



## ЭКОНОМИЧЕСКИЙ И СОЦИАЛЬНЫЙ СОВЕТ

Распространение: Общее  
30 апреля 2012 г.  
Оригинал: английский

### Европейская экономическая комиссия ООН

Конференция европейских статистиков

60-е пленарное заседание

Париж, 6-8 июня 2012

2010 раунд переписи – инновации и уроки

### Уроки проведения комбинированной переписи для будущего развития социальной статистики

#### Записка Федерального статистического управления Германии

##### *Резюме*

В данном документе описан опыт, приобретенный при проведении Переписи 2011 в Германии, где были использованы новые методы сбора данных. Этот новый метод состоит в комбинации административных регистров и полевых обследований. Самой сложной проблемой нового дизайна переписи явилось то, что при объединении данных из разных источников на уровне отдельных лиц необходимо частично использовать текстовую информацию, включающую имя и адрес, поскольку не существует ни единых персональных идентификационных номеров, ни идентификационных номеров для зданий. Опыт по объединению данных без числовых идентификаторов при проведении Переписи 2011 может быть использован для разработки предложений по дальнейшему развитию нового дизайна переписи. Этот опыт может также быть полезен другим странам, которые еще не начали использовать данные из административных регистров для целей переписи в связи с похожими трудностями в отношении комбинирования данных из административных регистров и данных полевых обследований.

## I. Правовые и политические требования при разработке новой модели переписи в Германии

1. Последняя перепись населения и жилого фонда в Германии до 2011 года была проведена в 1987 году как полное обследование всей страны, когда все жители были опрошены интервьюерами. Эта перепись вызвала массовые недовольства в некоторых группах населения, включая успешную жалобу на нарушение конституции, что привело к запрету на передачу данных обратно в администрации. В соответствии с этим запретом, который действует и поныне, данные, которые были предоставлены Федеральной и региональным статистическим службам для производства статистики, не должны были направляться обратно в администрации. Также в соответствии со своим решением Федеральный конституционный суд обязал законодателя рассмотреть специальный вопрос при принятии решений о методах сбора данных в ходе будущих переписей населения. Этот вопрос таков: в свете дальнейшего развития методов официальной статистики и социальных исследований, а также методов обработки данных, является ли все еще необходимым проведение традиционной полной переписи, или же информация, необходимая для получения результатов, может быть получена с использованием других методов сбора данных, которые позволят снизить нагрузку на респондентов?

2. Таким образом, после 1987 года законодатель обязал статистические органы разработать новые методы проведения переписи. Следующие общие требования должны были быть приняты во внимание: где возможно, следует использовать существующие административные данные для целей переписи, и затраты должны быть существенно ниже, чем при проведении переписи традиционным методом, при этом качество основных результатов переписи, в частности, показателей численности населения, должно быть не хуже, чем в предыдущих переписях.

3. В результате был разработан дизайн переписи, который предполагал использование регистров в сочетании с полевыми обследованиями. Первым шагом в выполнении этой цели явилось создание регистра адресов и зданий. Он содержит все адреса в Германии, где имеются жилые помещения. Этот регистр служит, во-первых, в качестве статистической совокупности для переписи зданий и жилого фонда, и, во-вторых, - в качестве основы выборки для обследования домашних хозяйств. Кроме того, регистр адресов и зданий используется в качестве основного организационного компонента для всех обследований Переписи 2011. Данные Переписи 2011 получают из следующих компонентов (Схема 1):

(а) Данные управлений по регистрации населения являются наиболее важными источниками данных для определения численности населения и демографической структуры населения, проживающего в частных домохозяйствах. Они объединяются в общий массив данных для Германии и проходят проверку на наличие неправильных и дублирующих записей<sup>1</sup>;

(б) Для жителей коллективных жилых помещений, институциональных организаций, жилых заведений и других подобных объектов, данные регистра населения содержат слишком много ошибок, так что данные на этих людей собираются с помощью полной переписи.

