

Распространение: Общее  
27 марта 2012 г.

Оригинал: английский

---

## Европейская экономическая комиссия

### Конференция статистиков Европы

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

Женева, 22-23 мая 2012 г.

Сессия 3: Наличие, полнота и качество данных из регистров и других источников

### Качество регистров, используемых в переписи в Нидерландах

#### Записка Статистического бюро Нидерландов<sup>1</sup>

##### *Резюме*

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. Однако комбинированное использование регистров и обследований в составе переписи также приводит к нескольким сложным задачам методологического характера. Одной из них является определение влияния качества источников. Например, сбор и ведение регистров выходит за рамки компетенции Статистического бюро. Поэтому важно, чтобы Статистическое бюро могло определять уровень качества используемых источников. Понимание качества используемых источников позволяет проводить тщательно продуманное сопоставление сравнимой информации в разных источниках.

Независимо от исследований по усовершенствованию виртуальной переписи Статистическое бюро Нидерландов активно занимается разработкой систематических методов определения качества регистров. В этих целях были разработаны рамки качества. Такие рамки состоят из трех точек зрения высокого

---

<sup>1</sup> Авторы - Eric Schulte Nordholt, Saskia Ossen и Piet Daas.

уровня на качество административных данных: с точки зрения источника, с точки зрения метаданных и с точки зрения данных. Для систематической оценки первых точек зрения был подготовлен контрольный список, который уже успешно применялся к нескольким источникам данных. В настоящее время основное внимание в рамках исследований направлено на разработку способов систематической оценки качества данных.

В данном документе мы рассматриваем возможности для использования полученных в ходе исследований по качеству регистров знаний для улучшения решений, принимаемых в виртуальной переписи о том, какие источники следует использовать для получения или оценки каких переменных. Это также полезно для исследований о качестве, поскольку таким образом можно протестировать применим ли разработанный подход к переписям.

## I. Введение

1. Все страны Европейского Союза (ЕС) будут проводить перепись в 2011 г. То, каким образом будет проводиться перепись, остается на усмотрение страны. Со времени проведения последней традиционной переписи в 1971 г. в Нидерландах проводится виртуальная перепись. Это означает, что переписные листы более не применяются, а вся соответствующая информация предоставляется из данных уже существующих регистров и обследований (Schulte Nordholt, 2004). Именно таким образом были проведены виртуальные переписи в 1981, 1991 и 2001 гг. Переписи в 1981 и 1991 гг. носили ограниченный характер. Данные, составленные по 1981 и 1991 гг., были гораздо менее детализированы по сравнению с таблицами переписи 2001 г. В 2001 г. Статистическое бюро опубликовало информацию по переписи на муниципальном уровне. При проведении переписи 2011 г. будет задействовано еще больше регистров и обследований. Регистр населения является основой для работы по интеграции, которая в результате позволит получить логически связанную и детализированную демографическую и социально-экономическую статистическую информацию по лицам и домохозяйствам.

2. Общая проблема при использовании административных регистров в статистических целях заключается в том, что данные в этих источниках собираются и ведутся другими организациями в отличных от статистических целях. Этот процесс выходит за рамки компетенции Статистического бюро Нидерландов. Это не только обуславливает высокую зависимость Статистического бюро Нидерландов, но также может повлиять на качество результатов деятельности Статистического бюро. Поскольку ожидается, что в будущем Статистическое бюро Нидерландов будет использовать большее число регистров в целях снижения административного бремени, были разработаны рамки качества, которые позволяют определять уровень качества внешних источников данных, таких как регистры, до их использования (Daas et al., 2009). Эти рамки были применены для изучения качества входных данных наиболее важных регистров, использованных при виртуальной переписи 2011 г. Результаты этих исследований и являются предметом данного документа. В следующих разделах представлены источники данных и переменные переписи 2011 г. в Нидерландах. В Разделе 3 более подробно описываются рамки качества. Далее обсуждаются результаты применения рамок качества. И, наконец, в Разделе 5 представлены некоторые выводы.

