправления ошибок, обнаруженных в регистрах, необходимо также использовать дополнительный выборочный опрос (см. Statistische Ämter des Bundes und der Länder, 2004).

Главная цель переписи 2011 года заключалась в использовании демографической информации, имеющейся в децентрализованных регистрах населения, и в дополнении и, при необходимости, корректировке этих данных путем их объединения с информацией из других регистров и обязательных первичных обследований. (Нижеследующая дискуссия основана на документах, представленных Бехтольдом в 2013 и 2016 годах; более подробное описание методологии приводится в Statistische Ämter des Bundes und der Länder, 2015). Путем комбинирования различных источников данных и с помощью методов автоматического формирования данных можно создать отдельный набор данных, содержащий всю необходимую переписную информацию по каждому лицу, каждому домохозяйству и каждому зданию с жилищами.

Для объединения данных различных частей набора переписных данных сначала был создан базовый регистр, содержащий список всех адресов, по которым на день проведения переписи существовали здания с жилой площадью. Этот регистр адресов и зданий служил ключевым инструментом для сбора всех данных в ходе переписи. Он также использовался в качестве статистической совокупности для процедуры формирования выборки частных домохозяйств и переписи жилищного фонда.

Основными источниками данных, использовавшимися в комбинированной модели (иллюстрацией которой служит диаграмма I.1), являлись:

- ↑ *Регистры населения, содержащие основные демографические данные, а также информацию о семейных связях по всем лицам, относящимся к целевой совокупности (около 86 млн записей данных). Данные муниципальных регистров населения были собраны по состоянию на критический момент переписи (9 мая 2011 года) и обновлены через три месяца с целью учета задержек в регистрации и снятии с регистрации. Данные регистров были объединены в общенациональный*

Диаграмма I.1 Немецкая модель переписи населения 2011 года



набор данных, который впоследствии был подвергнут проверке для определения того, были ли люди зарегистрированы более чем по одному единственному или основному месту жительства по состоянию на критический момент переписи. Если такие случаи выявлялись в крупных муниципалитетах (с населением не менее 10 000 человек), они автоматически корректировались с использованием самой последней информации. Множественные места жительства в малых муниципалитетах (с населением менее 10 000 человек) проверялись с помощью опроса по почте. То же самое относится и к случаям, когда лицо было зарегистрировано только по вторичному месту жительства.

- ↑ *Дополнительное выборочное обследование домохозяйств*, охватывающее почти 10% населения, было использовано для корректировки регистровых данных в муниципалитетах с населением 10 000 человек и более после того, как регистры были скорректированы по множественным местам проживания. При расчете численности населения крупных муниципалитетов учитывался уровень ошибок в регистрах населения (завышение и занижение охвата), выявленных в ходе обследования домохозяйств. Выборка была сформирована таким образом, чтобы данные о численности населения крупных муниципалитетов удовлетворяли целевой предельной ошибке выборки в размере 1% при уровне достоверности 95%. Метод, применявшийся для оптимизации процесса выборки, разрабатывался индивидуально для каждого муниципалитета, и размер выборки варьировался от 2,1% до 45,6% и значительно различался даже в случае муниципалитетов схожего размера. В случае муниципалитетов с населением менее 10 000 человек был проведен опрос тех домохозяйств, которые были определены как требующие уточнения, путем комбинирования и анализа регистровой информации и информации, полученной в ходе обследования жилищного фонда.

Помимо цели расчета численности населения, дополнительное обследование домохозяйств было также использовано для сбора данных по дополнительным переписным переменным, требуемым в соответствии с законодательством ЕС, которые не были включены в регистры (в частности, касающимся участия в рынке труда и уровня образования). Данные по дополнительным признакам переписи были собраны во всех муниципалитетах (не только с числом 10 000 или более). Размер выборки был призван обеспечить публикацию на уровне КТЕС-3.