(с) На уровне отдельных лиц данные регистра Федерального агентства занятости и данные о государственных служащих использовались для пополнения демографических переменных, полученных из регистров населения и специальных обследований;

(d) Проведение выборочного обследования менее чем 10% населения обеспечивает качество данных регистра путем определения для каждого муниципального образования процента устаревших и пропущенных записей и учета их при определении численности населения по муниципальным образованиям<sup>2</sup>. В ходе обследования также собирают дополнительные данные, которые невозможно получить из регистров<sup>3</sup>;

(е) Данные о количестве и структуре жилых домов и жилых помещений собирают посредством проведения почтового обследования всех собственников;

(f) Наконец, путем объединения персональных данных проводят процедуру формирования данных по домохозяйствам путем привязки данных из регистров населения, которые включают информацию о семейных отношениях, к данным о жилых помещениях из переписи зданий и жилого фонда.

4. Методы проведения переписи обеспечивают полный охват целевой совокупности<sup>4</sup>

---

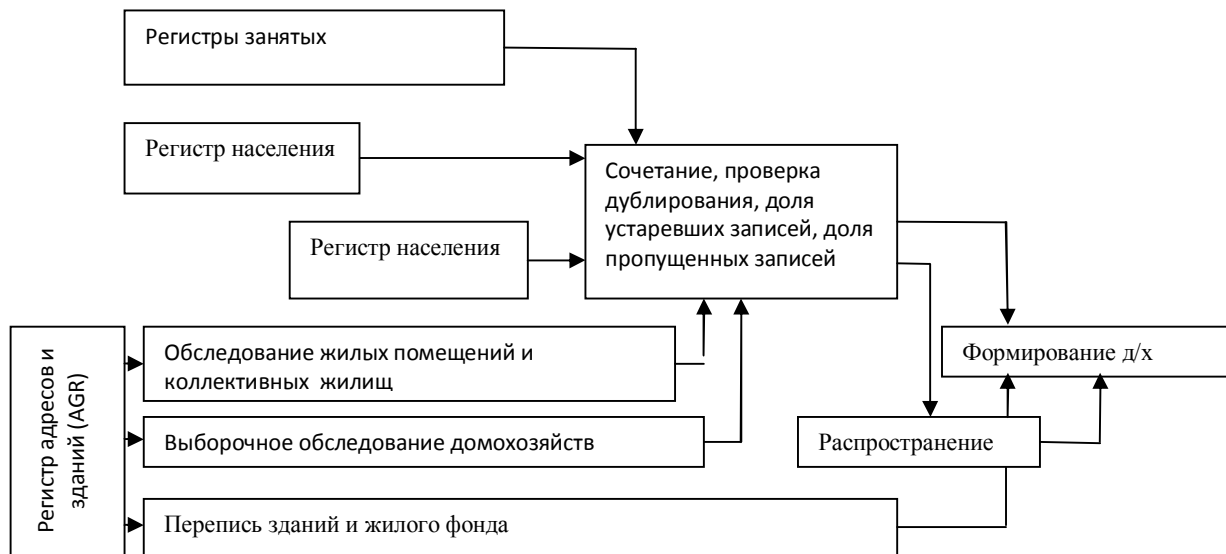
<sup>1</sup> Случаи дублирования могут касаться лиц, которые зарегистрировали основное место жительства в нескольких муниципалитетах

<sup>2</sup> Метод обеспечения качества данных регистра населения применяется в муниципальных образованиях с населением 10 000 человек и выше. В муниципальных образованиях с населением меньше 10 000 человек, контроль качества осуществляется путем проведения обследования для выявления расхождений и уточнения места проживания.

<sup>3</sup> Дизайн выборки и модель для обследования домашних хозяйств были разработаны в рамках специального проекта Университета Трира и GESIS - Института Социальных Наук им. Лейбница и в Мангейме.

<sup>4</sup> Охвачены жители в основном или втором месте жительства, но официальный показатель численности рассчитывается только на основе учета жителей по основному или единственному месту жительства.

