

Распространение: Общее

27 марта 2012 г.

Оригинал: английский

Европейская экономическая комиссия

Конференция европейских статистиков

Заседание экспертной группы ЕЭК ООН – Евростат по проведению переписи с использованием регистров

Женева, 22-23 мая 2012 года

Сессия 4: Методология оценки отсутствующей в регистрах информации

Методика, используемая для оценки таблиц переписей населения на основе неполной информации

Записка Статистического бюро Нидерландов¹

Резюме

Таблицы переписей населения в Нидерландах формируются на основании данных из множества различных источников. С тех пор, как последняя перепись на основе сплошного учета населения была проведена в 1971 г., готовность населения к участию в переписи резко снизилась. Статистическое бюро Нидерландов нашло альтернативу в виде виртуальной переписи населения, используя имеющиеся регистры и обследования. Виртуальная перепись дешевле по сравнению с ранее проведенными переписями в Нидерландах и социально более приемлема. Нидерланды занимают уникальное положение в европейской переписи населения раунда 2011 г.: нидерландский подход основан на регистрах и более не существует переписных листов, отсутствующая информация оценивается исходя из уже имеющихся исследований и в качестве ключевого метода подготовки набора таблиц переписи применяется методика повторного взвешивания. Результаты таблицы сопоставимы не только с предыдущими переписями в Нидерландах, но и с другими европейскими странами, принимающими участие в переписи раунда 2011 г.

¹ Автор – Eric Schulte Nordholt

I. Введение

1. В Нидерландах микроданные объединяются в таблицы переписи населения через два года после года переписи. В Нидерландах это осуществляется не путем сплошного опроса проживающих лиц, а на основании данных, уже имеющихся в Статистическом бюро Нидерландов. Таким образом, голландский налогоплательщик платит намного меньше за проведение переписи. Расходы на традиционную перепись составили бы несколько сотен миллионов евро, тогда как сейчас затраты составляют «всего» несколько миллионов. Сюда входят расходы на все подготовительные работы, например, разработку, обновление и тестирование методологии и сопутствующего программного обеспечения. Расходы, связанные с регистрами, не включаются, в отличие от анализа результатов. Регистры обновляются не для переписей населения, а для других целей. Страны, проводящие традиционную перепись населения, часто обосновывают огромные расходы на проведение переписей теми последствиями, которые результаты переписи имеют на распределение государственных средств между регионами. Кроме того, проведение виртуальных переписей было бы во многих странах невозможно в связи с отсутствием достаточных данных в регистрах.

2. Помимо финансового аспекта, существуют и другие важные различия между традиционной и виртуальной переписями населения, проводимыми в Нидерландах. Несмотря на обязательность традиционной переписи, определенная часть населения не участвует (выпадение ответов на уровне единиц), а часть, которая участвует, не отвечает на некоторые вопросы (выпадение ответов на вопросы). Стоит попробовать провести корректировку выпадения ответов методами взвешивания и вменения. Известная проблема с традиционными переписями состоит в том, что участие носит ограниченный и избирательный характер. Традиционные методы корректировки не дают требуемых достоверных результатов. Последняя традиционная перепись населения в Нидерландах (в 1971 г.) столкнулась с большим сопротивлением вторжению в частную жизнь при сборе информации о населении, проживающем в Нидерландах. Это усугубило проблему с отсутствием ответов и привело к ожиданию, что непредставление ответов на вопросы переписи будет еще более распространенным при проведении очередной традиционной переписи в Нидерландах (Corbey, 1994). Виртуальная перепись населения не вызвала почти никаких возражений и проблема с отсутствием ответов играет свою роль только при проведении обследований, данные которых используются. Если отсутствие ответов может быть скорректировано при обследовании, то его можно корректировать на избирательность такого обследования в ходе переписи, где оно используется.

