



**Экономический
и Социальный Совет**

Distr.
GENERAL

ECE/CES/GE.41/2007/6
22 March 2007

RUSSIAN
Original: ENGLISH

ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

Группа экспертов по переписям населения и жилищного фонда

Десятая сессия

Астана, 4-6 июня 2007

Пункт 3 (b) предварительной повестки дня

**ТЕХНОЛОГИЯ ПЕРЕПИСИ: НОВЫЕ РАЗРАБОТКИ И ИХ ВЛИЯНИЯ
НА МЕТОДОЛОГИЮ ПЕРЕПИСИ**

Увязка данных в целях переписи в Нидерландах

Документ представлен Нидерландами*

Настоящее Совещание проводится совместно с Евростатом

Резюме

Бюро Конференции европейских статистиков (КЕС) на своем совещании, состоявшемся в Вашингтоне, округ Колумбия, 19-20 октября 2006 года, одобрило обновленный круг ведения Руководящей группы по переписям населения и жилищного фонда и план деятельности КЕС в области переписей населения и жилищного фонда. Бюро КЕС также приняло решение о том, что Руководящая группа будет координировать работу по различным типам совещаний. Настоящий документ подготовлен по просьбе Руководящей группы по переписям населения и жилищного фонда для представления и обсуждения на Совместном совещании ЕЭК/Евростата по переписям населения и жилищного фонда, которое состоится в Астане (Казахстан) 4-6 июня 2007 года. Данный документ послужит основой для обсуждения на заседании, посвященном теме "Технологии проведения переписи: последние изменения в проведении переписи и их последствия для методологии".

* Настоящий документ был подготовлен Статистическим управлением Нидерландов.

ВЫДЕРЖКА

1. Для подготовки таблиц переписи 2001 года в Нидерландах были сведены воедино данные из многих различных источников. С момента проведения последней переписи, основанной на сплошном учете, которая проводилась в 1971 году, готовность населения участвовать в переписях резко снизилась. Статистическое управление Нидерландов нашло альтернативу в форме виртуальной переписи, в которой используются имеющиеся регистры и обследования. Ее табличные результаты сопоставимы не только с итогами предыдущих переписей Нидерландов, но и переписей, проводившихся другими странами в рамках цикла переписей 2000 года.
2. Для переписи 2001 года потребовалась более детальная информация, чем в случае предыдущих циклов переписей. Накопленный опыт работы с данными из различных административных регистров для использования в статистических целях позволил Статистическому управлению Нидерландов разработать базу данных социальной статистики (БДСС), в которой содержится согласованная и подробная демографическая и социально-экономическая статистическая информация об отдельных лицах и домохозяйствах. Основным элементом БДСС является регистр населения. В основе БДСС лежит процесс увязки на микроуровне ряда административных регистров и результатов выборочных обследований. Процесс микроинтеграции обеспечивает непротиворечивость, согласованность и полноту данных БДСС. Для получения информации, отсутствующей в регистрах, по-прежнему необходимы выборочные обследования.

Ключевые слова: перепись; микроинтеграция; увязка на микроуровне

I. ВВЕДЕНИЕ

3. Первая перепись в Нидерландах была проведена в 1795 году для целей создания избирательных участков. В то время объединенные провинции Нидерландов еще являлись республикой, а границы страны отличались от нынешних. После падения Наполеона Нидерланды стали королевством, и переписи начали организовываться каждые десять лет. Первая перепись в Королевстве Нидерландов была проведена в 1829 году. До создания Статистического управления Нидерландов состоялось шесть переписей в 1839, 1849, 1859, 1869, 1879 и 1889 годах, причем отвечало за их проведение министерство внутренних дел. Статистическое управление Нидерландов было создано в 1899 году, и ему сразу же было поручено проведение восьмой переписи. В XX веке было проведено еще шесть традиционных переписей – в 1909, 1920, 1930, 1947, 1960 и 1971 годах. Три

самые последние переписи (1981, 1991 и 2001 годов) основывались не на сплошном учете, а на использовании регистров и результатов обследований, имевшихся в распоряжении Статистического управления Нидерландов.