Схема 1  
Дизайн переписи



## II. Приемлемость нового метода проведения переписи

### A. Одобрение населением

5. По сравнению с тем стрессом, который испытали политики и статистики в результате открытого неприятия последней переписи населения 1987 года в Западной Германии большими группами общества, перепись этого года протекала, фактически, при отсутствии какой-либо оппозиции. Это явилось не только результатом изменения юридических и общественных условий (защита данных была значительно усилена после переписи населения 1987 г., случаи с нецелевым использованием персональных данных пока не происходили в официальной статистике, да и многие люди в настоящее время довольно открыты в отношении своих персональных данных), но и результатом использования нового метода проведения переписи. Поскольку были использованы данные регистра населения, только десятая часть населения бала непосредственно затронута обследованием домашних хозяйств.

6. Почтовое обследование зданий и жилого фонда было ограничено группой населения с более слабым оппозиционным настроем (собственники жилья). Количество запросов от собственников недвижимости, которые выражали беспокойство, потому что не получили вопросник вовремя, было гораздо больше, чем количество тех, кто не выполнил своего обязательства по участию в обследовании.

### B. Значительное недоверие со стороны местных органов

7. Среди местных администраций наблюдается явное неприятие результатов переписи, которые еще даже не были опубликованы. Одна из причин этого состоит просто в том, что понимание основных принципов нового метода проведения переписи требует знаний в области теории выборочных обследований, тогда как люди способны понять основные принципы традиционной переписи и регистравого учета, не обладая специальными знаниями. Понимание методов выборки и распространения результатов, применяемых в Переписи 2011, потребует хорошего знания методологии выборочных обследований.

8. Поскольку официальный показатель численности населения является одним из важных факторов в системе «уравнивания доходов», то этот показатель является основой для определения финансовых ресурсов региональных и местных властей. Поэтому некоторые муниципалитеты уже сейчас готовятся к юридическим действиям, чтобы опротестовать официальные показатели численности населения, которые будут получены в ходе Переписи 2011. Основная критика направлена на то, что показатель численности населения, вместо того, чтобы быть полученным в ходе поголовного учета, основан на оценке, которая включает ошибку выборки. В ходе выборочного обследования на уровне муниципалитетов определяется доля неактуальных (избыточный охват) или пропущенных (недостаточный охват) записей в данных регистра населения. Эта доля затем принимается во внимание при определении официального показателя численности населения. Уровень ошибки выборки в официальном показателе численности населения не будет известен до тех пор, пока перепись не будет завершена. Однако в Законе о переписи сказано, что дизайн переписи должен быть таков, чтобы относительная стандартная ошибка результатов Переписи не превышала 0,5%.

### **III. Объединение данных из нескольких источников без персональной идентификации**

9. Несмотря на то, что сбор и обработка данных для Переписи 2011 из разных источников еще не закончены, уже возникают проблемы, которые нужно будет преодолеть при объединении данных на уровне адресов и отдельных лиц. В Германии отсутствуют персональные идентификационные номера (ID) и, принимая во внимание законодательство о защите данных, в настоящее время трудно себе представить, что даже в долгосрочной перспективе могут быть введены стандартные персональные ID. Опыт, приобретенный в ходе Переписи 2011, показывает, однако, что такое положение не должно препятствовать проведению переписи, когда данные из традиционных статистических обследований объединяются с данными административных регистров. Объединение данных достигается с использованием ряда переменных, которые призваны идентифицировать человека (среди прочих, имя при рождении, фамилия, имя, дата рождения, место рождения и адрес). Однако для этого требуется гораздо больше усилий, чем для объединения данных при помощи цифровых идентификаторов.

10. Центральным компонентом Переписи 2011 в Германии является регистр адресов и зданий (*Anschriften- und Gebäuderegister – AGR*), который содержит все адреса зданий с жилыми помещениями в Германии (схема 1). Это временный статистический регистр, который был сформирован специально для целей Переписи 2011 и может быть использован только для проведения этой переписи. В соответствии с законодательством, регистр и все данные должны быть уничтожены самое позднее через шесть лет после даты переписи. При составлении AGR, была использована адреса с геопривязкой, полученные от землеустроительных органов, данные из регистра населения и регистра Федеральной службы занятости. Регистры населения поддерживаются децентрализованно около 12 500 муниципалитетов в примерно 5 400 вычислительных центрах. Построение регистра потребовало произвести объединение источников данных путем сопоставления текста (названий улиц).