3. Виртуальная перепись в Нидерландах начинается позднее, чем в других странах, где проводятся традиционные переписи населения. На самом деле нет смысла начинать основной этап Проекта переписи населения, пока не будут доступны все источники; некоторые регистры становятся доступными сравнительно поздно. Тем не менее, Нидерланды, вероятно, подготовят таблицы переписи населения быстрее большинства других стран, участвующих в европейском раунде переписи населения. На самом деле, в Нидерландах сроки передачи полного набора таблиц в Евростат, координирующий усилия всех стран-членов Европейского союза (ЕС) и Европейской ассоциации свободной торговли (ЕАСТ), одни из самых коротких. У Нидерландов есть то преимущество, что не надо проверять и корректировать поступающие переписные листы. Однако необходимо понимать, что для некоторых переменных имеется только выборочная информация, что означает невозможность соблюдения уровня детализации в некоторых таблицах переписи населения.

4. В настоящее время в Нидерландах преимущества виртуальной переписи населения в отношении затрат и проблем с непредставлением ответов компенсируют отсутствие некоторых деталей по сравнению с традиционной переписью. Кроме того, не вся необходимая информация всегда будет доступна для пользователей в традиционных переписях населения. Это связано с тем, что традиционные методы корректировки, такие как взвешивание и вменение, иногда не делают поправку на ограниченное и избирательное участие. Это означает, что по некоторым ячейкам в таблицах нельзя получить достоверные результаты.

5. В четырех скандинавских странах (Дания, Финляндия, Норвегия и Швеция) переменных в регистрах больше, чем в Нидерландах. Таким образом, здесь проблема недостаточной детализации результата не играет существенной роли. Кроме того, в Австрии и Словении большинство переменных переписи доступны в регистрах. Большинство других стран находятся в таком же положении, что и Нидерланды, где некоторые переменные, важные для переписи населения, можно найти в регистрах, в то время как другие переменные имеются только на выборочной основе. Именно поэтому вызывает интерес голландский подход с объединением регистров и обследований и использованием современных статистических методов и сопутствующего программного обеспечения для составления таблиц. Конечно, важно, чтобы статистические бюро могли использовать регистры, имеющие отношение к переписи населения. Для Статистического бюро Нидерландов это предусмотрено новым законом о статистике, который вступил в силу в 2004 г. Тем не менее, в ближайшие годы Статистическому бюро Нидерландов придется поддерживать хорошие связи с организациями, которым принадлежат регистры, созданные за последние пятнадцать лет. Своевременное представление соответствующих переменных для Статистического бюро Нидерландов очень важно для получения статистических данных.

6. В Нидерландах перепись населения 1971 г. была последней переписью, в которой использовались переписные листы. В 1970-ые гг. Парламент Нидерландов принял решение о том, что традиционные переписи больше не будут проводиться в Нидерландах. В то же время огромный интерес к результатам переписи населения сохранялся. Данные переписи населения, собранные в 1981 и 1991 гг., были гораздо менее подробны, чем данные за 1971 г. и набор таблиц виртуальной переписи населения 2001 г. Для получения более подробных таблиц для переписи населения 1981 г. для обследования рабочей силы этого года использовалась более высокая доля выборки, чем обычно. Позже, такая дополнительная выборка ОРС для целей переписи не выполнялась, и при переписях использовались только обычные данные ОРС. Перепись населения 1991 г. была в значительной степени основана на регистрах в сочетании с обследованием рабочей силы 1991 г. и обследованием спроса на жилье 1989/1990 гг. В 1991 г. Закон о переписи населения утратил силу, официально отменив обязанность Статистического бюро Нидерландов проводить перепись населения один раз в десять лет (Corbey, 1994). Обязательство предоставлять данные переписи за 2001 г. не было предусмотрено для европейских стран, но возможность отказа Нидерландов от сбора данных переписи населения для международных организаций в отличие от других европейских стран не рассматривалась. В отличие от 1981 и 1991 гг., Статистическое бюро Нидерландов опубликовало данные переписи населения за 2001 г. на муниципальном уровне.