4. Первоначально переписи преследовали две цели. Во-первых, они проводились для исправления ошибок в муниципальных регистрах учета населения. Во-вторых, они использовались для получения дополнительной информации о социально-экономических явлениях в стране. Поскольку Нидерланды проводят перепись на основе регистров, первая цель уже не ставится. Кроме того, качество центрального регистра населения (РН), объединяющего в себе регистры населения всех муниципалитетов, с течением времени значительно повысилось. Это объясняется тем, что ведение муниципалитетами своих регистров населения на постоянно обновляемой основе стимулируется распределением среди муниципалитетов средств центрального правительства, которое, как правило, производится на основе численности населения по данным из местных регистров. Еще одна причина заключается в том, что без регистрации в РН жизнь в голландском обществе сопряжена с большими практическими трудностями. Поэтому как муниципалитеты, так и рядовые граждане в достаточной мере заинтересованы в поддержании высокого качества РН. Подтверждением этих слов являются недавно принятые в Роттердаме меры по повышению качества муниципального регистра населения в отношении отдельных старых кварталов города. Вторая цель не теряет своей актуальности, и многие результаты переписей публикуются в историческом или международном контексте. В настоящее время данные переписей пользуются популярностью для проведения сопоставлений между странами.

5. Первая традиционная перепись населения в Королевстве Нидерландов, объявленная королевским декретом, состоялась в 1829 году. Она положила начало 140-летнему периоду в истории Нидерландов, в ходе которого было проведено 14 традиционных переписей. Последняя перепись этого цикла была проведена в 1971 году. С тех пор наблюдается растущее недоверие к целям сбора правительством всякого рода информации о своих гражданах. Это ознаменовало собой период, когда общество стало менее склонным к сотрудничеству, что заставило Статистическое управление Нидерландов искать альтернативы традиционной переписи населения. Вместо регистрации Статистическое управление Нидерландов для получения необходимой переписной информации стало изучать возможности использования административных регистров и выборочных обследований в качестве новых источников данных.

6. В ходе переписей 1981 и 1991 годов данные брались из Регистра населения. Данные о социально-экономических характеристиках, такие, как трудовая активность и образовательный уровень, были получены в рамках обследования рабочей силы. Однако,

поскольку эти источники использовались обособленно, вопросу непротиворечивости информации на микроуровне не уделялось достаточно внимания. Кроме того, табличные итоги одного источника могли отличаться от соответствующих итогов другого. Для преодоления проблемы непротиворечивости табличные результаты подвергались повторному взвешиванию до уровня итоговых цифр регистра населения (описание опыта переписи 1981 года см. в работе Vliegen and Van de Stadt, 1988; анализ обработки данных переписи 1991 года см. в работе Corbey, 1994). Данные, полученные в 1981 и 1991 годах, были намного менее подробными, чем набор таблиц переписи 2001 года. В отличие от 1981 и 1991 годов Статистическое управление Нидерландов опубликовало переписные данные за 2001 год на муниципальном уровне.

7. Для разработки программы переписи 2001 года Статистическое управление Нидерландов использовало новый подход, ранее никем не использовавшийся в Европе. Одним из наиболее важных достижений 1990-х годов в области официальной статистики стало создание возможности разработки все большего объема социально-экономической статистики на основе административных регистров. В настоящее время в наличии имеются всеобъемлющие подробные данные о занятости и социальном страховании. Благодаря увязке на микроуровне и микроинтеграции демографических и социально-экономических данных и широкого спектра административных регистров и выборочных обследований Статистическое управление Нидерландов создало Базу данных социальной статистики (БДСС). Данная БДСС содержит непротиворечивую подробную информацию о лицах, домохозяйствах, местах работы и (социальных) пособиях. Таким образом, она являлась надлежащим источником данных для проведения переписи населения 2001 года. Непротиворечивость (под)итогов выборочных обследований и регистров достигается с помощью метода повторного взвешивания, представляющего собой новый передовой метод взвешивания.

8. Настоящий документ в значительной степени опирается на главы 1 и 13 работы Schulte Nordholt et al. (2004). В разделе II более подробно обсуждается перепись 2001 года. В разделе III описывается метод, использовавшийся для расчета итогов переписи 2001 года. В разделе IV обсуждаются вопросы комбинирования источников данных, в частности процесса увязки на микроуровне и процесса микроинтеграции. В разделе V рассказывается о создании БДСС. В разделе VI изложены некоторые заключительные замечания.