## II. Источники данных и переменные

3. Регистр населения (РН) составляет основу переписи. Информация из других регистров и обследований дополняется для получения в итоге всех переменных переписи 2011 г. Важно осознавать, что регистры меняются с течением времени, равно как и их качество. Например, при проведении переписи в 2001 г. новый регистр жилья (РЖ) еще отсутствовал, но он будет использоваться при переписи 2011 г. Ожидается, что часть информации в новом РЖ может заменить информацию, которая в рамках переписи 2001 г. обеспечивалась двумя другими источниками данных; а именно старым регистром жилья и обследованием жилищных условий (SHC). Кроме того, фискальный регистр и регистр социального обеспечения в Нидерландах также изменились со времени переписи 2001 г. Эти источники данных были объединены и будут использоваться вместо ранее применявшегося обследования занятости и доходов (ОЗД). Мы надеемся, что этот новый комбинированный регистр может использоваться вместе с регистром пособий по безработице (РПБ) и регистром социального обеспечения (PCO) для получения большинства категорий переменной по текущему виду экономической деятельности. В дополнение к информации регистров, для переписи 2011 г. сохраняет свою значимость некоторая информация, обеспечиваемая обследованием рабочей силы (ОРС).

4. Решения в отношении того, какие источники данных использовать для получения различных переменных в рамках переписи 2011 г. основываются преимущественным образом на качестве источников, содержащих информацию по переменным. В данном документе сравнивается ряд регистров по ограниченному набору из трех переменных. Такими переменными являются: самый высокий уровень полученного образования, текущий вид экономической деятельности и информация о жилье.

5. *«Самый высокий уровень полученного образования»* является важной переменной. Информацию по этой переменной можно найти в ОРС. Тем не менее, ОРС в Нидерландах содержит информацию лишь по небольшой доле (примерно 1%) населения в возрасте от 15 до 64 лет за один календарный год. Информацию по гораздо большему числу людей можно найти в регистре образования (РО). Однако, информация в РО Нидерландов менее актуальная по сравнению с ОРС. В идеале, объединяется информация из обоих источников. Для переписи информация из одного из этих источников может быть достаточной для подготовки надежных и непротиворечивых таблиц.

6. *Текущий вид деятельности* является переменной, которая включает большое количество разных категорий таких как, например, занятые, безработные и домохозяйки. Информация о занятых лицах поступает из информации регистра. Информация о безработных в соответствии с определением Международной организацией труда (МОТ) может быть получена на основании данных обследования ОРС. В качестве альтернативы можно получить данные по безработице из регистра, содержащего данные по пособиям, а именно из РПБ и PCO. Информация в этих регистрах является полной, но не имеет точного определения безработицы, необходимого для переписи. Здесь вопрос для изучения заключается в следующем: какая информация является наилучшей для переписи 2011 г.: выборочная информация из ОРС с правильным определением или комплексная информация из регистров с приблизительным значением официального определения?

7. *Информация о жилье* может быть получена из нового РЖ. Как отмечалось выше, данный регистр ранее не использовался в целях переписи. Недостатком этого регистра является то, что в нем не достает некоторой информации. Поскольку

некоторые переменные в РЖ также имеются в других источниках (например, в земельном регистре), вопрос заключается в том, какой из источников следует использовать для получения конкретных переменных переписи.

8. Краткий обзор, представленный выше, показывает, что такие источники как РО, РПБ, РСО, РЖ и РН обеспечивают полезную информацию для получения одной из необходимых переменных. В данном документе изучается текущее состояние и качество информации об уровне образования, текущему виду деятельности и жилью, имеющейся в регистрах (и ОРС), с применением рамок качества для регистров.