7. Причина, по которой Статистическое бюро Нидерландов составило набор таблиц переписи населения 2001 г., связана с джентльменским соглашением с Евростатом, позволявшим сопоставить таблицы Переписи населения 2001 г. среди европейских стран. Результаты переписи 2001 г. в Нидерландах также были сопоставлены с предыдущими переписями населения в стране. Такая работа проводилась и в прошлом. Общее ощущение после проведения переписи населения

2001 г. в Европе было таким, что джентльменского соглашения для продолжения проведения десятилетних переписей в Европе было недостаточно. Это послужило основанием для придания официального статуса Раунду переписи населения 2011 г. посредством четырех регламентов (Европейская комиссия, 2008, 2009, 2010а и 2010б). Эти четыре регламента гармонизировали определения населения, переменные переписи и их категории, гиперкубы переписи (многомерные таблицы) и метаданные в рамках ЕС. Кроме того, был уточнен технический формат, в котором должны быть представлены данные; также все страны должны подготовить качественный отчет с описанием примененной методологии.

II. Методика составления

8. Статистическое бюро Нидерландов проводит виртуальные переписи с 1981 г. Основу таких переписей составляет Центральный регистр населения (РН), объединяющий все муниципальные регистры населения. Результаты переписи населения охватывают лиц, проживающих в Нидерландах на день проведения переписи. Лица, проживающие в Нидерландах на начало этого дня в соответствии с РН, «учитываются» в виртуальной переписи. Большинство людей в Нидерландах живет в частных домохозяйствах, часть проживает в институциональных домохозяйствах. В 2001 и 2011 г. данные переписи РН на 1 января (день Переписи) используются в качестве основы для набора таблиц. Набор таблиц сфокусирован на учете частотности, а не на количественной информации.

9. В девяностые годы прошлого века в Статистическом бюро Нидерландов была создана система социальной статистической базы данных (ССБД). ССБД включает в себя интегрированные микроданные по занятому и samozанятому населению. ССБД представляет собой набор интегрированных файлов микроданных с последовательными и подробными демографическими и социально-экономическими данными о людях, домашних хозяйствах, рабочих местах, пособиях и жилье. Благодаря выполнению микроинтеграционных процессов до сохранения данных, ССБД не содержит внутренних противоречивых данных. Для переписи населения 2001 года некоторые переменные (например, количество рабочих мест) были получены из крупного Обследования занятости и доходов (ОЗД). Проведение такого обследования было прекращено в 2006 г. и с тех пор Статистическое бюро Нидерландов больше полагается на фискальные данные и данные о социальном обеспечении, которые были включены в ССБД. Численность населения, работающего по найму, в таблицах дано на конец года, предшествующего году переписи. Для определения конца года в качестве справочной даты для закрепления количества рабочих мест занятого населения в Нидерландах используется день в декабре. Какой-либо день в год проведения переписи не может быть использован в качестве справочной даты для численности занятого населения, поскольку базы данных ССБД на год переписи еще не доступны на момент проведения переписи. ССБД использовала информацию о рабочих местах занятого населения из регистров. Если одно лицо работает на нескольких работах одновременно, он(-а) может несколько раз учитываться в регистре занятого населения. В наборе таблиц используются характеристики основной работы, где основное место работы было определено как работа с самой высокой валовой заработной платой для целей социального страхования.

10. Различные переменные, например, род деятельности и уровень образования, взяты из обследования рабочей силы (ОРС). Для переменной «уровень образования» в переписи населения 2011 г. помимо ОРС частично используется информация из регистров (на основе результатов обследования). Чтобы получить достаточно

записей, информацию о людях из ОРС в год, предшествующий году переписи объединяют с ОРС, проводимым в год переписи населения. Таким образом, предполагается, что показатели по переменным ОРС стабильны в течение двухлетнего периода. Между переписями 2001 и 2011 гг. ОРС изменило форму, став панельным обследованием. Сейчас также можно использовать панельный характер обследования: имеются данные из разных волн и данные ближе к дню переписи населения могут быть использованы для составления таблиц.

