



ЭКОНОМИЧЕСКИЙ И СОЦИАЛЬНЫЙ СОВЕТ

Распространение: Общее
30 апреля 2012 г.
Оригинал: английский

Европейская экономическая комиссия ООН

Конференция европейских статистиков

60-е пленарное заседание

Париж, 6-8 июня 2012

2010 раунд переписи – инновации и уроки

Уроки переписи населения и жилого фонда в Польше

Записка Центрального статистического управления Польши

Резюме

В этом документе представлен опыт Центрального статистического управления Польши по проведению Национальной переписи населения и жилого фонда в 2011 году.

Проведение переписей требует больших средств и значительного участия граждан. В данном материале показано, как Центральному статистическому управлению Польши удалось сократить расходы на перепись и снизить нагрузку на респондентов при сохранении высокого качества результатов.

Здесь описаны решения в области технологий и программного обеспечения, использованные для объединения данных из разных источников, включая административные регистры, замены бумажных переписных листов на электронные вопросники и использования передовых информационных технологий, таких как геоинформационные системы.

I. Национальная перепись населения и жилого фонда в Польше

1. Проведение периодических переписей населения и жилого фонда очень важно, и переписи следует проводить в разных экономических и социальных условиях. Однако независимо от экономической ситуации в стране и развития общественного сознания, проведение переписи обычно влечет высокие затраты и требует значительной включенности граждан. Поэтому цель состоит в том, чтобы сократить расходы на перепись и снизить нагрузку на респондентов при сохранении высокого качества результатов.
2. Следует отметить, что изменение способов производства статистических данных – не только на национальном, но и на глобальном уровне – отражается на качестве и эффективности проведения переписи. Кроме того, одним из руководящих принципов является минимизация нагрузки на респондентов, что отражается на организации проводимых обследований. На проведение переписи влияют новые информационные и телекоммуникационные технологии, а также постоянно совершенствующиеся альтернативные виды информации, что дает возможность внести изменения в

организацию переписей.

3. Организация и проведение Национальной переписи населения и жилого фонда 2011 (NSP 2011) в Польше очень выиграли от развития информационного общества, что было продемонстрировано в ходе переписи. Последний раунд переписи показал, что технологические инновации в статистике не только возможны, но и ведут к повышению качества результатов, минимизации нагрузки на респондентов и снижению затрат.

4. NSP 2011 была разработана и впервые проведена с применением смешанной модели, то есть на основе административных данных и данных, полученных от респондентов исключительно с использованием электронных вопросников. В результате бумажные переписные листы были упразднены.

5. Использование административных источников, уже существующих в рамках государственных административных структур, было выбрано в качестве отправной точки для проведения NSP 2011. Такое решение было принято для проведения NSP 2011, и это оказалось одним из важных факторов успеха. Использование существующей и значительно усовершенствованной системы регистров, синхронизированной со статистическими системами и поддерживаемой продолжающимися обследованиями, позволило воспользоваться данными, которые были собраны ранее в ходе других обследований, а также для целей учета.

6. Проведение переписи требовало использования интегрированных статистических процессов, эффективного сотрудничества в области методологии, развития ИТ - инфраструктуры, инструментария и метаданных.

7. Наряду с использованием ИТ - систем государственных учреждений, было задействовано три новых канала сбора данных (с общим названием САХI):

(а) Интервью при помощи компьютера через интернет (CAII) – это онлайн вопросник для самостоятельного заполнения, который включает проверку респондентом данных, полученных из административных источников в течение определенного периода времени и, при необходимости, исправление их и внесение недостающей информации;

(б) Телефонное интервью с использованием компьютера (CATI) - телефонный опрос с использованием компьютера, проводимый интервьюером-статистиком;

(с) Личное интервью с использованием компьютера (CAPI) – опрос, проводимый переписчиком с вводом информации при помощи карманного компьютера.

8. Все три канала сбора данных используют адаптируемые электронные вопросники, которые обеспечивают высокое качество данных на этапе сбора. Электронные вопросники заполнялись и корректировались с использованием технологии САХI. При помощи приложения по работе с вопросниками (которое было доступно на мобильном устройстве или через Интернет браузер) проводилась проверка правильности заполнения анкеты; в ряду прочего проводился логический и арифметический контроль.

9. Кроме того, существенным источником первичных данных для переписи NSP 2011 явились данные административной системы.