### **III. Рамки качества**

9. Рамки качества для регистров были разработаны в целях стандартизации определения различных компонентов качества административных регистров (Daas et al., 2009). Рамки качества состоят из трех взглядов высокого уровня на качество. Эти три взгляда обеспечивают полный обзор компонентов качества (Daas et al., 2010). Эти взгляды считаются гиперасpekтами (Karr et al., 2006) и называются следующим образом: Источник, Метаданные и Данные. Каждый гипераспект состоит из нескольких аспектов, а каждый аспект содержит несколько показателей качества. Показатель качества измеряется или оценивается одним или более методами, которые могут быть качественными или количественными (Daas et al., 2009).

10. Статистический орган, планирующий использовать административный регистр, должен начать с изучения качества информации, которая позволяет регулярно использовать источник данных. Эти компоненты качества размещены в гипераспекте «Источник» рамок качества.

11. Второй гипераспект в рамках качества, гипераспект «Метаданные», сконцентрирован на концептуальных и связанных с процессом компонентами качества метаданных источника. Перед использованием статистическому органу важно полностью понимать компоненты качества, связанные с метаданными, поскольку неправильное понимание существенно влияет на качество результата, основанного на данных источника.

12. Для оценки показателей качества в гипераспектах «Источник» и «Метаданные» был разработан контрольный список. Он включен в работу Daas et al. (2009). Контрольный список помогает пользователю в ходе применения методов измерения к каждому из показателей качества в обоих гипераспектах. Отвечая на вопросы в контрольном списке, указывается «значение» каждого метода для каждого показателя, от «хорошо» до «плохо». Для оценки части, касающейся метаданных, необходимо, чтобы пользователь имел в виду конкретное использование, в нашем случае таковым является перепись 2011 г. Следующим шагом является определение качества данных.

13. Показатели для оценки качества данных в регистре входят в гипераспект «Данные». Здесь основное внимание показателей направлено на качество данных в регистрах как входных данных в статистическом процессе. Аспекты качества для данных касаются технических проверок, точности, полноты, своевременности и интегрируемости (Daas et al., 2011).

### **IV. Результаты проверки качества**

14. К вышеуказанным регистрам был применен контрольный список, ссылающийся на гипераспекты «Источник» и «Метаданные». После этого был сделан

первый шаг для применения показателей, соответствующих гипераспекту «Данные». В данном разделе вначале обсуждаются результаты применения контрольного списка к различным регистрам. Далее представлены предварительные выводы касательно оценки качества в гипераспекте «Данные». Основной упор в исследовании был сделан на имеющуюся в регистрах информацию об *уровне образования, текущем виде деятельности и жилье*.

## А. Источник и метаданные: применение контрольного списка

15. Контрольный список был применен к следующим регистрам: РО, РПБ, РСО, РЖ и РН. Результаты оценки, полученные в гипераспектах «Источник» и «Метаданные», представлены в Таблицах 1 и 2 соответственно.

**Таблица 1: Результаты оценки по гипераспекту «Источник»**

Аспекты	Источники данных				
	РО	РПБ	РСО	РЖ	РН
1. Поставщик	+	о	о	+	+
2. Актуальность	о	+	+	о	+
3. Неприкосновенность частной жизни и безопасность	+	+	+	+	+
4. Представление	-	+	+	+	+
5. Процедуры	о	о	о	+	+

**Таблица 2: Результаты оценки по гипераспекту «Метаданные»**

Аспекты	Источники данных				
	РО	РПБ	РСО	РЖ	РН
1. Ясность	+	+	+	+	+
2. Сопоставимость	-	о	о	+	+
3. Уникальные ключи	+	+	+	о	+
4. Обработка данных	+	+	+	+	+

15. В обеих таблицах оценки указаны на уровне аспекта. Оценки по аспектам были получены посредством выбора наиболее часто наблюдаемой оценки по каждому методу измерения в каждом аспекте. Используемые обозначения оценок: хороший (+), удовлетворительный (о), плохой (-) и неизвестный (?); промежуточные оценки создаются путем сочетания символов с косой чертой (/) в качестве разделителя.