11. Для таблиц жилья в переписи 2001 г. мы использовали данные РН на 1 января 2001 г., регистра жилищного фонда 2001 г. и обследования жилищных условий (SHC) 2000 г. Для переписи 2011 г. информация о жилье взята из регистра жилья и новых источников, таких как Базовые регистрации адресов и зданий (BAV). Для таблиц 2001 г. о пригородном транспортном сообщении мы использовали данные РН, ОЗД 2000 г. и данные ССБД 2000 г. Для таблиц 2011 г. о пригородном транспортном сообщении мы будем опираться на данные РН по месту жительства и данные ССБД по месту работы.

12. Переменные регистров данных РН и ССБД доступны в интегрированном виде, например, возраст, пол, семейное положение и занятость. Переменные обследования доступны только для части населения, например, самый высокий уровень полученного образования и проживание в арендуемом или собственном жилье. Австрия, Дания, Финляндия, Норвегия, Словения и Швеция недавно сообщили, что они планируют провести Перепись 2011 г. полностью на основе регистров. Хотя у всех них нет некоторых данных, объем такой отсутствующей информации будет меньше, чем в Нидерландах. Таким образом, для компенсации соответствующей недостающей информации применяются методы вменения. Для Нидерландов этот вариант не подходит, поскольку у нас отсутствует немного больше информации, чем в шести странах, проводящих Перепись 2011 года полностью на основе регистров.

13. Еще одну проблему для стран, работающих на основе регистров, составляет разница между официальными определениями переменных переписи 2011 года и переменных в регистрах. Важным примером является численность безработных. По определению Международной организации труда (МОТ) люди могут считаться безработными только в том случае, если они активно ищут работу. Из регистра, как правило, нельзя сделать вывод, действительно ли получатель социальной выплаты и пособия по безработице активно ищет работу. Кроме того, подход, основанный на регистрах, не учитывает безработных, которые не получают социальное пособие (например, потому что другие члены семьи получают достаточный доход). Для переписи населения 2001 г. в Нидерландах было использовано официальное определение МОТ, и результаты выборки были применены ко всему населению. Учитывая более детальные гиперкубы в Переписи раунда 2011 г., Статистическое управление Нидерландов серьезно рассматривает возможность применения определения безработицы только на основе регистров. Таким образом, можно добиться уровня детализации в таблицах по текущей активности (для которых безработица является одной из категорий). Если Статистическое бюро Нидерландов действительно примет такое решение, информация о безработице в гиперкубах переписи Статистического бюро Нидерландов будет сопоставима с другими странами, работающими на основе регистров, но сравнение с другими европейскими странами станет более проблематичным.

III. Повторное взвешивание

14. Может возникнуть вопрос, почему для решения проблемы отсутствующих данных просто не применить метод массового вменения (заполнение допустимых

значений для всех недостающих показателей). Важным преимуществом метода массового вменения является то, что после него любой пользователь сможет воспроизвести результаты при использовании этого файла с вменением. Тем не менее, метод вменения неэффективен для применения результатов обследования ко всему населению. Степеней свободы для поддержания достаточно емкой модели вменения, которая бы учитывала все значимые модели данных между выборкой и регистровыми переменными, недостаточно. Только в том случае, если интерес представляет расчет итогов подмножеств совокупности, определяемых объясняющими переменными в модели, подход вменения приводит к несмещенным на структуру выборки, а, следовательно, и надежным оценкам (по крайней мере, если дисперсии достаточно малы) (Kroese и Renssen, 2000).

15. Основным методом Переписи населения 2011 г. в Нидерландах является повторное взвешивание (RW), описанное Houbiers и др. (2003) и Houbiers (2004). Цель повторного взвешивания состоит том, чтобы исключить численное несоответствие между табличными оценками из разных источников. Он основан на повторном применении регрессионной оценки и дает новый набор весовых коэффициентов для каждой оцениваемой таблицы. Пусть y – переменная, из которой должны быть получены параметры совокупности – общие или средние – посредством набора объясняющих переменных x из регистра. Оценка линейной регрессии среднего значения по совокупности для y определяется следующим образом:

$$\hat{Y}_{REG} = \hat{Y}_d + b'_s (\bar{X}_p - \hat{X}_d); \quad b_s = (X'_s D_s X_s)^{-1} X'_s D_s y_s,$$