II. ПЕРЕПИСЬ 2001 ГОДА

9. В 2003 году для подготовки таблиц нидерландской переписи 2001 года применялось комбинирование данных. В Нидерландах это было сделано путем использования данных,

которыми уже располагало Статистическое управление Нидерландов, а не путем опроса жителей в рамках сплошного учета. Благодаря этому затраты на проведение переписи оказались для голландских налогоплательщиков значительно меньше. Для проведения традиционной переписи потребовалось бы ориентировочно 300 млн. евро, а при использовании этого метода расходы составляют "всего лишь" примерно 3 миллиона. Эта смета включает в себя расходы на все подготовительные мероприятия, такие, как разработка новой методологии и сопутствующего программного обеспечения. В отличие от расходов на проведение анализа полученных результатов расходы на регистры в этой смете не предусмотрены, поскольку регистры уже имеются в наличии. Регистры обновляются не для целей переписи, а для решения других задач. Экономия средств на расходах, связанных с переписями, возможна лишь в странах, обладающих достаточной регистровой информацией.

10. По итогам переписи 2001 года подготовлено 40 обширных таблиц. 28 из них содержат данные по Нидерландам в целом, а 9 - данные на уровне КОРОП (КТЕС-3) и 3 - данные на муниципальном уровне (КТЕС-5). Все 40 таблиц разбиты на ряд категорий. Восемь таблиц посвящены жилищному фонду, две таблицы - маятниковой миграции, а остальные 30 таблиц содержат демографические данные, касающиеся рода занятий, уровня образования и экономической активности. Кроме того, данные, касающиеся демографических характеристик, жилищного фонда и рабочей силы, скомпилированы на субгородском - районном - уровне по 10 крупным городам, участвующим в Аудите городов II (Statistics Netherlands, 2003).

11. Виртуальная перепись в Нидерландах началась позднее, чем в других странах, в которых проводилась традиционная перепись. Приступить к реализации проекта переписи 2001 года, не имея в наличии всех источников, было нецелесообразно; некоторые регистры были готовы относительно поздно. Тем не менее Нидерланды быстрее большинства других стран, участвовавших в цикле переписей 2000 года, завершили составление 40 таблиц переписи. Более того, Нидерланды одними из первых направили полный набор из 40 таблиц в Евростат, координирующий представление материалов всеми государствами - членами Европейского союза (ЕС), присоединяющимися странами и государствами - членами Европейской ассоциации свободной торговли (ЕАСТ). Нидерланды имели то преимущество, что входящие переписные листы не нуждались в проверке и корректировке. Однако необходимо отметить, что по некоторым переменным имеется только выборочная информация, чем и была обусловлена невозможность достижения в некоторых нидерландских таблицах необходимой степени детализации.

12. Страны Северной Европы (Дания, Исландия, Норвегия, Финляндия и Швеция) в настоящее время имеют больше переменных в регистрах, чем Нидерланды. Поэтому проблема недостаточной детализации в итоговых показателях не играет там существенной роли. Кроме того, некоторые страны Северной Европы провели (ограниченное) обследование для получения недостающих переменных в регистрах. Большинство других стран находится в сходном с Нидерландами положении, когда одни переменные, необходимые при проведении переписи, могут быть взяты из регистров, а другие переменные доступны только на базе выборочных обследований. Именно этим объясняется значительный интерес к нидерландскому подходу, основанному на комбинировании регистров и обследований и на применении современных статистических методов и компьютерных программ для составления таблиц. Разумеется, исключительно важно, чтобы статистические бюро имели возможность использовать регистры, релевантные для целей переписи. В случае Статистического управления Нидерландов эта возможность была расширена в новом законе о статистике 2003 года. Тем не менее в ближайшие годы Статистическому управлению Нидерландов необходимо будет наладить надежные контакты с кураторами регистров. Своевременное представление соответствующих переменных для Статистического управления Нидерландов имеет исключительно важное значение для составления статистики.