16. Результаты в Таблице 1 показывают, что на уровне аспектов общие оценки по большинству источников данных достаточно хорошие в гипераспекте «Источник». Исключением является РО, здесь низкая оценка наблюдается по *представлению*. Это обусловлено низкой регулярностью представления (не чаще одного раза в год). РО имеет также лишь удовлетворительную (о) оценку по *актуальности*, поскольку этот источник не удовлетворяет всем потребностям переписи в информации. Этот регистр серьезно страдает от избирательной неполноты охвата (см. следующий подраздел). РПБ и РСО имеют лишь удовлетворительную оценку по *поставщику* и *процедурам*, вследствие иногда проблематичной неясной цели провайдера данных и высокого риска зависимости Статистического бюро Нидерландов. РЖ имеет удовлетворительную оценку по *актуальности*, поскольку этот источник не удовлетворяет всем потребностям в информации; в нем отсутствуют некоторые переменные (например, является ли жилье собственным или арендуемым). РН имеет только хорошие оценки.

17. Результаты в Таблице 2 показывают, что на уровне аспектов общие оценки источников данных также достаточно хорошие по большинству аспектов в гипераспекте «Метаданные». Аспекты *ясность* и *обработка данных* показывают только хорошие результаты. И вновь, РО является единственным источником данных с низкой оценкой. Данный источник данных имеет низкую оценку по *сопоставимости*, поскольку переменные промежутков сложно трансформировать в моменты времени, используемые Статистическим бюро Нидерландов. РЖ имеет лишь удовлетворительную оценку по *уникальным ключам* вследствие сложности сравнимости уникальных источников в этом источнике. Это существенно препятствует сочетанию данного источника данных с другими источниками информации. РПБ и РСО имеют удовлетворительные оценки по *сопоставимости*, что обуславливается временными различиями в отчетных периодах. Положительным исключением из всего этого вновь является РН, который имеет только хорошие оценки.

18. В целом, результаты оценки пяти источников данных показывают, что необходимо уделить внимание аспектам качества, связанным с *поставщиком*, *актуальностью*, *процедурами* и *сопоставимостью*. Результаты по РН свидетельствуют о том, что контролировать каждый аспект качества в гипераспектах «Источник» и «Метаданные» возможно. В отношении остальных источников данных можно утверждать, что результаты по одному или более аспекту качества в обоих гипераспектах требуют внимания. Было сделано заключение, что использование регистров в переписи 2011 г. не представляет больших проблем.

## **В. Данные: первые результаты оценки**

19. В данном разделе обсуждаются предварительные результаты применения показателей, относящихся к гипераспекту «Данные». В имеющемся наборе данных исходные данные были предварительно обработаны в ограниченном объеме и привязаны к РН. Все данные имели одинаковую базовую дату – 1 января 2008 г. Это предполагает, что все показатели, относящиеся к таким аспектам как: *технические проверки*, *сроки* и *интегрируемость* не рассматриваются в данном документе. Анализ направлен на рассмотрение аспектов *полноты* и *точности*.

### **1. Аспект полноты**

20. Анализ по полноте будет сосредоточен на следующих переменных: *уровень образования* (получаемый из РО) и *текущий вид экономической деятельности* (получаемый из РПБ, РСО и ОРС). Переменные, относящиеся к жилью, будут доступны позднее, в связи с чем не обсуждаются здесь.

21. Для того, чтобы получить первое впечатление об уровне *неполного охвата* информации, имеющейся по этим переменным, мы предполагаем, что население состоит из всех лиц в РН, мы рассматриваем по скольким строкам отсутствует значение. Конечно, это может ввести в заблуждение, если рассматриваемые переменные применимы только к части населения. Однако, это не касается переменных, рассматриваемых здесь, поскольку категории этих переменных таковы, что каждый человек принадлежит к какой-либо категории. В Таблице 3 представлен обзор количества отсутствующих значений по обеим изученным переменным.

