



ЭКОНОМИЧЕСКИЙ И СОЦИАЛЬНЫЙ СОВЕТ

Распространение: Общее
27 марта 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

Проблемы перехода от традиционной переписи к переписи на основе регистров в Австрии

Записка Статистического управления Австрии

Резюме

В этом документе представлены основные этапы перехода от традиционной переписи к переписи на основе регистров в Австрии. Кроме того, здесь описан структурный подход к оценке качества административных данных. В ходе трехэтапного процесса (первичные данные, объединенные данные, импутированные данные) получают показатели качества, которые должны относиться ко всей имеющейся информации о качестве для каждого признака. Для того чтобы гарантировать применимость такого подхода для оценки качества переписи на основе регистров 2011 г., процедура была протестирована с использованием данных статистики рынка труда 2009, полученных из регистров.

Важным преимуществом этого подхода является то, что оценка качества и обработка данных независимы. Это дает возможность применить методы для массивов данных из регистров. Кроме того, сотрудничество между Статистическим управлением и органами, ответственными за регистры, усиливается, так как держатель данных участвует в процессе оценки качества.

I. Эволюция: от традиционной переписи к переписи на основе регистров в Австрии

1. В мае 2011 года в Австрии была проведена последняя традиционная перепись населения наряду с переписью зданий и жилого фонда, а также перепись местных производственных единиц. Объединенная перепись охватывала приблизительно 2 млн. зданий, 3,8 млн. жилых помещений и 3,3 млн. домашних хозяйств, состоящих из 8,1 млн. человек; перепись проводилась с использованием бумажных переписных листов. Объединенная перепись явилась очень сложным и дорогим мероприятием.

2. Принимая во внимание сложные требования к традиционной переписи, в последнее время рассматривается вопрос о все более широком использовании административных данных. Обработка данных, которые уже были зарегистрированы административными органами, имеет массу преимуществ по сравнению со сбором данных обследований; эти преимущества включают сокращение затрат, снижение нагрузки на респондентов и быстрый доступ к данным. Использование данных из существующих регистров

должно обеспечить оптимальное отражение реальности при разумных затратах путем объединения регистров с использованием уникальных переменных для увязки, повышения качества данных и гармонизации определений. Очевидным преимуществом является регулярная актуализация регистров для отслеживания любых изменений в данных, описывающих единицы и их характеристики (Statistics Finland, 2004, стр. 10).

3. Все большее число Национальных статистических служб поддерживают переписи на основе регистров как замену традиционной переписи. В 2000 г. Совет министров Австрии принял решение ввести новый метод проведения переписи в 2011 г., и в 2006 г. законодательство вступило в силу. Австрия находится в числе шести европейских стран (вместе со Швецией, Финляндией, Норвегией, Данией и Словенией), которые провели в 2011 г. перепись на основе регистров.

4. На основании оценки численности населения, полученной в ходе переписи, определяются денежные суммы бюджетного выравнивания между федеральными и местными финансовыми органам, а также число избирателей в будущих выборах. Кроме того, данные о поездках к месту работы, образовании и занятости предоставляют важную информацию для выработки экономической и социальной политики. Однако в административных источниках данных могут применяться определения, которые не соответствуют потребностям национальной статистической службы, даже если качество данных не вызывает сомнений (см. United Nations, 2007, р. 3). Таким образом, статистические службы должны решать, соответствуют ли данные изучаемым вопросам.