где \bar{X}_p and \bar{Y}_p – средние значения по совокупности для x и y соответственно, в то время как \hat{X}_d and \hat{Y}_d это их оценки, основанные на расчетных весовых коэффициентах, а b_s – оценочный вектор коэффициентов регрессии. X_s – матрица выборочных наблюдений по x -переменным, а y_s – вектор наблюдений по переменной y . D_s – диагональная матрица с расчетными весовыми коэффициентами. Вместо этих традиционных оценок регрессии, в процедуре повторного взвешивания используется набор коэффициентов в виде

$$b_w = (Z'_s W_s Z_s)^{-1} Z'_s W_s y_s,$$

где Z_s – матрица выборочных наблюдений по переменным на полях таблицы с переменной y , а W_s – диагональная матрица с регрессионными весовыми коэффициентами последнего шага взвешивания для расчета оценки регрессии. Средние значения предельных переменных z были уже оценены в предыдущей таблице или известны из регистра. Обозначив эти оценки или встречаемость в регистре как \hat{Z}_{RW} , оценка повторного взвешивания определяется следующим образом:

$$\hat{Y}_{RW} = \hat{Y}_{REG} + b_w (\hat{Z}_{RW} - \hat{Z}_{REG}).$$

Можно показать, что весовые коэффициенты записей в микроданных адаптируются таким образом, что новая табличная оценка соответствует всем предыдущим табличным расчетам (Knottnerus и Van Duin, 2006).

16. Для применения метода повторного взвешивания мы используем последнюю версию (2.0) программного пакета VRD, разработанного Статистическим бюро Нидерландов. Буквы VRD обозначают Vullen Reference Database (Справочная база

данных для заполнения), и цель данного приложения состоит в заполнении и управлении справочной базой данных. Основными функциями VRD являются оценка таблиц с помощью повторного взвешивания, добавление этих таблиц в справочную базу данных, а также извлечение агрегированных показателей из справочной базы данных. При условии небольших независимых выборок можно оценить дисперсии табличных значений. Оценка таблиц происходит не в самом VRD, а в Bascula автоматически, при этом пользователь VRD не видит этого в явном виде. Оценка таблиц и дисперсий может быть сделана в пакетном или интерактивном режиме.

17. Для того, чтобы как можно точнее оценить каждую таблицу, каждая оценка основана на максимально возможном количестве записей. Таблицы, которые содержат только переменные из регистров, рассчитываются из регистров. Таблицы, содержащие, как минимум, одну переменную из обследования, оцениваются из самой большой возможной комбинации регистров и обследований. Сочетание регистров и обследований образует блоки, из которых оцениваются переписные таблицы.

18. Для блоков, составляемых на основе данных обследования, весовые коэффициенты должны определяться для получения оценок для всего населения. Эти весовые коэффициенты зависят от:

- точного состава соответствующего блока (один или несколько обследований);
- дизайна обследования (-й);
- корректировки обследования (-й) на отсутствие ответов;
- уменьшения дисперсии с помощью вспомогательной информации;
- достижения согласованности.

19. Полная согласованность не всегда возможна, например, при наложении слишком большого количества ограничений. В некоторых случаях возможна полная согласованность, но это приводит к очень большому колебанию весов и, следовательно, резко увеличивает дисперсию. В этих случаях лучше ограничить публикуемую детализацию.

20. При составлении переписных таблиц мы адаптируем веса блоков на каждом этапе VRD с помощью всех регистров и таблиц, оцененных ранее из блоков. Таким образом, все таблицы взаимно согласованы. Каждая таблица должна быть рассчитана из крупнейшего блока, из которого может быть определена таблица. Если все таблицы оцениваются таким образом с корректными весовыми коэффициентами, результаты таблиц взаимно согласованы. Начиная каждый раз с крупнейшего блока, получаются самые подробные возможные таблицы переписи.