- ↑ В отношении лиц, проживающих в специальных учреждениях, таких как коммунальные заведения, учреждения по уходу, общежития или аналогичные типы жилья, переписная информация собиралась методом сплошной переписи, поскольку расхождения и отсутствие регистрации по этой подгруппе населения в регистрах населения вели к высокому уровню погрешности. Адреса, которые были сочтены деликатной или потенциально стигматизирующей информацией, например психиатрические больницы или тюрьмы (и указывавшиеся как «конфиденциальные специальные учреждения»), были обособлены от неконфиденциальных специальных учреждений, таких как студенческие общежития. В специальных конфиденциальных учреждениях конфиденциальность сбора данных обеспечивалась специальной процедурой, и сбор данных проводился лишь по ограниченному набору переменных.
- ↑ Из-за отсутствия регистров зданий и жилищ, охватывающих всю Германию, данные по обязательным переменным переписи жилищного фонда ЕС пришлось получать путем проведения почтового обследования зданий и жилищ, которым были охвачены все владельцы объектов недвижимости (в отношении в общей сложности чуть менее 20 млн зданий с жилыми помещениями данные были собраны от примерно 19 млн владельцев). Кроме того, перепись зданий и жилья охватывала вспомогательные переменные (число лиц, проживающих в жилище, и имена двух жильцов), которые использовались в процессе формирования домохозяйств (см. ниже).
- ↑ Значительная часть информации о занятости населения была взята из регистров Федеральной службы занятости (около 36 млн работников, уплачивающих взносы по социальному страхованию) и из административных файлов государственных служб (около 3 млн государственных служащих, судей и военнослужащих). Эти регистры также использовались для дополнения демографической информации, полученной из регистров населения, результатов выборочного обследования домохозяйств и обследования адресов в неконфиденциальных специальных учреждениях. В сочетании с регистром адресов и зданий эта информация составляла справочный массив данных.
- ↑ Для получения информации о структуре домохозяйств и семей и их жилищных условиях (такая информация не включена ни в один регистр) пришлось комбинировать данные по различным

компонентам переписи с помощью так называемой процедуры формирования домохозяйств. В рамках этой многоступенчатой процедуры информация о лицах, полученная из регистров населения, результатов выборочного обследования домохозяйств и обследования, проведенного в специальных учреждениях, использовалась для формирования домохозяйств и увязки их с жилищами с помощью данных переписи жилищного фонда.

Объединение наборов данных из различных источников по отдельным лицам стало одним из главных вызовов переписи населения 2011 года, поскольку это пришлось делать в отсутствие уникальных персональных идентификационных номеров, имеющих в различных регистрах. Один уже существующий набор идентификационных номеров для целей налоговых органов имелся в некоторых регистрах, но не мог быть использован в силу юридических ограничений. Поэтому для увязки соответствующих записей различных наборов данных использовалась персональная и адресная информация, такая как имя, пол, дата рождения, код муниципалитета, почтовый индекс, название улицы и номер дома.

Извлеченные уроки и изменения в переписи населения 2021 года

Задним числом можно отметить, модульная концепция немецкой переписи 2011 года, сочетающая использование регистров и первичных статистических данных, оказалась успешной. Качественная оценка показала, что эта модель хорошо зарекомендовала себя и (с учетом предложений по ее совершенствованию) должна служить основой для проведения переписи в 2021 году. В то же время был сделан вывод о том, что качество регистров населения все еще не считается достаточным для целей переписи, в связи с чем вновь будет необходимо провести дополнительное выборочное обследование домохозяйств в качестве механизма корректировки избыточного и недостаточного охвата в регистрах (и для сбора данных по переменным, не охваченным регистрами). Таким образом, анализируя результаты переписи 2011 года, можно сделать вывод о том, что выборочное обследование домохозяйств остается неотъемлемой составной частью оптимизации модели переписи. Модель была сочтена научно приемлемой и обеспечила высокую точность. В то же время данная модель значительно снизила нагрузку на респондентов и затраты на сбор данных для статистических управлений по сравнению с предыдущей традиционной переписью. Еще одним новым элементом, который оказался приемлемым и точным способом определения данных о семьях и домохозяйствах на уровне локальных единиц, стала процедура формирования домохозяйств.