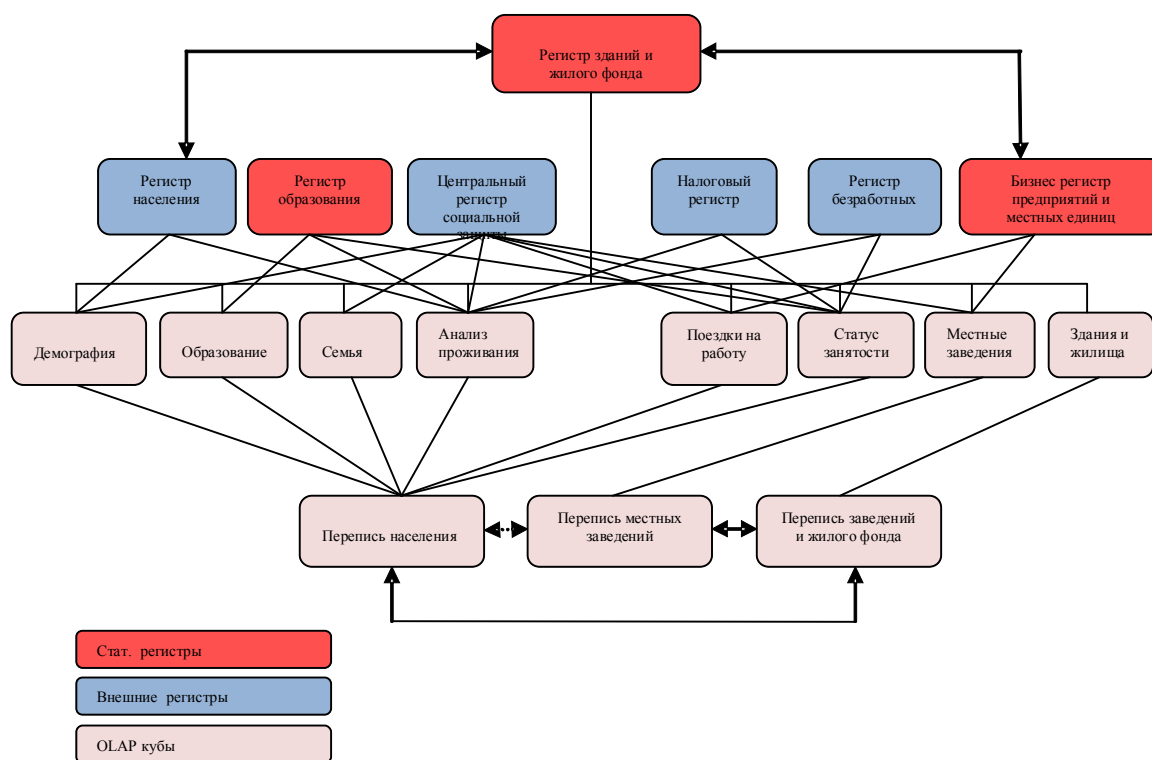
А. Принцип избыточности

5. Одна из основных проблем – выбор подходящего источника данных для получения нужной информации. Цель переписи на основе регистров состоит в охвате всех переменных, которые ранее предоставлялись в ходе традиционной переписи. В этом смысле Перепись 2001 явилась толчком к созданию некоторых источников данных, например Центрального регистра населения, Регистра зданий и жилого фонда или Регистра уровня образования. Каждый муниципалитет вел свои записи и обычно данные даже не были введены в электронные системы. С введением в 2001 г. регистра жилого фонда возник централизованный регистр населения, и муниципалитеты стали обязаны предоставлять свои записи. Последняя традиционная перепись в Австрии была основана на информации из вновь созданного регистра населения. В 2004 г. Регистр зданий и жилого фонда был в первый раз синхронизирован с регистром населения. Кроме того, в ходе проведения переписи 2001 был создан регистр уровня образования.

6. На рис. 1 показаны все основные регистры и их связь с соответствующими темами. Источники данных, закрашенные красным цветом, поддерживаются Статистическим управлением Австрии, остальная информация предоставляется из внешних источников, как, например, Службой занятости Австрии (AMS) или Службой социального обеспечения (SVA). Центральный регистр населения является основой переписи, поскольку единицами анализа являются отдельные лица с основным местом жительства в Австрии. Для обеспечения качества результатов переписи основной регистр поддерживают семь регистров для сравнения. Информация об этих семи видах административных единиц предоставляется 35 держателями данных и, в основном, используется для перекрестных проверок, а также для предоставления информации, которая отсутствует в основном регистре. (Berka et al., 2010, стр. 300).