11. Регистр AGR описывает совокупность зданий и жилой фонд. Он предоставляет информацию, необходимую для распределения вопросников (имя и адрес собственника здания), а также основу выборки для обследования домохозяйств. Кроме того, он используется для сбора и обработки данных для объединения данных из разных источников на уровне отдельных лиц. Обычно это объединение представляет собой двухэтапный процесс. На первом этапе записи с персональными данными увязываются с соответствующим адресом (используя такие переменные, как почтовый индекс, официальный код муниципального образования, суб-муниципальное образование, название улицы, номер дома с дополнениями, если есть), и только на втором этапе происходит сопоставление лиц с заданным адресом.

12. При объединении данных для формирования AGR, имеется три слабые точки. Во-первых, переименования улиц (например, в случае изменения территориального статуса), во-вторых, разное написание названий улиц (в трех источниках данных и даже в рамках одного источника), и, в-третьих, разное использование дополнительных символы в номерах домов. В большинстве случаев разное написание названий улиц было уточнено с использованием компьютерной программы, например, путем стандартизации сокращений или удаления специальных знаков и чисел. Однако оставалось много случаев, в которых нужно было разбираться вручную. Для приблизительно 1,2 млн. улиц в Германии существовало около 5,1 млн. вариантов написания в трех источниках данных, использовавшихся для построения AGR. Используя полученные знания о различном написании названий улиц, был составлен тезаурус улиц, который использовался при обновлении AGR при поступлении новых данных. Это было необходимо, потому что новые записи содержали старые, нестандартизованные названия улиц, и не

разрешалось передавать стандартизованные написания обратно в агентства, поддерживающие регистры. Таким образом, когда новые записи включались в AGR, в большинстве случаев не было необходимости ни проводить автоматическую стандартизацию, ни – особенно – трудоемкую ручную проверку.

#### **IV. Дальнейшее развитие нового метода проведения переписи**

##### **A. Использование стандартных идентификаторов адресов в административных регистрах**

13. Объединение разных источников данных можно было бы проводить гораздо более эффективно при проведении комбинированных переписей в будущем, если бы переменные адреса были бы введены в административные регистры, используемые для переписи. Очевидный код, который можно здесь использовать – это географические координаты. В качестве отправной точки для разработки такого списка адресов может быть использован тезаурус улиц. Такой тезаурус был разработан для Переписи 2011. Однако прежде необходимо решить ряд вопросов. В частности, существует проблема лицензий, а также вопрос о том, как можно эффективно организовать процесс актуализации тезауруса улиц (в основном, добавления новых названий улиц). Официальный список адресов может быть увязан со списком муниципалитетов, который поддерживается федеральной и региональными статистическими службами. Фактически, это означает, что статистические службы будут предлагать бесплатно списки адресов на хорошо понятных либеральных лицензионных условиях при помощи современных интерфейсов программирования приложений (APIs). В свою очередь это обеспечит то, что стандартизованный набор адресов будет иметься в распоряжении статистической службы при обращении к административным данным.

##### **B. Отказ от обновления результатов переписи**

14. Для получения текущих данных о демографической структуре населения Германии результаты последней переписи обновляются на детальном региональном уровне (в разбивке до муниципального уровня) и по демографическим переменным (возраст, пол, семейное положение и гражданство). Это делается с использованием данных о естественном движении населения (рождения и смерти) и статистики прибытия и убытия (только движение между муниципальными образованиями). Такое обновление данных по численности населения приводит к ошибкам, особенно, если проводится в течение долгого времени.