Поэтому изменения, внесенные в перепись 2021 года, учитывают опыт проведения переписи 2011 года с точки зрения аспектов качества и возможности снижения сложности и, следовательно, обеспечения более своевременных результатов. Кроме того, изменения направлены на то, чтобы облегчить понимание результатов и повысить общую приемлемость модели (как для пользователей, так и для общественности). Основные изменения, предложенные для модели переписи 2021 года (изображенной на диаграмме I.2), являются следующими:

Диаграмма I.2 Модель переписи 2021 года



- ↑ Взаимодействие различных компонентов переписи необходимо планировать на раннем этапе, с тем чтобы обеспечить комплексный технический подход, объединяющий индивидуальные компоненты модели. Поэтому результаты различных компонентов обследований и переписи будут объединены в центральный массив данных вместо того, чтобы хранить их отдельно, как это было в 2011 году. При этом данные могут быть перепроверены и подтверждены на ранней стадии обработки. Расхождения и неправдоподобия можно устранить с помощью правил или даже ручной проверки. Это поможет как повысить качество данных, так и сократить объем ресурсов, необходимых для их увязки.
- ↑ Использование бумажных вопросников будет сокращено с помощью жесткой стратегии «приоритет режиму онлайн». Это является важным компонентом дальнейших усилий по сокращению расходов, повышению своевременности и минимизации нагрузки на респондентов за счет упрощения перемещения респондентов по вопроснику.
- ↑ Построение регистра адресов начнется раньше, и один из источников данных больше не будет использоваться. В 2011 году для сбора информации об адресах были задействованы три основных источника: данные регистров населения, Федерального картографического агентства и регистры Федерального агентства занятости. Последние не будут использоваться в 2021 году, поскольку с 2011 года в этот источник не было добавлено дополнительных адресов, в то время как многие перекрестные проверки оказались необходимыми по причине различий в написании названий городов, улиц и номеров домов.
- ↑ В 2011 году данные Федерального агентства занятости также использовались для формирования данных о занятости населения. Эти данные были высокого качества, однако пользователи жаловались на сложность анализа, поскольку расходящиеся данные о занятости публиковались в зависимости от того, были они основаны только на комбинированной модели или только на обследовании домохозяйств. При рассмотрении занятости в более широком смысле этот источник пришлось анализировать в сочетании с обследованием домохозяйств, чтобы охватить также самозанятых или безработных. Поскольку эта дополнительная сложность сбивала с толку многих пользователей, использование регистров Федерального агентства занятости в 2021 году будет прекращено.
- ↑ Многие муниципалитеты подвергли критике различные модели корректировки числа людей, зарегистрированных в крупных городах (с населением 10 000 человек и более) и малых городах. Несмотря на то, что различные модели давали результаты ожидаемого с самого начала качества, использование таких различных моделей послужило препятствием для принятия их пользователями. Кроме того, с учетом ретроспективного опыта пришлось оптимизировать модель корректировки регистров населения в малых муниципалитетах.
- ↑ Схема взвешивания дополнительного обследования домохозяйств была ориентирована главным образом на высокоточное число жителей. Формирование результатов по переписным переменным, которые отсутствовали в регистрах, было сочтено лишь второстепенным приоритетом при разработке оценок. Процедура взвешивания должна быть оптимизирована, с тем чтобы свести к минимуму любой риск смещения переписных переменных, отсутствующих в регистрах.

Справочная литература

Bechtold S. (2013): Lessons learned from a mixed-mode census for the future of social statistics. Paper presented at the 97th DGINS Conference on 'New conceptual design for household and social statistics', Wiesbaden, 27 September 2011. Имеется по адресу https://www.destatis.de/EN/AboutUs/Events/DGINS/Document_Destatis_MixedModeCensus.pdf?__blob=publicationFile (дата посещения 14 июля 2017 года).

Bechtold S. (2016): 'The 2011 Census Model in Germany'. In: Comparative Population Studies D1-D9 (date of release: 04.08.2016). Имеется в DOI: 10.12765/CPoS-2016-07en (дата посещения 14 июля 2017 года).