7. Принимая во внимание независимость различных регистров, а также автономные процессы сбора данных, источники часто содержат разные значения для одних и тех же переменных. Поэтому принцип избыточности используется для обеспечения качества путем получения информации о поле, национальности и возрасте из как можно большего количества источников. Метод, разработанный в Статистическом управлении Австрии, нацелен на определение одного конкретного регистра в качестве основного для получения информации об определенной переменной, тогда как регистры для сравнения используются для подтверждения значений из основного регистра (Lenk, 2008, стр. 3). Хотя регистры предоставляют достаточную информацию о большинстве характеристик из традиционной переписи, некоторые переменные не могут быть включены в перепись на основе регистров. Например, продолжительность ежедневных поездок на работу, разговорный язык или религиозная принадлежность не отражаются ни в одном из имеющихся регистров.

Рисунок 1:
Регистры и темы переписи на основе регистров в Австрии



В. Защита конфиденциальности данных

8. Поскольку имена и номер социального страхования не включены в доставляемые данные, уникальный идентификатор необходимо использовать для объединения данных из разных источников. Для обеспечения конфиденциальности данных необходимо ввести для каждого сектора персональный идентификационный номер для официальной статистики (bPIN OS) перед тем, как отправлять данные в Статистическое управление Австрии. Каждый административный сектор в Австрии, например «социальное обеспечение», «налоги», или «социальное обеспечение» имеет свой код bPIN. Эти 172-значные коды, которые должны служить для защищенной передачи данных между государственными органами в рамках электронного правительства, получает Австрийская комиссия по защите данных из регистра PIN кодов (Stammzahlregister).

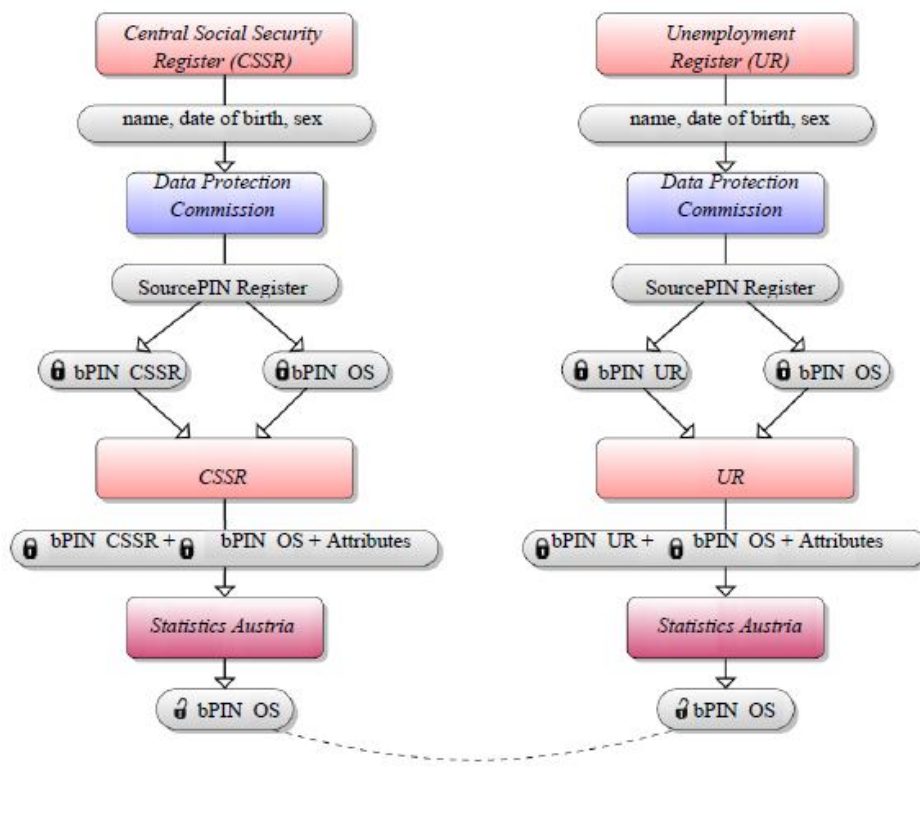
9. На рис. 2 показано обеспечение защиты конфиденциальности данных путем получения кодов для секторов и PIN кодов для официальной статистики. Держатели данных запрашивают эти PIN коды у Комиссии по защите данных для каждого лица, сообщая имя, пол, дату рождения, место рождения и адрес. PIN-коды для секторов (bPIN), а также PIN-коды для официальной статистики (bPIN OS) получают из Регистра PIN-кодов с использованием специального очень сложного алгоритма, разработанного Комиссией по защите данных. Для держателей регистров, кроме Статистического управления Австрии, код bPIN OS предоставляется только в зашифрованной форме. Держатель регистра должен отправлять данные вместе с зашифрованным кодом bPIN OS в Статистическое управление Австрии. Кроме того, держатель регистра сообщает свой собственный секторальный зашифрованный PIN-код для каждого лица для того, чтобы