15. Поэтому следует попытаться в среднесрочной перспективе заменить обновление данных по населению в межпереписной период на регулярные оценки (как минимум, раз в год) регистров населения федеральной и региональными статистическими службами. Для этого необходимо, чтобы были выполнены, по крайней мере, два следующих требования:

- (a) Избыточный охват и недостаточный охват регистров населения по оценке Переписи 2011 должны находиться в приемлемых пределах;
- (b) Значительное улучшение качества сетевого регистра может быть достигнуто путем регулярных (как минимум один раз в год) проверок и корректировок регистров, не объединенных в сеть, на наличие дублирующих записей.

16. При определении содержания регистров населения следует лучше учитывать требования официальной статистики. Так информация, которая должна присутствовать в регистрах населения, такова: для всех лиц – их бывшие места проживания в течение последних пяти лет; для натурализованных лиц – их бывшее гражданство. Такая информация необходима для разграничения целевой совокупности переписи в соответствии с международными стандартными определениями и для учета миграции.

##### **C. Построение постоянного регистра зданий и жилого фонда**

17. На перепись зданий и жилого фонда приходится половина всех затрат на проведение Переписи 2011 в Германии. При проведении в Германии переписи на основе регистра необходимо проводить это крупное первичное статистическое обследование, так как в Германии отсутствует регистр зданий и жилого фонда для всей страны. Кроме того, при помощи переписи зданий и жилого фонда в Германии следует получить вспомогательные переменные, которые могут быть использованы для формирования данных по домохозяйствам и объединения их с данными по жилым помещениям. Для этой цели собирают не только данные о зданиях и жилом фонде, но также сведения об одном-двух арендаторах/владельцах

каждого жилого помещения.

18. Перепись зданий и жилого фонда 2011 проводилась в форме почтового обследования собственников жилых зданий (в Германии почти 20 миллионов зданий с жилыми помещениями). С этой целью был создан регистр адресов, содержащий имя и адрес собственников для всех адресов. Данные о собственниках жилых помещений должны были проверяться в административных регистрах, например, при помощи данных муниципалитетов об уплате налога на недвижимость или административных данных предприятий по предоставлению коммунальных услуг и управлению отходами. Эти исследования оказались очень трудоемкими. Кроме того, качество данных о собственниках может сильно различаться по регионам. Плохое качество части данных и, в особенности, недостаточная актуальность данных о собственниках, не только служили причинами огромных усилий по сбору данных, но также приводили к значительным проблемам в части принятия их респондентами.

19. Статистические службы Германии хотели бы распространить успешный опыт использования регистра населения и регистра предприятий для включения других областей. В связи с этим предпринимаются шаги для достижения такого положения, когда данные о зданиях, жилых помещениях и домохозяйствах могут быть получены из регистров, что будет означать быстрый и дешевый сбор данных для переписи. Регистр зданий и жилого фонда, который еще только должен быть построен, позволит сделать это.

20. Такой регистр должен будет содержать полные и актуальные данные о зданиях и жилых помещениях и соответствующую базовую информацию, необходимую для выполнения требований регламента ЕС о переписях (год постройки, размер жилого помещения, структура собственности, тип отопления).

21. Построение такого регистра в первый раз может быть поручено федеральной и региональным статистическим службам Германии, так как у них имеется большой опыт построения регистра адресов и зданий для Переписи 2011. Более того, без нового регистра зданий и жилого фонда, статистическим органам придется опять через несколько лет создавать регистр адресов и зданий для Переписи 2021. Такой регистр будет полезен не только для органов официальной статистики, но и для государственной администрации. Поэтому его не следует вести как статистический регистр, потому что в этом случае не будет разрешено передавать данные обратно в администрацию, его скорее нужно вести вне защищенной сферы официальной статистики.

22. Существование такого регистра заметно снизит объем работы для статистических служб каждые десять лет, когда должны проводиться переписи. Кроме того, он будет полезен для других областей официальной статистики. Например, обновление статистики по строительству сможет достичь качества и оперативности, которые пока не были доступны. Также можно будет отказаться от дополнительного обследования жилищной ситуации, которое проводится как часть микропереписи, потеряв при этом только несколько переменных, таких как арендная плата и источник энергии для отопления; это позволит сократить количество первичных статистических обследований и проводить только те, которые абсолютно необходимы.