13. В цикле переписей 2000 года участвовали более 50 стран - членов Европейской экономической комиссии Организации Объединенных Наций. Многие страны выбрали один из дней в 2001 году в качестве своего базового дня, хотя эти базовые дни в основном не совпадают. Поскольку для завершения всеми странами своих таблиц, необходимых международным организациям, требуется немало времени, Нидерланды взяли на себя инициативу провести некоторые простые сопоставления между девятью европейскими странами, которые относительно быстро справились с подготовкой набора таблиц для Евростата и пожелали присоединиться к проведению сравнительных анализов (см. главу 12 в работе Schulte Nordholt et al., 2004). Кроме того, проводились сопоставления результатов переписи 2001 года в Нидерландах с результатами предыдущих общенациональных переписей Schulte Nordholt, 2005)

III. МЕТОДИКА ПРОВЕДЕНИЯ ПЕРЕПИСИ 2001 ГОДА

14. Настоящая виртуальная перепись относится к 2001 году. Базовым элементом переписи является центральный регистр населения (РН), который объединяет в себе все муниципальные регистры населения. Регистр населения (РН) содержит демографическую информацию обо всех жителях Нидерландов (Prins, 2000).

15. Даже если перед РН поставлена задача оптимальной регистрации данных о каждом гражданине, он не может претендовать на безупречность. Граждане могут менять место жительства и "забывать" уведомлять об этом власти. Вследствие этого муниципальные регистры населения не всегда содержат обновленную информацию. Еще одним примером ошибочной регистрации в РН является случай регистрации по различным адресам двух лиц, которые фактически проживают вместе. Они имеют финансовые стимулы для регистрации по различным адресам, если одно лицо является занятым, а второе получает пособия по линии социального обеспечения. Это объясняется тем, что лицо, получающее пособия утратит их, если орган социального обеспечения выяснит факт их совместного проживания.

16. Одной из важных групп населения, которая отсутствует в РН, являются лица, проживающие в стране без регистрации, причем многие из них нелегально. Эта группа населения не охватывается переписью населения. Нелегальные резиденты создают проблему для статистических органов, поскольку, с одной стороны, они участвуют в экономической деятельности и в качестве таковых учитываются в экономической статистике. С другой стороны, они не охватываются демографической статистикой. Однако вероятность их регистрации в рамках традиционной переписи является весьма малой. Статистическое управление Нидерландов предприняло попытку оценить численность нелегальных жителей, однако получение такой информации оказалось на практике весьма трудным делом. Подготовленная Статистическим управлением Нидерландов официальная оценка численности незаконных резидентов по состоянию на 1 января 2001 года характеризуется широким диапазоном: от 46 000 до 116 000 человек (Hoogteijling, 2002).

17. С РН увязан ряд комплексных обследований и регистров. Для данного процесса увязки используют только точные совпадения, опирающиеся на так называемый уникальный номер социального обеспечения и налогообложения (SoFi). Данная интегрированная система называется системой Базы данных социальной статистики (БДСС). Исходно она была разработана для проведения виртуальных переписей, но в настоящее время также используется для расчета различного рода показателей социальной статистики.

18. В качестве основы для подготовки набора таблиц использовались данные РН по состоянию на 1 января 2001 года. Различные переменные, такие, как показатели родов, занятие и уровень образования, были взяты из результатов обследования рабочей силы (ОРС). Для получения переменных по рабочим местам использовались результаты крупного обследования занятости и доходов (ОЗД). Для получения достаточных учетных данных была объединена информация по отдельным лицам из ОРС 2000 года и ОРС

2001 года. Для построения таблиц по жилищному фонду мы использовали данные РН по состоянию на 1 января 2001 года, регистр жилищного фонда 2001 года и обследование жилищных условий (ОЖУ) 2000 года.

19. Некоторые переменные из наборов данных РН и БДСС имеются в интегрированном виде. Примерами являются данные о возрасте, поле, семейном положении и занятости. В случае обследований переменные имеются только в отношении части населения. Примерами являются данные о наивысшем достигнутом уровне образования (ОРС) и данные о том, арендует ли респондент жилье, в котором он проживает, или является его владельцем (ОЖУ).