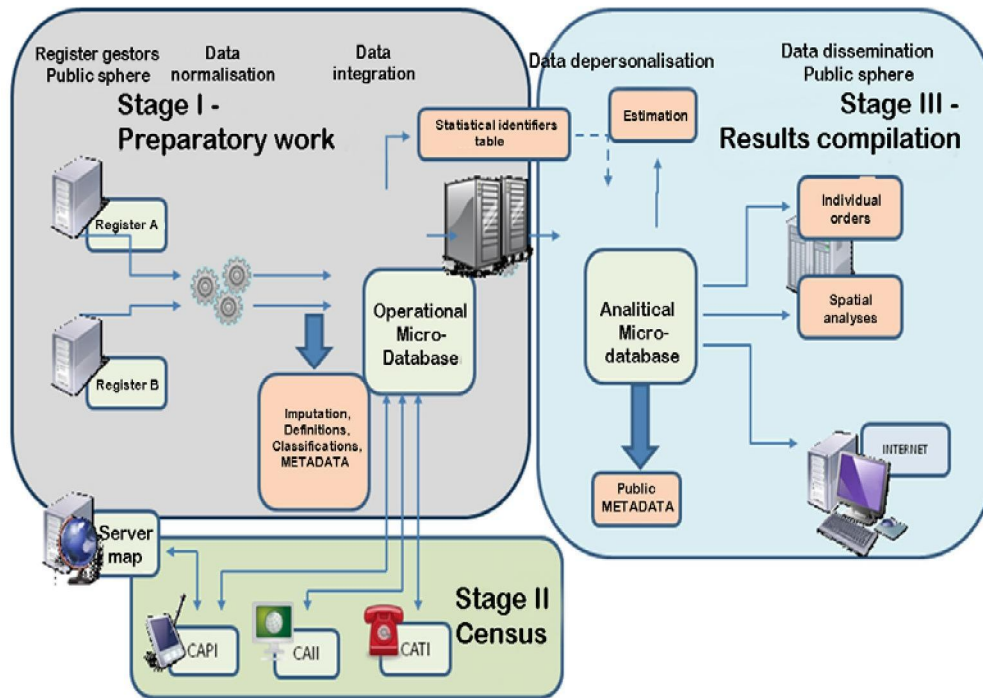
10. Для того чтобы использовать данные из административных и неадминистративных источников для целей переписи NSP 2011, статистики проделали большую работу, чтобы привести эти данные в подходящий вид для формирования статистических данных. Для достижения требуемого качества проводилось редактирование данных, и предпринимались меры по их интеграции. Наибольшие усилия требовались для проведения очистки данных. Процедуры очистки данных использовали алгоритмы, разработанные программистами и администраторами баз данных, и включали проверку достоверности, стандартизацию, устранение дублирования и импутацию данных. Эти методы позволили повысить качество данных Переписи 2011 и сделать их более удобными для пользователей.

11. В соответствии с Законом о переписи все государственные организации, поддерживающие ИТ системы, должны предоставлять данные, которые будут использованы для целей переписи. Закон определяет масштаб и временные рамки для предоставления таких данных.

12. В рамках переписи NSP 2011, было использовано 27 регистров, которые ведут 15 государственных учреждений. Кроме того было использовано три массива данных неадминистративного характера.

13. Причины для использования данных административных систем в Польше были следующими:
- (a) Экономические причины – требование к эффективности: минимизация стоимости статистической продукции, включая снижение административной нагрузки на респондентов;
 - (b) Риск роста неответов в статистических обследованиях, включая переписи;
 - (c) Интенсивное развитие ИТ - систем на основе передовых технологий в государственных учреждениях.
14. Использование административных систем имело многочисленные преимущества, которые включали следующие (но не ограничивались ими):
- (a) Эффективное использование административных систем;
 - (b) Сокращение затрат на проведение переписи;
 - (c) Сокращение нагрузки, связанной с передачей данных;
 - (d) Повышение безопасности данных;
 - (e) Дальнейшая гармонизация обследований;
 - (f) Наличие информации для переписи в годовых регистрах;
 - (g) Наличие данных из административных регистров на всех уровнях территориального деления;
 - (h) Возможность определить ошибки дублирования во вводимых данных;
 - (i) Создание баз микроданных для возможных косвенных оценок и моделирования на уровне единиц наблюдения;
 - (j) Совершенствование методов оценки для малых областей;
 - (k) Улучшение сопоставимости и надежности статистических данных.
15. Данные из административных источников были использованы для целей переписи:
- (a) Данные о зданиях, жилых помещениях и лицах; реестр адресов; основа выборки;
 - (b) В качестве прямого источника данных для переписи (персонализация переписных листов).
16. Для того чтобы администраторы смогли передавать данные из разрозненных систем с использованием средств телекоммуникаций, Центральное статистическое управление Польши создало электронную платформу для сбора и обработки данных, а также сетевое приложение для прямой передачи данных посредством электронных средств через безопасное соединение. Эти решения были также применены при сборе данных о жилом фонде от управляющих недвижимостью.
17. Перепись на основе регистра была проведена с использованием ряда ИТ-систем, содержащих большие объемы данных о единицах, включенных в перепись. Системы, содержащие данные обо всей (или почти всей) совокупности, включенной в перепись, играли главную роль. Результаты переписи были получены на основании многих систем различной значимости с учетом надлежащего качества. Результаты были получены путем передачи значений из административной системы или посредством получения оценок на основе максимально возможной комбинации систем, включая данные, полученные методом опроса с использованием компьютера (САХІ).
18. Для того чтобы обеспечить оптимальное применение передовых информационных и телекоммуникационных технологий, необходимо было создать новую архитектуру переписи. Для целей разработки программы и проведения переписи ЦСУ Польши разработало специальную ИТ - систему (ISS). Система состояла более чем из 10 компонентов, и была реализована разными подрядчиками. Она обеспечила ИТ - инструментарий для всех операций. Система ISS объединила разные технологии (приложения для мобильных терминалов и телефонных опросов, специальные базы данных, хранилища данных, инструментарий для анализа и учета данных).