21. Результаты пяти модельных исследований, тестирующих различные аспекты повторного взвешивания, можно найти в работе Van Duin и Snijders (2003). Knottnerus и Van Duin (2006) приводят формулы дисперсии для оценки повторного взвешивания (RW) и тестируют оценки RW в различных условиях. Использование оцениваемых полей результатов оценок для определения того, какие ячейки можно опубликовать, а какие нужно скрыть, объясняется в следующем разделе.

22. Остается открытым вопрос о том, как можно оценить небольшие территории случае отсутствия доступной информации из регистра. Текущее исследование проводится по дополнительным методам для минимизации проблем пустых ячеек. Нули в ячейках, основанные на выборочных данных, не обязательно означают, что данные на уровне населения отсутствуют. Тем не менее, со стандартными методами взвешивания нули в выборке не могут быть увеличены до положительного значения ячейки в совокупности.

23. Интересным вариантом является использование методов оценки небольших территорий для оценки значений ячеек, которые нельзя оценить адекватно. Оценка

небольших территорий, т.е. для получения действительной и эффективной оценки параметров населения для субнациональных территорий, как географических территорий, так и категорий в классификациях на дезагрегированном уровне, представляет собой задачу, которую можно правильно выполнить с помощью административных регистров. Для эффективной оценки небольших территорий записи из административных источников могут играть роль вспомогательной информации. Теоретическую основу для оценки небольших территорий можно найти у Rao (2003). Национальное статистическое управление (ONS) в Великобритании исследовало применение этого метода в контексте своей Программы районной статистики. Это важная инициатива для сведения и предоставления статистических данных на уровне небольших районов. В каждом случае проведения косвенных оценок небольших территорий особое внимание уделялось спецификации модели. Некоторые экспериментальные синтетические оценки были опубликованы в Великобритании, другие же находятся в процессе оценки. Возможно, что методы повторного взвешивания и оценки малых территорий можно объединить в раунде переписи населения 2011 г. Тем не менее, вопрос о том, как сохранить согласованность набора таблиц, остается открытым.

24. Еще одной интересной опцией является макроинтеграция (Mushkudiani, Daalmans и Pannekoek, 2012). Макроинтеграция широко используется для согласования макроданных, как правило, в виде крупных многомерных таблиц, полученных из различных источников. Традиционно эти методы широко применяются в области макроэкономики, особенно при составлении национальных счетов. Методы макроинтеграции получили со временем развитие и превратились в разнообразные техники для разрешения интеграции данных из различных источников на макроуровне. Проводимый в настоящее время проект исследования в Статистическом управлении Нидерландов делает попытки применять методы макроинтеграции при согласовании таблиц виртуальной переписи.

IV. Как публиковать таблицы?

25. Голландская перепись населения частично составлена на основе выборочных данных. Таким образом, для некоторых результатов переписи следует учитывать пределы погрешности. Благодаря надежности результатов, для значений ячеек, основанных на выборке из переписи населения, применяются эмпирические правила. Точные пределы погрешности не могут быть даны, поскольку блоки составлены на основании обследований и из-за сложной конструкции этих обследований. Эмпирические правила выводятся на основе допущений о том, что два набора данных ОРС (за год переписи и год, предшествующий году переписи) формируют одну выборку и что «вероятности включения» для данной выборки заданы весовыми коэффициентами блока ОРС. Эмпирические правила в Переписи населения 2001 г. для записей наблюдений из ОРС выглядят следующим образом:

- ячейки таблицы с данными по менее, чем 10 человекам, всегда убираются.
- ячейки таблицы с данными по 25 или более человекам всегда публикуются.
- ячейки таблицы с данными по 10-24 человекам публикуются только в том случае, если они являются частью разбивки (по возрасту или полу), в которой нет ячеек с данными по менее, чем 10 человекам и минимум 50 процентов ячеек в разбивке имеют данные по более, чем 25 человекам. Порог в 25 человек соответствует расчетной относительной погрешности не более 20 процентов (т.е. оцениваемые пределы составляют максимум до 40 процентов).