20. Для того чтобы каждая таблица могла быть рассчитана с максимально возможной точностью, в качестве основы для каждой оценки используется максимально возможное количество учетных данных. Таблицы, в которых содержатся только переменные, взятые из регистров, рассчитываются на основе регистровых данных. Таблицы, в которых содержится хотя бы один параметр, взятый из обследования, рассчитываются на основе максимально возможной комбинации регистров и обследований.

21. Для обеспечения непротиворечивости между таблицами мы использовали метод повторного взвешивания. Он позволяет получать новый набор весов по каждой расчетной таблице и опирается на повторное применение регрессионной оценки. При использовании этого метода весовые коэффициенты учетных записей в микроданных адаптируются таким образом, чтобы новое табличное значение согласовывалось со всеми предыдущими табличными значениями. При использовании этого метода весовые коэффициенты учетных записей в микроданных адаптируются таким образом, чтобы новое табличное значение согласовывалось со всеми предыдущими табличными значениями.

22. Данные переписи 2001 года относятся к лицам, проживавшим в Нидерландах по состоянию на 1 января 2001 года (учетные единицы). В виртуальной переписи были "учтены" лица, проживавшие в Нидерландах на начало этого дня согласно РН. Большинство голландского населения живет в частных домохозяйствах, а остальная часть - в институциональных домохозяйствах. Отраженная в таблицах численность лиц, работающих по найму, относится к концу 2000 года, и в отношении этого показателя 22 декабря 2000 года было выбрано в качестве базовой даты, с тем чтобы зафиксировать имевшееся в Нидерландах количество рабочих мест по найму. Установить базисный день в 2001 году для определения числа работающих по найму было невозможно, поскольку наборы данных БДСС за 2001 год не были своевременно подготовлены для использования в переписи 2001 года. Для данных БДСС использовалась информация о местах работы

работников по найму, взятая из регистров. Если какое-то лицо работает одновременно на нескольких работах, оно может несколько раз фигурировать в регистре занятых по найму. В наборе таблиц используются характеристики основной работы. Основная работа занятого по найму была определена как работа с самой высокой валовой заработной платой для целей социального страхования.

IV. КОМБИНИРОВАНИЕ ИСТОЧНИКОВ ДАННЫХ: УВЯЗКА НА МИКРОУРОВНЕ И МИКРОИНТЕГРАЦИЯ

A. Увязка на микроуровне

23. Большинство текущих административных регистров обладают единым ключом увязки. Речь идет о так называемом номере социального страхования и налогообложения (SoFi), который представляет собой личный идентификатор каждого (зарегистрированного) жителя Нидерландов, а также лиц, проживающих за границей, которые получают доход из Нидерландов и должны выплачивать на него соответствующий налог голландским налоговым органам.

24. Во избежание ошибочного использования номера SoFi, Статистическое управление Нидерландов производит его перекодирование в целях статистической обработки в так называемый регистрационный идентификационный номер (PIN). Личные идентификаторы, такие, как дата рождения и адрес, заменяются возрастом, контрольной датой и PIN-адресом. Это производится в соответствии с нормами Управления по защите данных Нидерландов с целью охраны личной жизни граждан.

25. Поскольку номер SoFi используется органами социального страхования и налоговыми органами, можно рассчитывать на то, что этот идентификатор обладает высоким качеством. Ограниченное количество номеров SoFi может регистрироваться с ошибочными значениями в файлах данных, в случае чего увязка с другими файлами может оказаться невозможной. Однако в целом процент совпадений близок к 100%. В некоторых случаях номера SoFi могут использоваться незаконно, например нелегальными работниками, что приводит к ошибочным совпадениям. В некоторых случаях имеются указания на ошибочную увязку. Одним из примеров может служить проведение увязки между регистром занятых и регистром населения, в результате которой оказывается, что работник является младенцем. Еще одним примером служит наличие в базе данных налоговой администрации сведений, указывающих на необычно высокий доход работника, в то время как на практике это является суммой доходов всех лиц, использующих один и тот же номер SoFi.