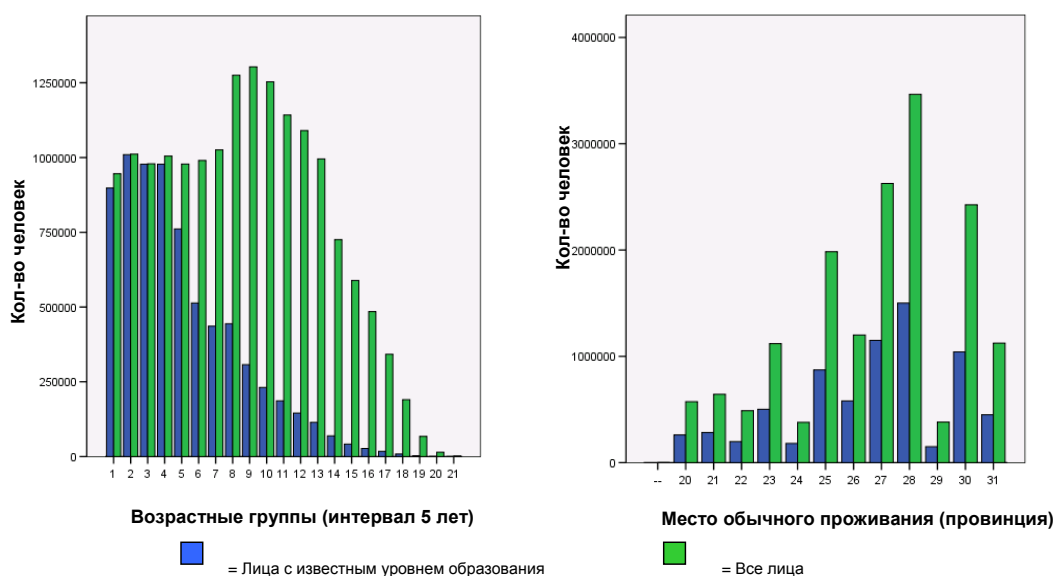
**Таблица 3: Обзор отсутствующих значений**

<i>Переменная</i>	<i>Кол-во отсут. значений</i>	<i>Доля отсут. значений (%)</i>
Уровень образования	9 238 212	56,3
Текущий вид экон. деят-ти	2 140 266	13,0

22. Неполный охват по такой переменной как *текущий вид экономической деятельности* неудивителен, поскольку по трем категориям этой переменной (т.е. безработные, домохозяйки и другие) в наших данных имеется информация только из ОРС (выборочного обследования). Еще больший уровень неполного охвата наблюдается по переменной «*уровень образования*». Это также не было неожиданностью, поскольку регистры, используемые для получения этой переменной, содержат в основном сведения о молодежи. Это позволяет предположить, что неполный охват очень избирателен в отношении возраста. Избирательность сведений, имеющихся по *уровню образования*, далее изучается на Рис. 1. Синие столбцы в гистограммах относятся к части населения, по которым имеется известное значение переменной по *уровню образования*. Зеленые столбцы относятся к населению в целом. На диаграмме слева люди сгруппированы по возрастам (с интервалом 5 лет), в то время как на гистограмме справа люди сгруппированы по такой переменной как «место обычного проживания» (провинции).

23. Графики четко показывают, что имеющиеся сведения касательно уровня образования очень избирательны с учетом возраста, т.е. гораздо больше сведений имеется по более младшим возрастным группам. С другой стороны, доли населения, проживающего в 12 провинциях, демонстрируют большое сходство между населением в целом и частью населения, по которым имеются сведения об *уровне образования*. Это предполагает, что сведения неизбирательны с учетом места проживания.