Bundesverfassungsgericht (1983): Urteil des Ersten Senats vom 15. Dezember 1983 auf die mündliche Verhandlung vom 18. und 19. Oktober 1983. (BVerfG 65, 1) Имеется по адресу https://www.zensus2011.de/SharedDocs/Downloads/DE/Gesetze/Volkszaehlungsurteil_1983.pdf?__blob=publicationFile&v=9 (дата посещения 4 июля 2017 года).

Scholz, Rembrandt D and Kreyenfeld, Michaela (2016): 'The register-based census in Germany: historical context and relevance for population research'. In: Comparative Population Studies - Zeitschrift für Bevölkerungswissenschaft 41 (2016), 2, pp. 175-204. Имеется в DOI: <http://dx.doi.org/10.12765/CPoS-2016-08en> (дата посещения 14 июля 2017 года).

Statistische Ämter des Bundes und der Länder (2004): Results of the census test. Statistisches Bundesamt: Wiesbaden.

Statistische Ämter des Bundes und der Länder (2015): Zensus 2011. Methoden und Verfahren. Statistisches Bundesamt: Wiesbaden. Имеется по адресу https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze_Archiv/2015_06_MethodenUndVerfahren.pdf?__blob=publicationFile&v=6 (дата посещения 14 июля 2017 года).

Глоссарий терминов, определений и сокращений

АБМ: Аналитическая база микроданных (Польша)

АБС: Австралийское бюро статистики

Административные данные: массивы данных, содержащие собранную главным образом для административных (не исследовательских или статистических) целей информацию. Данные такого типа собирают государственные ведомства и другие организации с целью регистрации, учета и проведения операций, обычно при предоставлении каких-либо услуг

Вероятностный метод: метод, включающий в себя случайный компонент, который, таким образом, не всегда приводит к одному и тому же результату

Воеводство: один из 16 административных районов Польши (эквивалент провинции или области)

Гиперкуб: многомерная статистическая таблица, состоящая, как правило, из четырех и более измерений

ГИС: географическая информационная система

Гмина: одна из 2 478 базовых единиц административно-территориального деления Польши (эквивалентна муниципалитету)

ГСПН: Генеральная совокупность переписи населения (Италия)

Детерминистский метод: метод без введения случайного компонента, который, таким образом, всегда приводит к одному и тому же результату

Доступность: показатель качества данных, характеризующий условия и способы получения, использования и интерпретации данных пользователями

ДСЗ: Департамент социальной защиты (Ирландия)

ДСО: двойная система оценки – статистический метод, основанный на проведении первичного и повторного обследования, используемый для оценки численности населения

ЕАСТ: Европейская ассоциация свободной торговли

Евростат: Статистическое управление Европейского союза

ЕС: Европейский союз

ЕСС: Европейская статистическая система

ЕЭК ООН: Европейская экономическая комиссия Организации Объединенных Наций

ИАОВ: Интегрированная адресная основа выборки (Италия)

ИР: Иммиграционный регистр (Португалия)

ИСМ: Интегрированная система микроданных (Италия)

ИСП: ИТ-система переписи (Польша)

ИСПСО: Интегрированная система переписи и социальных обследований (Италия)

ИСТАТ: Национальный институт статистики Италии

ИТ: информационная технология

Качество процесса: качество статистического процесса, оцениваемого по используемым методам, эффективности затрат и нагрузке на респондентов

КЕС: Конференция европейских статистиков

КНТС: Королевская налогово-таможенная служба (Англия и Уэльс)

Комбинированная перепись: перепись, в которой определенная информация о численности и характеристиках населения берется из административных источников данных, ведущихся в нестатистических целях, а другая информация, отсутствующая в таких источниках, собирается непосредственно от лиц и домохозяйств с помощью полной или частичной регистрации переписчиками или других выборочных обследований

ЛАЕ: локальная административная единица. Классификация локальных административных районов, используемая Евростатом. Уровни 1 и 2 соответствуют тем районам, которые ранее были классифицированы соответственно на уровне 4 и 5 Классификации территориальных единиц для целей статистики (КТЕС)