идентифицировать соответствующую запись на случай дальнейших запросов от Статистического управления Австрии (Lenk, 2008, p. 5).

10. Каждый держатель данных может декодировать свой bPIN в 28-значный номер, следовательно, только национальная статистическая служба может расшифровать bPIN OS и использовать его в качестве общей связующей переменной для входящих данных из регистров. Описанная процедура зашифровки персональных идентификаторов была впервые применена в пилотной переписи 2006 года.

Рисунок 2

Защита конфиденциальности данных в переписи на основе регистров



II. Проведение предварительной переписи

A. Пилотная перепись 2006

11. В 2006 г. Статистическое управление Австрии провело пилотную перепись в масштабе полной переписи населения, домов и жилых помещений и мест работы для оценки перехода от традиционной переписи к переписи на основе регистров. Впервые была запрошена информация от различных держателей данных для объединенной административной статистики. Не считая небольших проблем с источниками данных, пилотная перепись прошла успешно. Демографические данные, данные об образовании и занятости оказались очень высокого качества и достоверности, а проблемы возникали с идентификацией мест работы для тех, кто регулярно совершает поездки к месту работы, и с распределением лиц по жилым помещениям (Statistics Austria, 2009, стр. 7). Статистическое управление Австрии провело проверку качества результатов путем сравнения их с результатами выборочного обследования на ту же дату (Lenk, 2008, p. 14).

12. В 2008 г. введение закона о бюджетном выравнивании стало поворотным пунктом в истории переписей в Австрии. Закон устанавливал распределение федеральных фондов по муниципалитетам в

зависимости от оценки численности населения, полученной в ходе пилотной переписи, что превратило ее из необязательного мероприятия в политически значимый инструмент.

В. Анализ проживания

13. Для обеспечения качества расширенной пилотной переписи было важно выявить пропущенные значения, многократные записи или ошибки измерения и устранить их для целей расчета. Эта процедура, известная как анализ проживания, обозначает пять причин для исключения лиц из переписи:

- (a) Человек умер до даты переписи, но все еще присутствует в Центральном регистре населения (ЦРН);
- (b) Имеется несколько записей по одному лицу на дату переписи, например два или более мест основного проживания в ЦРН;
- (c) Человек был зарегистрирован с основным местом проживания в Австрии менее 90 дней, что связывает численность населения с минимальным пребыванием;
- (d) Выявление так называемого «туризма при переписи», то есть лица, которые зарегистрированы в муниципалитете не дольше 180 дней, но зарегистрированы в других муниципалитетах до и после отчетного периода;
- (e) Случаи возможных расхождений без основного места жительства, то есть лица, зарегистрированные только в ЦРН, но больше ни в одном административном источнике.

14. Последний случай требует уточнения в национальной статистической службе, поскольку признаков проживания недостаточно (см. Табл. 1). Данные по лицам, которые охвачены только ЦРН, должны быть уточнены, а лица опрошены в письменной форме¹. В нашем примере это лицо с идентификационным номером ID3459. В рамках пилотной переписи письма для уточнения были посланы около 45 000 человек, из которых 9 000 подтвердили свое постоянное место жительства в Австрии. Наконец, около 0,5 % первоначального населения не было переписано, включая примерно 10 000 технических удалений (см. пп. 1-3). Муниципалитеты в Австрии должны быть проинформированы о непереписанных лицах, а регистрационные органы могут доказать проживание этих лиц или, возможно, удалить их из регистра жителей. В соответствии с результатами пилотной переписи в 2006 г., около 80% непереписанных лиц были удалены из регистров местными муниципалитетами.