**Рисунок 1: Уровень образования по возрастным группам – выявлена неполнота охвата по некоторым возрастным группам.**



24. Другим показателем, связанным с полнотой, является избыточность, т.е. наличие нескольких записей объектов. Для изучения того, имеется ли в данных избыточность, мы искали строки в записях, которые показывали одинаковые значения по всем переменным (включая уровень образования и текущий вид

деятельности) за исключением личного идентификатора. В наших тестовых данных обнаружилось 67 644 «дублирующих записей», что составляет 0,4% всех данных. Дальнейший анализ дублирующих записей показал, что большинство дублирующих записей соответствуют лицам, живущим в институциональных учреждениях. Например, лица, живущие в домах для престарелых, имеют один и тот же адрес, относятся к одной возрастной категории и т.д. Учитывая возможность того, что люди в институциональных учреждениях, действительно имеют одни и те же значения по ограниченному набору переменных, имеющихся в нашей тестовой базе данных, мы решили сконцентрировать свое внимание в будущем исследовании на выборочную часть дублирующих записей, которые не относятся к людям, живущим в институциональных учреждениях.

## 2. Аспект точности

25. В части точности в данном разделе мы рассматриваем наличие неоднозначных (двусмысленных) значений в данных. Вновь основное внимание было уделено таким переменным как *уровень образования* и *текущий вид экономической деятельности*. По этим переменным особый интерес представляет связь с возрастом. Для наглядности, для получения высших уровней образования необходимо время. Поэтому, например, человек в возрасте 18 лет (обычно) еще не имеет степени доктора наук. Кроме того, ожидается, что (практически) только более взрослые люди будут иметь значение равное 3 (получатели пенсий или дохода от капитала) по *текущему виду экономической деятельности*.

26. Для того, чтобы проанализировать сохранение этих ожидаемых связей между рассматриваемыми переменными и возрастом, были созданы перекрестные таблицы. Результаты представлены в Таблицах 4 и 5. При толковании этих результатов следует учитывать то, что в особенности по переменной «уровень образования» многие значения отсутствуют и что количество отсутствующих значений зависит от возраста (см. первый столбец Таблицы 4). Вследствие этого, не так много выводов можно сделать, например, из итогов по какой-либо ячейке таблицы. Несмотря на это, в отношении переменной «уровень образования» правомерно сделать заключение, что для самой молодой части населения Нидерландов значение переменной отсутствует или оно имеет значение, равное «не применимо». Самое молодое население, которое достигло шестого уровня образования (Второй уровень третичного образования), находится в возрастной группе 20-25 лет. Кроме того, имеются некоторые лица, достигшие пятого уровня образования (первый этап третичного образования) уже в возрастной категории 15-20 лет. Однако, большинство людей достигает этого уровня в более старшем возрасте. Можно с предосторожностью сделать вывод, что большинство молодых людей продолжают обучение после достижения первого уровня. Это соответствует ожиданиям, поскольку в Нидерландах дети обязаны ходить в школу до 16 лет.