Метаданные: информация о содержании, структуре, качестве и других соответствующих характеристиках регистра

Микроданные: в настоящих Руководящих принципах обозначают информацию в регистре, относящуюся к одной единице или нескольким единицам

МОВ: местный орган власти (Англия и Уэльс)

МТП: Министерство труда и пенсий (Англия и Уэльс)

НИД: Национальная инфраструктура данных (Ирландия)

НР: Налоговый регистр (Австрия)

НСЗ: Национальная служба здравоохранения (Англия и Уэльс)

НСС: Национальная статистическая система (Португалия)

НСУ: Национальное статистическое управление

Обезличивание: процесс защиты конфиденциальности персональной информации путем удаления всех уникальных идентификаторов из сведений об учетных единицах

ОБМ: Оперативная база микроданных (Польша)

ОБОН: обследование охвата населения (Англия и Уэльс)

ОВ: основная выборка (Италия)

ОЕВ: окончательные единицы выборки (Италия)

ОРС: обследование рабочей силы

ОЧН: оценка численности населения (Португалия)

Перепись де-факто: перепись, основанная на регистрации лиц по месту их пребывания на контрольную дату переписи

Перепись де-юре: перепись, основанная на подсчете лиц по месту их обычного жительства на контрольную дату переписи

ПЖ: признаки жизни

ПИН: персональный идентификационный номер

ПНЖФ: перепись населения и жилищного фонда

ПП: пробная перепись

Пунктуальность: показатель качества данных, характеризующий расхождение между датой публикации результатов и целевой датой (датой, к которой должны были быть представлены данные)

РАГС: Регистр альтернативной гражданской службы

РАЛ: Регистр активности лиц (Ирландия)

РБ: Регистр безработных (Австрия)

РВА: Регистр владельцев автомобилей (Австрия)

РГС: Регистры государственных служащих федерального государства и земель (Австрия)

РД: Регистр домохозяйств (Словения)

РДП: Регистр детских пособий (Австрия)

Регистр активности: регистр, содержащий информацию о различных формах активности резидентов, указывающих на их присутствие в стране или районе. Под такой активностью может пониматься, например, занятость или иной экономический статус, получение пособий или пенсий, или статус учащегося

Регистр населения: регистр жителей страны

Регистр: систематизированный набор данных уровня единиц наблюдения, способ организации которых делает возможным их обновление. Обновление – это обработка поддающейся идентификации информации с целью создания, актуализации, исправления или расширения регистра, т.е. отслеживание любых изменений в данных, описывающих единицу наблюдения или ее признаки

Регистровая перепись: перепись, в которой источником данных о численности и характеристиках населения служит информации из административных источников, предназначенных для нестатистических целей. Никакая информация не собирается напрямую у лиц или домохозяйств

Релевантность: показатель качества данных, характеризующий, в какой степени статистические данные отвечают текущим и потенциальным потребностям пользователей

РЗЖ: Регистр зданий и жилищ (Австрия)

РИК: Защищенный идентификационный ключ (Ирландия)

РН: Регистр народонаселения

РН: Регистр недвижимости (Словения)

РП: Регистр предприятий и местных подразделений (Австрия)

РП: Регистр призывников (Австрия) РАГС: Регистр актов гражданского состояния (Португалия)

РПСП: Регистр получателей социальной помощи (Австрия)

РСДО: Расширенная система двойной оценки (Италия)

РУО: Регистр уровня образования (Австрия)

РУС: Регистр учащихся и студентов (Австрия)

Своевременность: показатель качества данных, характеризующий период между моментом получения информации и событием или явлением, которое она описывает

СДУЖ-ЕС: статистика доходов и условий жизни Европейского союза

СИН: Статистический идентификатор (Словения)

Синхронизация: передача данных, записанных на портативных устройствах на местах, на центральный сервер

СНДЖ: Статистический набор данных о жилищах (Португалия)

СНДН: статистический набор данных о населении

Согласованность: показатель качества данных, характеризующий степень, в которой данные переписи могут комбинироваться различными путями и для разных целей со статистической информацией из других источников