26. Эмпирические правила для записей из SHC имеют одинаковую форму. Тем не менее, применяются несколько более высокие пороговые значения в связи с тем, что размер выборки SHC несколько более ограничен, чем в ОРС. Для табличных ячеек, в которых домохозяйства или жилища являются счетной единицей, для голландской переписи населения 2001 г. применяются аналогичные эмпирические правила.

27. Следует ожидать, что для переписи 2011 г. могут быть получены аналогичные эмпирические правила. Проблемы с конфиденциальностью во многом будут решены, когда все решения о скрытии ячейки таблицы принимаются согласно этим эмпирическим правилам. Большинство переменных переписи не чувствительны и большинство переписных таблиц являются частотными таблицами, где вопросы конфиденциальности меньше влияют на детальность результата, чем на количественные таблицы. Различные варианты должной защиты всего результата Раунда переписи 2011 года будут обсуждаться на семинаре ESSnet по контролю за статистическим раскрытием (SDC) для данных переписи населения, который состоится в Люксембурге 19-20 апреля 2012 г. Этот семинар является частью ESSnet по общим инструментам и согласованной методологии SDC в рамках Европейской статистической системы (ЕСС).

28. Все государства-члены ЕС должны провести перепись населения 2011 г. Для большинства национальных статистических институтов это стало масштабной операцией, потребовавшей большого объема работы и затрат. Все страны должны закрепить и защитить результаты переписи населения в виде гиперкубов. Даже форматы, используемые для передачи данных, отличаются в зависимости от страны. Тем не менее, в конце концов, все данные должны быть преобразованы в формат SDMX и переданы в Евростат. Евростат выпустил DSD (определения структуры данных) и сделал их доступными для всех государств-членов. Индивидуальные проверки стран на предмет конфиденциальности оцениваются в контексте данной ESSnet. Хотя фактически все гиперкубы будут переданы в Евростат только в 2014 г., к этому сроку должны быть предприняты все меры защиты.

29. Будет полезно, если европейские страны обмениваются опытом и переймут друг у друга подход к конфиденциальности переписи населения. Сейчас мы сталкиваемся с риском, что многие страны осторожничают и защищают слишком много информации. Это может затруднить подведение европейских итогов. Кроме того, ситуация, когда во всех странах скрываются разные ячейки в гиперкубах, может привести к тому, что некоторые суммы нельзя рассчитать на европейском уровне. Семинар ESSnet в апреле 2012 г. также проводится для решения этой проблемы. На этот семинар были приглашены докладчики из разных стран для того, чтобы они высказали свое мнение по проблеме конфиденциальности европейской переписи населения. После этого будет организовано открытое обсуждение, в качестве модератора на котором будет выступать сотрудник Евростата. Этот семинар будет информативным и даст рекомендации по методам сохранения конфиденциальности для европейской переписи населения. Это позволит сделать огромный шаг вперед в согласовании защищаемого результата расчетов между европейскими странами. Таким образом, при минимальных потерях информации можно получить все более сопоставимые результаты переписи населения.

V. Выводы

30. Виртуальная перепись оказалась успешной концепцией в Нидерландах. Она имеет много преимуществ по сравнению с традиционной переписью населения. Расходы на перепись в настоящее время значительно ниже, при этом получаемые в результате данные о Нидерландах могут быть сопоставлены с результатами

предыдущих голландских переписей населения и результатами других стран, принимающих участие в том же раунде переписей. Статистическое управление Нидерландов в настоящее время проводит уже четвертую виртуальную перепись населения. Тем не менее, голландские данные, полученные в 1981 и 1991 гг., носили гораздо более ограниченный характер, чем набор таблиц в переписях населения 2001 и 2011 гг. Кроме того, они были в значительной степени основаны на регистрационном учете населения в сочетании с уже существующими обследованиями рабочей силы и жилищных условий.