Таблица 4: Перекрестные таблицы по переменной «Уровень образования в

Ageclass	Educational Attainment								
	Missing	0	1	2	3	4	5	6	9
1: [0, 5)	47674	0	0	0	0	0	0	0	898187
2: [5, 10)	1414	0	0	0	0	0	0	0	1009745
3: [10, 15)	1275	0	0	0	0	0	0	0	977689
4: [15, 20)	27192	147	258459	539275	163161	43	16684	0	0
5: [20, 25)	216918	1830	14448	126841	405943	2987	209128	4	0
6: [25, 30)	476523	3800	10723	36161	147537	2779	312380	79	0
7: [30, 35)	589228	4661	11692	22976	104633	3041	288873	217	0
8: [35, 40)	830625	5136	12811	27478	112666	5832	280059	399	0
9: [40, 45)	996017	4888	12890	31923	100996	7906	148114	485	0
10: [45, 50)	1022110	4596	13128	33829	79051	8578	91443	502	0
11: [50, 55)	955668	4245	14695	32938	58110	7787	68153	409	0
12: [55, 60)	944786	4018	16745	30682	41449	6790	45224	392	0
13: [60, 65)	881054	3536	16656	28929	30744	5720	28315	335	0
14: [65, 70)	656135	2962	11408	19367	18166	3353	13914	190	0
15: [70, 75)	547283	2059	8049	12352	10552	1818	6867	101	0
16: [75, 80)	457713	1254	5361	8994	6520	1094	3900	51	0
17: [80, 85)	324613	493	3356	6650	3922	617	2321	35	0
18: [85, 90)	181535	201	1911	3531	1811	262	909	13	0
19: [90, 95)	64749	85	852	888	496	61	268	7	0
20: [95, 100)	14174	22	201	157	78	10	40	2	0
21: [100, ∞)	1526	2	7	16	8	1	5	0	0

## сравнении с возрастными группами

Примечание: Уровни образования: (0) Отсутствие официального образования, (1) Уровень 1 МСКО. Начальное образование, (2) Уровень 2 МСКО. Первый этап среднего образования, (3) Уровень 3 МСКО. Второй этап среднего образования, (4) Уровень МСКО 4. Послесреднее нетретичное образование, (5) Уровень МСКО 5. Первый этап третичного образования, (6) Уровень МСКО 6. Второй этап третичного образования, (9) Не применимо (лица младше 15 лет).

Таблица 5: Перекрестная таблица по переменной «Текущий вид экономической

Ageclass	Current activity status							
	Missing	0	1	2	3	4	5	6
1: [0, 5)	0	945861	0	0	0	0	0	0
2: [5, 10)	0	1011159	0	0	0	0	0	0
3: [10, 15)	0	978964	0	0	0	0	0	0
4: [15, 20)	34911	0	482180	33	0	487533	11	293
5: [20, 25)	113286	0	716411	106	0	147395	190	711
6: [25, 30)	142149	0	818167	107	0	28396	486	677
7: [30, 35)	163141	0	856030	129	0	4506	744	771
8: [35, 40)	216807	0	1053407	180	0	2418	1138	1056
9: [40, 45)	228634	0	1070204	228	0	1853	1076	1224
10: [45, 50)	236102	0	1013249	242	0	1134	1076	1434
11: [50, 55)	262473	0	875724	253	1	504	1261	1789
12: [55, 60)	330898	0	714959	263	39705	232	1776	2253
13: [60, 65)	390062	0	343089	122	256826	78	2348	2764
14: [65, 70)	8730	0	88209	1	628490	16	3	46
15: [70, 75)	5306	0	35690	1	548059	3	0	22
16: [75, 80)	3822	0	14705	0	466339	2	0	19
17: [80, 85)	2166	0	5897	0	333936	0	0	8
18: [85, 90)	1115	0	2360	0	186690	0	0	8
19: [90, 95)	405	0	662	0	66339	0	0	0
20: [95, 100)	162	0	136	0	14386	0	0	0
21: [100, ∞)	97	0	18	0	1450	0	0	0

## деятельности» в сравнении с возрастными группами

Примечание: Текущий вид экономической деятельности: (0). Лица младше минимального возраста для экономической деятельности, (1) Занятые, (2) Безработные, (3) Получатели пенсии или дохода с капитала, (4) Экономически неактивные студенты, (5) Домохозяйки, (6) Другие.

27. В Таблице 5 число безработных, домохозяек и других получены из ОРС, что означает, что по этим категориям имеются только выборочные сведения. Результаты, показанные по этим категориям, не взвешены к общей численности населения. Таблица 5 соответствует тому, что пенсионный возраст в Нидерландах обычно составляет 65 лет, т.е. в возрастной группе 60-69 лет наблюдается очевидный пик записей со значением 3 (получатели пенсий или дохода от капитала) по текущему виду экономической деятельности. Можно также увидеть, что доля населения со

значением 1 (занятые) существенно сокращается по достижению ими возраста 65 лет, хотя, похоже, что есть некоторые лица старше 100 лет, которые все еще работают. Значение 4 (экономически неактивные студенты) также соответствует ожиданиям, поскольку этот вид экономической деятельности наблюдается в основном среди населения младше 25 лет.