15. Для переписи на основе регистров в 2011г. Статистическое управление Австрии применяет такую же процедуру как в пилотной переписи 2006 г. Анализ проживания был начат в январе 2012 г. В первом раунде было разослано около 54 000 писем для уточнения основного места проживания. Второй раунд намечен на сентябрь 2012 г., так как будет завершена передача информации от внешних держателей данных.

Таблица 1 . Признаки проживания в анализе проживания

бPIN OS	CPR	CSSR	TR	UR	SWR	CAR
.
.
ID 3457	+	+	+	-	-	-
ID 3458	+	+	-	+	-	-
ID 3459	+	-	-	-	-	-
ID 3460	+	-	-	-	+	-
.
.

CPR= Центральный регистр населения, CSSR = Центральный регистр социальной защиты,
TR = Налоговый регистр, UR = регистр безработных, SWR = регистр получателей социальной

¹ Письмо во всех случаях содержит только один вопрос: «Ваше основное место жительства на дату переписи находилось в Австрии? «Да» или «Нет»

помощи, CAR = Регистр получателей пособий на детей

С. Статистика рынка труда на основе регистров

16. Особенно широкие возможности для формирования статистики в реальном времени дает применение на ежегодной основе статистики рынка труда, основанной на регистрах. Начиная с 2009 г., данные, касающиеся демографии, образования, поездок на работу и рынка труда, предоставляются ежегодно и позволяют получить подробную информацию о социально-экономической ситуации в Австрии даже на муниципальном уровне. Например, частые запросы административных данных позволяют проводить глубокий анализ изменения ситуации с безработицей молодежи от года к году или динамики различий в уровне образования в разбивке по полу. Используемые источники данных и процессы для формирования этих показателей аналогичны методам, применяемым в переписи на основе регистров. Около 35 различных административных и статистических источников увязаны для получения интересующих переменных. При реализации статистики рынка труда на основе регистров национальная статистическая служба соответствует стандартам современной статистики и анализа данных. Данные даже имеются в открытом доступе в статистической базе данных на вебсайте Статистического управления Австрии (www.statistik.at).

III. Соображения по поводу подхода к оценке качества

17. Короткий переходный период от традиционной переписи к переписи на основе регистров был трудной задачей для Австрии. В некоторых странах период постепенной замены данных обследования административными данными длился около 20 лет (например, в Финляндии, см. Ruotsalainen (2008)). Это позволило некоторым странам провести широкое обсуждение возможностей для оценки качества в рамках национальной статистической службы, а также среди держателей данных, в то время как в Австрии график перехода был довольно жестким. Поскольку национальная статистическая служба не отвечает за ведение регистров внешних данных, необходимо было проводить оценку качества в процессе проведения переписи на основе регистров. Анализ качества данных регистра должен удовлетворять нескольким требованиям: прозрачности, точности и выполнимости. Наш подход предоставляет рамки для анализа качества административных данных с использованием трех разных измерений гиперкуба для получения индексов качества (см. Verka et al., 2012). Этот подход тесно связан с потоком данных, но независим от обработки данных, что обеспечивает тот факт, что обработка не подвергается влиянию, а лишь оценивается.

18. Поток данных переписи на основе регистров в Австрии имеет три уровня: первичные данные (то есть регистры), объединенный массив данных (База данных переписи, CDB) и импутированный массив данных (окончательный массив данных, FDP). На рис. 3 показан процесс обработки данных, начиная с получения первичных данных от разных держателей административных данных. Информация увязана с использованием уникального ключа (bPIN OS) и объединена в кубы данных в CBD. Далее данные CBD пополняются импутированными значениями по результату анализа ответов на отдельные вопросы, что завершает построение FDP, который содержит фактические данные и оценки.

A. Оценка качества на уровне регистра

19. Информацию о качестве на уровне первичных данных получают, используя три измерения гиперкуба: Документацию (HDD), Предварительную обработку (HDP) и Внешний источник (HDE). До того как увидеть данные, HDD описывает процессы, относящиеся к качеству, в органе, ведущем регистр, а также документацию о данных (то есть метаданные). Для контроля степени уверенности и надежности держателей данных используют вопросник, содержащий 16 открытых вопросов и 9 оцениваемых вопросов. Национальная статистическая служба может, таким образом, проводить проверку методов сбора данных или наличия законного принуждения к регистрации данных, что может оказывать значительное влияние на качество данных. На эти вопросники отвечают эксперты из соответствующих органов, ведущих регистры, так что результаты должны быть убедительными.