31. Метод повторного взвешивания успешно используется для получения последовательного набора таблиц переписи населения 2001 г., а также будет основным методом для переписи 2011 г. Каждая таблица рассчитывается из крупнейшего блока, из которого может быть определена таблица. Все таблицы для переписи населения оцениваются таким же образом с корректными весовыми коэффициентами, и поэтому результаты таблицы взаимно согласованы. Начиная каждый раз с крупнейшего блока, получаются самые подробные возможные таблицы переписи. До составления таблиц этим методом свое значение по-прежнему сохраняет микроинтеграция различных источников в ССБД. В процессе микроинтеграции данные проверяются, и некорректные данные адаптируются. Существует твердое убеждение, что микроинтегрированные данные обеспечат более надежные результаты, поскольку они основаны на максимальном количестве информации. Кроме того, охват подгрупп населения будет лучше, потому что при отсутствии данных в одном источнике можно использовать другой источник. Еще одно преимущество микроинтеграции и повторного взвешивания состоит в том, что пользователи статистической информации больше не будут путаться, поскольку по каждому социально-экономическому явлению будет приведена одна цифра вместо нескольких цифр в зависимости от использованных источников.

Библиография

Corbey, P., 1994. Exit the population Census. *Netherlands Official Statistics*, Volume 9, summer 1994, pp. 41-44.

Duin, C. van and V. Snijders, 2003. Simulation studies of repeated weighting. Discussion paper 03008, Statistics Netherlands, Voorburg / Heerlen. <http://www.cbs.nl/NR/rdonlyres/203C85C6-7075-47A0-97BA-A3B748D393FE/0/Discussionpaper03008.pdf>

European Commission, 2008. Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses. Official Journal of the European Union, L218, pp. 14-20.

European Commission, 2009. Commission Regulation (EC) No 1201/2009 of 30 November 2009 implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdown. Official Journal of the European Union, L329, pp. 29-68.

European Commission, 2010a. Commission Regulation (EU) No 519/2010 of 16 June 2010 adopting the programme of the statistical data and of the metadata for population and housing censuses provided for by Regulation (EC) No 763/2008 of the European Parliament and of the Council. Official Journal of the European Union, L151, pp. 1-13.

European Commission, 2010b. Commission Regulation (EU) No 1151/2010 of 8 December 2010 implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses, as regards the modalities and structure of the

quality reports and the technical format for data transmission. Official Journal of the European Union, L324, pp. 1-12.

Houbiers, M., 2004. Towards a social statistical database and unified estimates at Statistics Netherlands. *Journal of Official Statistics*, Volume 20, No. 1, pp. 55-75.

Houbiers, M., P. Knottnerus, A.H. Kroese, R.H. Renssen and V. Snijders, 2003. Estimating consistent table sets: position paper on repeated weighting. Discussion paper 03005, Statistics Netherlands, Voorburg / Heerlen. <http://www.cbs.nl/NR/rdonlyres/6C31D31C-831F-41E5-8A94-7F321297ADB8/0/discussionpaper03005.pdf>

Knottnerus, P. and C. van Duin, 2006. Variances in repeated weighting with an application to the Dutch Labour Force Survey. *Journal of Official Statistics*, Volume 22, No. 3, pp. 565-584.

Kroese, A.H. and R. H. Renssen, 2000. New applications of old weighting techniques, constructing a consistent set of estimates based on data from different sources. *ICES II, Proceedings of the second international conference on establishment surveys, survey methods for businesses, farms, and institutions, invited papers*, June 17-21, 2000, Buffalo, New York, American Statistical Association, Alexandria, Virginia, United States, pp. 831-840.

Mushkudiani, N., J. Daalmans and J. Pannekoek, 2012. Macro-integration techniques with applications to census tables and labour market statistics. Discussion paper, Statistics Netherlands, The Hague / Heerlen. <http://www.cbs.nl/NR/rdonlyres/AD653253-647D-4FFD-AFC4-67E2BDE602EE/0/201201x10pub.pdf>

Rao, J.N.K., 2003. Small area estimation. Wiley, New York, United States.