Архитектура информационной системы переписи



19. В системе ISS были применены различные решения для обеспечения высокого уровня безопасности данных переписи. Были приняты некоторые организационные меры, обязующие каждого, кто принимает участие в работе по переписи, обеспечивать конфиденциальность статистических данных и тем самым гарантировать защиту персональных данных. В части обработки данных переписи, включая данные из административных регистров, в ЦСУ было разработано большое количество правил и теоретических решений, которые затем были внедрены в практику.

20. Мобильное приложение служило в качестве вспомогательного инструмента для переписчиков. Этот инструмент был полезен на всех территориальных уровнях проведения переписи: на центральном уровне (ЦСУ), на уровне *воеводства* (статистические офисы воеводств) и на уровне *гмины* (статистические офисы гмин). Центральный диспетчер (член Центра управления переписью, ответственный за мониторинг и координацию NSP 2011 на центральном уровне) и региональный (воеводств) диспетчер (координатор сбора данных переписи, ответственный за мониторинг и координацию сбора данных переписчиками методом CAPI) получили помощь в виде специального диспетчерского приложения. Подобное приложение использовалось руководителями работ по переписи на уровне гмин (сотрудники местного статистического офиса, которые обеспечивали постоянную поддержку переписчиков, работающих в их районах. Они пользовались мобильными терминалами для облегчения связи с переписчиками и являлись первыми помощниками переписчиков). За бесперебойным функционированием приложений наблюдали администраторы систем переписи, которые работали в каждом статистическом офисе, а также в ЦСУ.

21. Безусловным прорывом явилась замена бумажных переписных листов электронными вопросниками. Это дало возможность повысить эффективность сбора данных. Переписчикам были выданы переносные устройства с загруженными приложениями для заполнения электронных вопросников и цифровыми картами. Использование цифровых карт и технологии GPS изменило планирование и управление операциями, как до начала переписи, так и во время ее проведения.

22. Данные из государственных геодезических, картографических и ортофото (обработанные фотографии аэрокосмической съемки) ресурсов были объединены со статистическими данными для создания основы выборки для переписей. Массивы состоят из статистических адресных точек с пространственной привязкой.

23. Цифровые карты облегчили работу переписчиков при передвижениях по местности, проверке списков и пр. Руководители работ по переписи в местных статистических органах, а также диспетчеры в воеводствах и в центре могут проверять по карте ход проведения переписи и, например, путь и

местоположение переписчика, используя GPS.

24. Включение координат и адресных точек в статистические данные позволило заменить использовавшуюся ранее систему, в которой происходило назначение переписчика на определенный участок (переписной участок), на назначение точек. Это изменение позволило более гибко группировать собранные данные даже для самых малых областей. Это также помогло созданию пространственно-ориентированной базы микроданных и позволило проводить гео-статистический анализ.

25. В соответствии с организационными принципами, до начала переписи были проведены предварительные раунды. В это время переписчики проверяли существование жилых помещений и дополняли реестр отсутствующими в нем адресными точками. Они также изучали район проведения переписи для того, чтобы устранить все возможные неясности, которые возникли при контроле данных на уровне *гмин*.

26. В мобильные терминалы было загружено приложение ГИС, которое позволяло вносить изменения и показывать на карте, среди прочего, местоположение переписчика и назначенные ему адресные точки (GPS). Используя мобильное приложение, переписчик мог изменить местоположение точки, удалить ее или добавить точку, не включенную в реестр. На этой предварительной стадии переписчики должны были также контролировать всю площадь приписанных им счетных участков. Это было особенно важно, когда контроль переписного участка *гмины* производился только на основе регистра, а переписчик был первым и часто единственным человеком, непосредственно участвующим в полевой работе. ГИС - приложение также активно использовалось в ходе переписи – для управления ее ходом. Оно позволило проводить мониторинг и контроль работы переписчиков, а также следить за их передвижениями в регионе (среди прочего, это помогало обеспечить безопасность).