Сопоставимость: показатель качества данных, характеризующий степень, в которой статистические данные сопоставимы между географическими районами и во времени

Статистический регистр: регистр, созданный для статистических целей. Обычно он создается путем преобразования данных из регистров и/или других административных источников данных

СУП: Статистическое управление Португалии

СУРС: Статистическое управление Республики Словения

Точность: показатель качества данных, характеризующий степень близости оценок к неизвестным «истинным» значениям

Традиционная перепись: перепись, основанная на прямом подсчете численности населения и сборе данных о его характеристиках посредством заполнения переписных листов самостоятельно респондентами или в ходе их опроса в бумажной форме или в электронном виде

УЗД: Орган по защите данных (Австрия)

Уникальный идентификатор/ключ: единый буквенно-цифровой идентификатор, увязывающий характеристику или переменную с определенной единицей (лицом, домохозяйством или жилым помещением) в самых разных регистрах или административных источниках данных

УНС: Управление национальной статистики (Англия и Уэльс)

ФАО: Продовольственная и сельскохозяйственная организация Объединенных Наций

Цифровой адрес: код, связывающий с адресом

ЦРН: Центральный регистр населения (например, в Австрии и Словении)

ЦРСО: Центральный регистр социального обеспечения (Австрия)

ЦСБ: Центральное статистическое бюро (например, в Израиле)

ЦСУ: Центральное статистическое управление (например, в Ирландии и Польше)

ЦХМД: Центральное хранилище метаданных (Польша)

ANVIS: ANagrafe Virtuale Statistica – ежегодная статистическая совокупность (Италия)

ьPIN OC: Отраслевой персональный идентификационный номер для официальной статистики (Австрия)

ьPIN: Отраслевой персональный идентификационный номер (Австрия)

CAII: автоматизированный опрос через Интернет

CAPI: компьютеризованный личный опрос

CATI: компьютеризованный опрос по телефону

CAWI: компьютеризованный веб-опрос

CAxi: компьютеризованный многорежимный опрос (Польша)

EHIS: информационная система образования (Эстония)

ID: идентификационный номер

NIC: номер Регистра актов гражданского состояния (Португалия)

NIF: номер налогового регистра (Португалия)

NISS: номер социального страхования (Португалия)

PAYE: сбор налога у источника (Англия и Уэльс)

PPSN: персональный номер Службы государственного Обеспечения (Ирландия)

UPRN: уникальный идентификационный номер объекта недвижимого имущества (Великобритания)

XML: расширяемый язык разметки. В сфере обработки данных этот язык определяет набор правил кодирования документов в формате, который является удобочитаемым для людей и машин

XSD: расширенные сервисы данных (Эстония)

Руководящие принципы использования регистров и административных данных в целях переписей населения и жилищного фонда

Существует много различных способов проведения переписи населения и жилищного фонда. Традиционный подход, основанный на прямом подсчете всех лиц и сборе информации с помощью переписных листов, представляет собой сложную и дорогостоящую операцию. В последние годы все большее число стран переходят на альтернативные методы проведения переписи. В большинстве случаев эти методы опираются в максимально возможной степени на использование имеющихся регистров и административных данных, что ограничивает или даже сводит к нулю прямой сбор информации для переписи.

Цель настоящей публикации заключается в предоставлении национальным статистическим управлениям и другим ответственным учреждениям рекомендаций по использованию регистров и административных данных в целях переписей, включая оперативные, практические, технические и правовые аспекты. Эта публикация также, как ожидается, будет полезна и для пользователей результатов переписи.

Настоящая публикация была подготовлена целевой группой, учрежденной Конференцией европейских статистиков (КЕС) и состоящей из экспертов национальных статистических управлений, работа которой координировалась Европейской экономической комиссией Организации Объединенных Наций. КЕС одобрила данную публикацию на своей пленарной сессии 2018 года.

Information Service
United Nations Economic Commission for Europe

Palais des Nations
CH - 1211 Geneva 10, Switzerland
Telephone: +41(0)22 917 12 34
E-mail: unece_info@un.org
Website: <http://www.unece.org>