26. Все файлы данных социальной статистики могут быть увязаны с РН. На практике это означает, что эти файлы данных косвенно увязаны со всеми другими через РН. Таким образом, РН может рассматриваться в качестве опорного элемента набора источников данных социальной статистики. Благодаря увязке РН с регистром занятости или РН с регистром социальных пособий обеспечивается увязка различных статистических единиц (лиц, рабочих мест, пособий). В данном случае могут существовать множественные взаимосвязи, поскольку одно лицо может иметь несколько рабочих мест или получать несколько социальных пособий.

27. В ходе выборочных обследований домохозяйств, таких, как ОРС, данные не имеют номера SoFi. Для этих обследований используется альтернативный ключ увязки, который во многих случаях опирается на комбинацию следующих личных идентификаторов:

- a) пол;
- b) дата рождения;
- c) адрес¹.

28. Такой тип ключа увязки обычно позволяет эффективно производить дифференциацию лиц. Однако он не обеспечивает 100% уникальную комбинацию идентификаторов. Увязка может привести к ошибке в случае близнецов одного и того же пола. Ошибочные совпадения могут также возникать в тех случаях, когда часть даты рождения или почтового индекса и номера дома неизвестны или ошибочны. Еще одним недостатком является то, что ключом увязки является не лицо, а соответствующий адрес, что может привести к проблемам увязки в случае недавнего переезда. При увязке РН и ОРС с использованием данного оперативного ключа и допуске вариации между источниками только по одной из переменных - пол, год рождения, месяц рождения или дата рождения, результат будет близок к 100% увязке данных ОРС.

29. В рамках своей стратегии увязки Статистическое управление Нидерландов пытается максимизировать число совпадений и минимизировать число расхождений. Исходя из этого для обеспечения более высокого коэффициента увязки предпринимаются дополнительные усилия по увязке остающихся и несовпадающих записей с помощью различных вариантов ключей увязки, примером чему может служить как использование номера дома, так и допущение вариации числовой составляющей почтового индекса. Для удержания вероятности несовпадения на минимально низком уровне в процесс увязки

¹ На практике комбинация почтового индекса (главным образом относящегося к улице) и номера дома используется в качестве заменителя адреса. Почтовый индекс в Нидерландах состоит из четырех цифр, за которыми следуют две буквы.

встроены определенные "страховочные" механизмы. Они обеспечивают еще один процент совпадений.

30. В конечном итоге примерно от 2% до 3% данных ОРС не могут быть увязаны с РН. Это довольно хороший результат, однако избирательность процесса увязки на микроуровне может быть преодолена. Если совпадающие записи относятся к определенной подсовокупности, это может привести к систематической ошибке в оценке, опирающейся на увязанные данные, поскольку они не описывают все население. Проведенный в прошлом анализ свидетельствует о том, что молодые люди в возрасте 15-24 года характеризуются более низким коэффициентом увязки в рамках выборочных обследований домохозяйств по сравнению с другими возрастными группами. Причиной этого является более частая смена места жительства, в связи с чем они часто оказываются зарегистрированными по бывшему адресу. Коэффициент увязки в случае лиц, проживающих в четырех крупных городах - Амстердаме, Роттердаме, Гааге и Утрехте, - является меньшим по сравнению с жителями других населенных пунктов. Этнические меньшинства также характеризуются более низкой вероятностью увязки, в частности по той причине, что их дата рождения во многих случаях регистрируется менее точно (Arts et al., 2000).

31. В настоящее время РН служит основой выборки для ОРС. Вследствие этого коэффициент увязки составляет почти 100% и без возникновения каких-либо проблем селективности.

В. Микроинтеграция

31. Успешная увязка РН со всеми другими вышеупомянутыми источниками данных позволяет получать намного более не противоречивую информацию о различных демографических и социально-экономических аспектах жизни каждого лица. Однако необходимо учитывать, что некоторые источники являются более надежными по сравнению с другими. Одни источники обладают более полным охватом по сравнению с другими, в результате чего различные источники могут содержать даже противоречивую информацию. С учетом этого важно признать сильные и слабые стороны всех используемых источников данных.

33. Поскольку между источниками существуют различия, нам необходим процесс микроинтеграции для проверки данных и корректировки ошибочных данных. Считается, что интегрированные данные обеспечивают более надежные результаты, поскольку они опираются на оптимальный объем информации. Кроме того, охват (под) совокупности будет лучшим, если данные, отсутствующие в одном источнике, могут быть взяты из

другого источника. Еще одним преимуществом интеграции является то, что пользователи статистической информацией будут иметь одну величину, описывающую каждое социальное явление, вместо создающих путаницу нескольких различных величин, в зависимости от того, какой источник использовался.