28. На основании этих таблиц можно с предосторожностью заключить, что значения переменных по рассматриваемым переменным являются точными. Хотя, конечно, по этому вопросу требуется гораздо больший анализ.

## V. Выводы

29. 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. На сегодня в Нидерландах было проведено три виртуальных переписей. 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. Также в отношении виртуальной переписи 2011 г. важно, что окончательные результаты сопоставимы как с течением времени, так и с другими странами. В связи с этим качество регистров, используемых в Нидерландах, имеют критическую важность для переписи 2011 г.

30. Перепись на основе регистров можно проводить в гораздо большем числе стран. Нужно отметить, что в большинстве стран не все переменные переписи могут быть получены из сведений, имеющихся в регистрах. Для этих переменных сохраняется необходимость в дополнительных обследованиях. Для того, чтобы можно было использовать регистры в статистических целях, должна быть возможность определить качество этих регистров. Результаты, представленные в этом документе, показывают, что рамки качества, разработанные для административных регистров, и соответствующий контрольный список являются ценными инструментами для оценки статистической пригодности таких источников данных. В предстоящие годы будет принято решение о том, как будут получаться разные переменные переписи в Нидерландах. В течение этого периода к данным в используемых регистрах будет применяться больше показателей из гиперспекта «Данные».

31. Большим преимуществом подхода, применяемого для построения массива данных виртуальной переписи (Schulte Nordholt, 2004), является использование микроинтеграции. В процессе микроинтеграции данные проверяются, и некорректные данные адаптируются. Это позволяет сократить число ошибок измерений. Внедрение методики повторного взвешивания решает проблему остальных несоответствий. С учетом детальной запрашиваемой информации в рамках переписи 2011 г., имеющихся источников для переписи в Нидерландах и наш первый опыт по применению рамок качества, несомненно, у нас будет очень интересный опыт с переписью на основе регистров 2011 г. в предстоящие годы, который привлечет внимание многих других стран.

## Библиография

Daas, P.J.H., S.J.L. Ossen, R.J.W.M. Vis-Visschers and J. Arends-Toth, 2009. Checklist for the Quality evaluation of Administrative Data Sources. Discussion paper 09042, Statistics Netherlands, The Hague / Heerlen. <http://www.cbs.nl/NR/rdonlyres/0DBC2574-CDAE-4A6D-A68A-88458CF05FB2/0/200942x10pub.pdf>

Daas, P.J.H., S.J.L. Ossen and M. Tennekes, 2010. The determination of administrative data quality: recent results and new developments. Paper for the European Conference on Quality in Official Statistics 2010, Helsinki, Finland.

Daas, P., S. Ossen, M. Tennekes, L.C. Zhang, C. Hendriks, K. Foldal Haugen, F. Cerroni, G. Di Bella, T. Laitila, A. Wallgren and B. Wallgren, 2011. Report on methods preferred for the quality indicators of administrative data sources. Second deliverable of workpackage 4 of the BLUE Enterprise and Trade Statistics project, 28 September 2011.

Karr, A.F., A.P. Sanil and D.L. Banks, 2006. Data quality: A statistical perspective. *Statistical Methodology*, Number 3, pp. 137-173.

Schulte Nordholt, E., 2004. Introduction to the Dutch Virtual Census of 2001. *The Dutch Virtual Census of 2001, analysis and methodology*, eds. E. Schulte Nordholt, M. Hartgers and R. Gircour, Statistics Netherlands, Voorburg, pp. 9-22.

---