20. Второе измерение гиперкуба HDP касается формальных ошибок в первичных данных. Ошибки диапазона в данных, неответы на отдельные вопросы и отсутствующие первичные ключи обнаруживаются на этом этапе оценки качества. Окончательный результат этого измерения гиперкуба представлен отношением годных к употреблению записей к общему количеству записей. Эта процедура также проводится для каждого признака в регистре. Если процент годных к употреблению записей для признака в

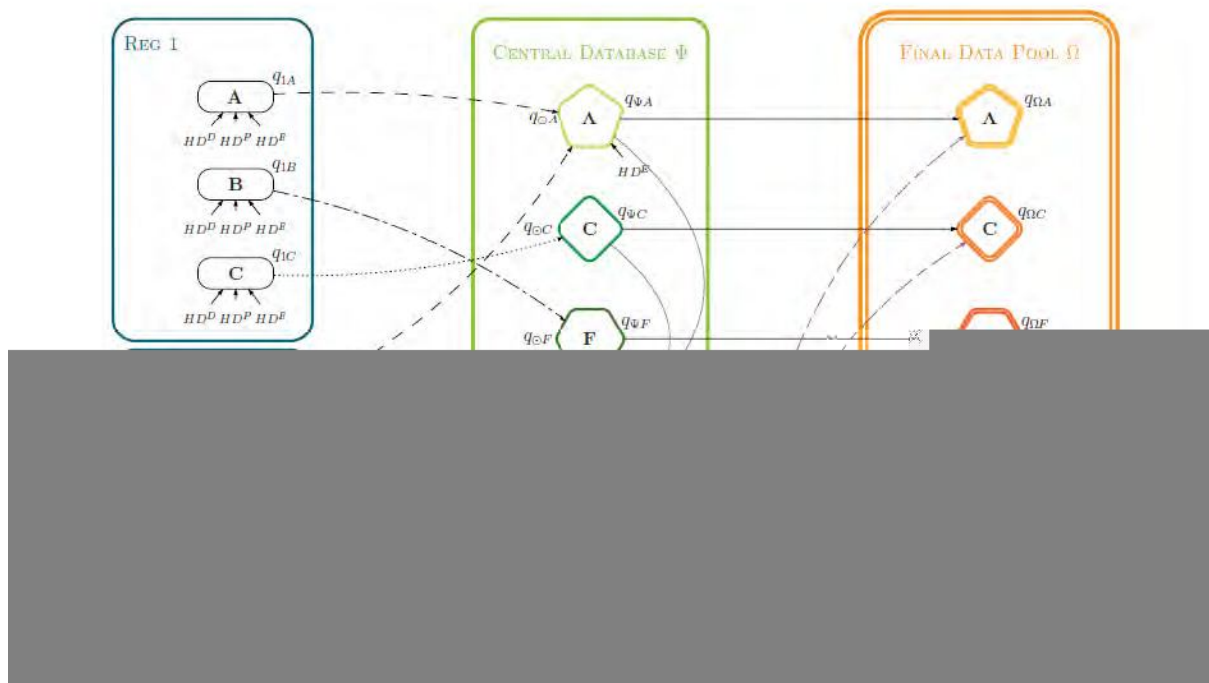
каком-то регистре меньше, чем этот процент для того же признака в другом регистре, то оценка качества будет ниже

21. Наконец, третье измерение гиперкуба, HDE обеспечивает сравнение между данными на основе регистров и внешним источником. В Австрии, микроперепись является общим ориентиром для репрезентативных обследований и считается лучшим имеющимся набором данных для сравнений. Проверка на согласованность с внешним источником предлагает третью оценку качества, которой является отношение между количеством совпадающих значений и общим количеством увязанных записей. Если рассматриваемый признак не представлен в микропереписи, мы полагаемся на мнения экспертов.