34. Процесс микроинтеграции источников данных состоит из следующих этапов (Van der Laan, 2000):

- a) согласование статистических единиц;
- b) согласование отчетных периодов;
- c) полнота совокупности (охват);
- d) согласование переменных в случае различий в определениях;
- e) согласование классификаций;
- f) корректировка неточностей измерения, когда соответствующие переменные по-прежнему не имеют одинакового значения после устранения различий в определениях;
- g) условные расчеты в случае отсутствия данных;
- h) расчет (переменных); создание переменных на основе различных источников данных;
- i) проверки на общую непротиворечивость.

Все этапы контролируются с использованием набора правил интеграции и являются полностью автоматизированным.

35. Ниже приводится пример микроинтеграции, опирающийся на сопоставление данных из регистра рабочих мест с данными из регистра пособий. Как рабочие места, так и пособия регистрируются по факту, а это означает, что информация о состоянии этих переменных может относиться к любому моменту года, а не только к контрольной дате. Аналитикам подразделения по ведению регистра рабочих мест известно, что даты начала и окончания работы на том или ином месте страдают некоторыми неточностями. Вместе с тем важно знать, имело ли то или иное лицо работу по состоянию на контрольную дату,

другими словами, являлось ли или нет соответствующее лицо занятым по найму. Регистр пособий в некоторых случаях позволяет определить период занятости более точно.

36. Предположим, что какое-то лицо стало безработным в конце ноября и начало получать пособие по безработице с начала декабря. Регистр рабочих мест может указывать на то, что это лицо потеряло работу в конце года, что, возможно, объясняется административными задержками или выплатами после потери работы. Считается, что регистрация пособий является более точной. При сопоставлении этих фактов "интегратор" может принять решение об изменении даты потери работы на конец ноября, поскольку маловероятно, что данное лицо одновременно имело работу и получало пособие в декабре. Такие решения требуют исключительно тщательного подхода. Если существуют убедительные контрдоказательства, предоставляемые другими переменными регистра рабочих мест, свидетельствующие о том, что данное лицо продолжало работать и в декабре, дата окончания работы, как правило, не будет корректироваться.

V. БАЗА ДАННЫХ СОЦИАЛЬНОЙ СТАТИСТИКИ (БДСС)

37. Процесс увязки на микроуровне и микроинтеграции всех имеющихся источников данных находит свое конечное воплощение в Базе данных социальной статистики (БДСС), которая представляет полный набор интегрированных файлов микроданных в своем окончательном виде. БДСС содержит непротиворечивые подробные демографические, социально-экономические данные о лицах, домохозяйствах, местах работы и (социальных) пособиях. Значительная часть статистической информации зарегистрирована по факту. Подробный анализ БДСС может быть найден в работе Arts and Hoogteijling (2002).

38. Если попытаться представить, как выглядит БДСС, то не следует думать, что это крупномасштабный файл с миллионами учетных записей и тысячами переменными. Было бы весьма неэффективно хранить интегрированные данные в качестве таковых. Кроме того, вопрос защиты данных не позволяет Статистическому управлению Нидерландов хранить такой большой объем информации в одном месте. Интегрированные файлы данных в своем окончательном виде хранятся отдельно. Существует лишь один комбинирующий элемент, каковым является ключ увязки RIN, присутствующий в каждом интегрированном файле. Таким образом, когда пользователи запрашивают набор переменных из совокупности БДСС, им будут предоставлены только файлы с запрошенными переменными. Они могут легко извлекаться из общей совокупности и интегрироваться с помощью ключа увязки.

VI. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

39. Статистическое управление Нидерландов модернизировало свои методы сбора и обработки данных для целей программы составления таблиц переписи 2001 года. Наиболее важными элементами нового подхода являются использование комбинации административных регистров и результатов выборочных обследований в качестве источников данных и применение повторного взвешивания, представляющего собой новую методологию расчета непротиворечивых таблиц на основе этого источника данных. Результатом этой работы стала виртуальная перепись, поскольку результаты по некоторым признакам населения опираются на оценки, а не на регистрацию.