II. Операционная база микроданных 2011

27. В соответствии с Законом о Национальной переписи 2011, в ЦСУ была разработана операционная база микроданных 2011. Система представляла собой инфраструктуру «оборудование – система – инструментарий» (оборудование, системное ПО, вычислительный инструментарий) и приложения (специальные компьютерные программы).

28. База данных включала данные, переданные в электронном формате от организаций, которые должны предоставлять данные в соответствии с Законом, включая те организации, которые поддерживают ИТ - системы. Данные о лицах, собранные через интернет, в ходе телефонного опроса или личных интервью переписчиками, поступали в базу данных. Она также обеспечила платформу для дальнейшей обработки данных.

III. Подсистема метаданных переписи

29. В подсистеме метаданных переписи хранятся необходимые метаданные, описывающие данные переписи и операции по проведению переписи, включая операции, необходимые для подготовки отчетов о качестве.

IV. Аналитическая база микроданных

30. Аналитическая база микроданных (АМБ) была сформирована для хранения деперсонифицированных данных переписи в их окончательном виде. АМБ используется в настоящее время, среди прочего, для проведения статистического анализа, результаты которого распространяются для широкой публики. АМБ дает возможность быстрого формирования агрегированных данных. Она содержит данные о миграции населения, рынке труда, поездках на работу, образовании, инвалидности, национальности и религии, а также о жилом фонде.

30. Следует отметить, что эффективному проведению переписи способствовала хорошая подготовка в методологической, организационной и логистической областях.

31. Сроки были определены правильно, поскольку этап сбора данных (определенный в Законе о переписи) не нужно было продлевать, и не было необходимости увеличивать бюджет. Подробный график проведения переписи NSP 2011 регулярно корректировался. Основной график включал около 250 пунктов. Детальные графики по задачам, например, проведение контрольной переписи и поддержка систем переписи, хранились в отдельных файлах. Эти графики включали несколько тысяч пунктов.

32. Основа выборки представляла собой иерархическую структуру в соответствии с территориальным делением страны и системой сбора данных переписи NSP 2011. Она включала центральный уровень (ЦСУ), уровень воеводств (региональные статистические офисы) и гмин (местные статистические офисы). Некоторые части структуры были постоянными, другие (на уровне воеводств и гмин) были образованы только для целей подготовки и проведения NSP 2011 в поле.
33. Тщательный анализ проведения переписи с учетом всех действующих лиц, включая полевой персонал, позволяет сделать выводы о дальнейшем использовании новой технологии. Эта технология оказалась хорошо применимой для обследований с использованием вопросников. Это позволяет снизить стоимость процесса, применить современные методы контроля, повысить качество собираемых данных, и, следовательно, снизить нагрузку на респондентов.
35. Замена бумажных переписных листов на электронные вопросники и введение новых методов сбора данных снизило стоимость проведения переписи в Польше. При традиционном проведении переписи затраты были бы выше.
36. Успешное применение современных методов сбора данных привело к тому, что ЦСУ Польши приняло решение о систематическом использовании данной архитектуры для других статистических обследований, использующих вопросники, в Польше.

V. Уроки и предложения

37. Впереди следующий раунд переписи. На основе приобретенного опыта, Польша будет использовать новые ИТ- технологии, что сделает его проведение еще более эффективными и инновационным.
38. До того как появятся новые технологии, отправной точкой будут служить методы, использованные в 2011 году. Следует приложить большие усилия для развития новой стратегии переписи, так чтобы гарантировать дальнейшие улучшения. Следует попытаться:
- (a) Сократить затраты на перепись;
 - (b) Эффективно использовать административные источники;
 - (c) Снизить социальную нагрузку, связанную с передачей данных;
 - (d) Усилить безопасность передачи данных;
 - (e) Улучшить согласованность и надежность статистических данных.
39. Польша и дальше будет развивать методы, использованные при проведении NSP 2011. ЦСУ будет уделять внимание развитию технологии, программного обеспечения и телекоммуникаций, но прежде всего, проведет оценку юридических аспектов сбора данных переписи и защиты конфиденциальности (особенно в том, что касается предоставления доступа к индивидуальным данным). Также важно развивать методологию для увязки данных, приходящих из разных источников и разных обследований (сплошные и выборочные обследования).
40. Более подробную информацию можно найти на вебсайте: www.e-census.stat.gov.pl