22. При наличии этих трех измерений качества, общий показатель качества для каждого признака и регистра может быть получен как средневзвешенное значение. Следовательно, этот показатель может отразить эффекты, связанные с качеством, начиная от формирования данных до поступления окончательных первичных данных в регистр.

Рис 3

Оценка качества для переписи на основе регистров



q_n	Показатель качества
Λ	Признак
\diamond	Признак из нескольких источников
\diamond	Уникальный признак
\diamond	Производный признак
HD^D	Документация
HD^P	Предварительная обработка
HD^E	Внешние источники
HD^I	Импутация

В. Оценка качества в базе данных переписи

23. Вся информация из регистров объединяется в Базе данных переписи, которая охватывает все

признаки, представляющие интерес для переписи. Поскольку информация о признаке может поступать более чем из одного источника, существуют правила, разработанные Статистическим управлением Австрии, по которым выбирается наиболее подходящая информация из регистров. Относительно оценки качества для Базы данных мы различаем три типа признаков по их происхождению:

- (a) Уникальные признаки, присутствующие только в одном регистре, например уровень образования (признак С на рис. 3);
- (b) Признаки, присутствующие во многих регистрах, например, пол (см. Признак А на рис. 3). Информация из нескольких источников объединяется в соответствии с набором правил для получения наиболее подходящего значения для признака в Базе данных переписи;
- (c) Производные признаки создаются на основе разных признаков, например, статус занятости (см. признаки F и G на рис. 3). Регистры не содержат информации по этим признакам в требуемом виде.

24. Подробное описание оценки качества для трех типов признаков в Базе данных переписи можно найти в работах Berka et al. (2010) и Berka et al. (2012).

25. В настоящее время основное внимание в исследованиях уделяется Окончательному массиву данных (FDP), который соответствует Базе данных переписи после применения импутации. Количество неотчетов на отдельные вопросы эффективно снижается посредством импутации, однако сам процесс импутации необходимо контролировать. Для этого используется информация из измерения гиперкуба «Импутация» (HD¹), которая является постоянной задачей.

С. Выводы и перспективы

26. В этом документе представлены основные этапы перехода от традиционной переписи к переписи на основе регистров в Австрии. Кроме того, здесь описан структурный подход к оценке качества административных данных. В ходе трехэтапного процесса (первичные данные, объединенные данные, импутированные данные) получают показатели качества для трех измерений гиперкуба. Эти оценки должны относиться ко всей имеющейся информации о качестве для каждого признака. Для того чтобы гарантировать применимость такого подхода для оценки качества переписи на основе регистров 2011 г., процедура была протестирована с использованием данных статистики рынка труда 2009, полученных из регистров.

27. Важным преимуществом этого подхода к оценке качества является независимость оценки качества и обработки данных. Отделение от процедуры обработки нужно для того, чтобы оценить процесс, не оказывая влияния на него. Это дает возможность применить методы для массивов данных из регистров. Кроме того, сотрудничество между Статистическим управлением и органами, ответственными за регистры, усиливается, так как держатель данных интегрирован в процесс оценки качества.

IV. Ссылки

Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H., & Schwerer, E. (2010). A quality framework for statistics based on administrative data sources using the example of the Austrian census 2011. *Austrian Journal of Statistics*, Volume 39, Number 4, 299-308.

Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H., & Schwerer, E. (2012). Combination of evidence from multiple administrative data sources: quality assessment of the Austrian register-based census 2011. *Statistica Neerlandica*, Volume 66, Issue 1, 18-33.

Lenk, M. (2008). *Methods of Register-based Census in Austria* (Tech. Rep.). Statistik Austria, Wien.

Ruotsalainen, K. (2008). *Finnish Register-based Census System* (Tech. Rep.). Statistics Finland.

Statistics Austria. (2009). *Bericht Über die Probezählung 2006 - Ergebnisse und Evaluierung* (Tech. Rep.). Statistics Austria.

Statistics Finland. (2004). *Use of registers and administrative data sources for statistical purposes* (Tech. Rep.). Statistics Finland.

United Nations. (2007). *Register-based statistics in the Nordic countries*. United Nations Economic Commission for Europe, New York and Geneva.