40. Новый способ разработки таблиц переписи доказал свою эффективность и является намного более дешевым по сравнению с дорогостоящими проектами переписей, осуществлявшимися в прошлом. Он не потребовал специальных мероприятий по сбору данных, поскольку источники данных, использовавшиеся в целях переписи 2001 года, уже являлись частью регулярной программы статистических работ Статистического управления Нидерландов. Большинство данных были взяты из регистров, и лишь небольшая дополнительная информация была получена с помощью выборочных обследований. Это означает, что реализация программы переписи 2001 года не создала дополнительной нагрузки на респондентов. Кроме того, продолжительность обработки данных в целях переписи 2001 года составила лишь незначительную часть от того объема времени, которого потребовала бы обработка в рамках традиционной переписи.

41. Одним из недостатков нового подхода является то, что размер выборки, используемый для проведения обследований, не всегда является достаточным для того, чтобы гарантировать надежные оценки по малым совокупностям, таким, как муниципалитеты. В рамках традиционной переписи такой проблемы не возникает.

42. Сопоставление нынешнего способа составления переписных таблиц с методами, использовавшимися в рамках переписей 1981 и 1991 годов, позволяет сделать вывод о том, что программа переписи 2001 года требует больше времени с точки зрения реализации, однако оценки в перекрестных таблицах данных регистров и результатов обследований являются более точными. Во-первых, непротиворечивость статистической информации намного возросла благодаря комбинированию источников данных. Во-вторых, эта методика позволяет использовать больше вспомогательной информации для расчетов оценок по сравнению с прошлым благодаря наличию большего числа регистров. В-третьих, намного больше усилий было посвящено обеспечению общей цифровой непротиворечивости.

Библиография

- Arts, C.H. and E.M.J. Hoogteijling. 2002. 'The Social Statistical Database of 1998 and 1999'. *Monthly Bulletin of Socio-economic Statistics*. Vol. 2002/12 (December 2002), pp. 13-21, 2002. [in Dutch]
- Corbey, P.I., 1994. 'Exit the population census'. *Netherlands Official Statistics*, Vol. 9 (Summer 1994); pp. 41-44, 1994.
- Hoogteijling, E.M.J. 'Illegal people in the Netherlands'. *Monthly Bulletin of Population Statistics*. Vol. 2002/03 (March 2002), page 21. [in Dutch]
- Laan, P. van der, 2000. 'Integrating Administrative Registers and Household Surveys'. *Netherlands Official Statistics*, Vol. 15 (Summer 2000): Special Issue, *Integrating Administrative Registers and Household Surveys*, ed. P.G. Al and B.F.M. Bakker, pp. 7-15.
- Prins, C.J.M., 2000. 'Dutch population statistics based on population register data'. *Monthly Bulletin of Population Statistics*. Vol. 2000/02 (February 2000), pp. 9-15.
- Schulte Nordholt, E., M. Hartgers and R. Gircour (Eds.), 2004. 'The Dutch Virtual Census of 2001, Analysis and Methodology'. Statistics Netherlands, Voorburg / Heerlen, July, 2004.
<http://www.cbs.nl/NR/rdonlyres/D1716A60-0D13-4281-BED6-3607514888AD/0/b572001.pdf>.
- Schulte Nordholt, E., 2005. 'The Dutch virtual Census 2001: A new approach by combining different sources'. *Statistical Journal of the United Nations Economic Commission for Europe*, Volume 22, Number 1, 2005, pp. 25-37.
- Statistics Netherlands, 2003. 'Urban Audit II, the implementation in the Netherlands'. Report, BPA no. 2192-03-SAV/II, Statistics Netherlands, Voorburg.
<http://www.cbs.nl/NR/rdonlyres/8C6E4C9D-4338-4E32-848B-8D43B9B3242D/0/urbanauditIINetherlands.pdf>.
- Vliegen, J.M. and H. van de Stadt, 1988. 'Is a Census still necessary? Experiences and alternatives'. *Netherlands Official Statistics*, Volume 3, No. 3, pp. 27-34.
