

Распространение: Общее  
18 апреля 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 3: Наличие, полнота и качество данных регистров и других источников**

### **Регистры населения в Германии – основной источник данных в переписи 2011 года**

#### **Записка Федерального статистического управления Германии<sup>1</sup>**

## **I. Модель переписи в Германии – краткий обзор**

1. В отличие от последней переписи населения перепись 2011 г. проводилась по новому методу. Для сокращения нагрузки на население по мере возможности использовались административные регистры. Это позволило провести оценку всех данных, содержащихся в регистрах, без необходимости напрямую связываться с гражданами, составляющими совокупность переписи. Тем не менее, информация, обеспечиваемая регистрами по переменным, охватывает всю Германию и полностью доступна для дальнейшего использования. Это позволит анализировать результаты переписи без необходимости проведения полного подсчета населения. Поэтому используются имеющиеся административные регистры и будут добавлены дополнительные компоненты обследования для того, чтобы обеспечить данные по всем переменным переписи. В результате, перепись 2011 г. затронула не всех жителей. Прогресс, достигнутый в информационных технологиях, позволяет осуществить методологический переход от традиционной переписи населения к новой модели переписи. Помимо такого преимущества как снижение нагрузки на население, регистры обеспечивают территориальные переменные, позволяя проводить соответствующий анализ и комбинацию результатов.

2. Перепись на основе регистров состоит из трех основных компонентов, обеспечивая сведения о населении, а также о зданиях и жилищном фонде в

---

<sup>1</sup> Автор - Stephanie Hirner.

Германии. Этими компонентами являются административные регистры, полная перепись зданий и жилищного фонда и выборочное обследование домохозяйств в конкретных районах. Другие переменные, которые не могут быть охвачены напрямую переписью на основе регистров, будут получены посредством статистических методов, применяемых в процедуре формирования домохозяйств. Процесс формирования сочетает информацию по всем компонентам переписи и составляет новые переменные, такие как тип домохозяйства или информация о структуре семьи.

## **II. Использование регистров населения в переписи 2011 г.**

3. В переписи 2011 г. регистры населения выполняют три основные функции.
4. С одной стороны, регистры населения использовались для построения регистра адресов и зданий, который является ведущим блоком данных для переписи 2011 г. в Германии и основой выборки для выборочного обследования домохозяйств. Для этого адреса зарегистрированного населения используются для создания полного регистра путем слияния с другими административными регистрами, такими как регистры занятости и регистры органов геодезии и картографии. Детальную информацию об этом регистре адресов можно найти в другом вспомогательном документе к данному заседанию «Регистр адресов и зданий – комбинация различных регистров».
5. С другой стороны, регистры населения использовались счетчиками для обследования. Для этого, а также для планирования и организации обследования регистры населения были собраны в 2010 г., т.е. за 6 месяцев до базовой даты переписи. Имея эту информацию, можно было планировать статистические территории и готовить организационные документы. Кроме того, сведения из регистров можно было использовать в качестве указателя по количеству проживающих, которых должны были опросить счетчики.
6. Наиболее важной функцией регистров населения в переписи 2011 г. является то, что они являются основным источником данных, охватывающими демографические и географические переменные переписи. Регистры населения имеются в Германии по территориям; лица, проживающие в Германии, обязаны зарегистрироваться по своему адресу проживания. Вследствие этого данные управлений по регистрации граждан являются основой для определения численности населения. Далее в настоящем документе основное внимание будет уделено использованию регистров населения в качестве основного источника данных для определения численности населения.

## **III. Регистры населения: характеристики и сбор данных**

### **A. Характеристики**

7. В настоящее время технический прогресс позволяет эффективно использовать данные, которые уже хранятся в административных регистрах. Наиболее важными административными регистрами, которые используются в переписи, являются регистры населения. Они являются основными источниками данных для получения демографических и географических переменных переписи. Они содержат сведения о каждом человеке, зарегистрированном как проживающий в Федеративной Республике Германия. Регистры населения ведутся управлениями регистрации населения, которые законодательно обязаны передать имеющиеся у них

совокупности данных в органы статистики в целях переписи. Поскольку управления регистрации населения являются частью государственного аппарата, наличие и полнота регистров гарантируется на очень высоком уровне. В соответствии с законом о регистрации населения в этих регистрах хранятся сведения о зарегистрированном местожительстве с указанием является ли данное местожительство единственным, основным или второстепенным местожительством, и в целях переписи на основе регистров соответствует обязательной переменной ЕС «место обычного проживания». Поэтому данные управлений регистрации граждан являются наиболее важными источниками данных для определения численности населения Германии. Кроме того, этот административный регистр содержит основные демографические данные по каждому человеку (пол, возраст, семейное положение, гражданство, место рождения).

8. Вследствие требования по регистрации в управлении регистрации населения все муниципалитеты в Германии ведут регистры населения, которые содержат сравнимые сведения. Регистры населения ведутся децентрализованно, поэтому каждый муниципалитет использует собственный регистр. В целом, основная структура регистров одинаковая, поскольку она основывается на единых параметрах совокупности данных.

9. Для оценки данных регистров населения, децентрализованные совокупности данных должны быть объединены в один массив данных. Согласно закону о регистрации для граждан, имеющих одно основное место проживания и одно или несколько второстепенных мест проживания, может быть зарегистрировано несколько адресов проживания. Но может так произойти, что лица зарегистрированы одновременно в нескольких муниципалитетах с единственным местожительством или основным местожительством или только с второстепенным местожительством. Такие ошибки могут произойти, когда регистрация и снятие с регистрации не производятся одновременно, или могут быть вызваны задержками в административной обработке, непредставлением человеком заявления на снятие с регистрации и т.д. После объединения регистров населения в один источник данных в целях переписи 2011 г. в первую очередь необходимо было выявить и скорректировать такие ошибки.

## **В. Сбор данных**

10. На основании Закона о переписи 2011 г. сбор данных регистров населения был организован децентрализованным образом на нескольких уровнях. На первом этапе управления по регистрации населения должны были передать свои массивы данных в статистический орган Земли (Länder). Для этого, исходя из институционального стандарта массивов данных в регистрах населения, было определено стандартизированное определение записи данных. На следующем этапе регистры были импортированы Землями в централизованную базу данных, администратором которой является Федеральное статистическое управление. Земли отвечали за обработку данных, опираясь на автоматизированную поддержку. Для интеграции и проверки достоверности данных было разработано приложение. При помощи этого приложения, предоставленного Федеральным статистическим управлением, проводились проверки правильности параметров, значений и комбинированных значений переменных. После передачи и проверки данных Федеральное статистическое управление имело доступ к различным децентрализованным данным. Эти регистры необходимо было консолидировать, используя информацию по адресам и персональные сведения в каждом массиве данных. Сначала адресные переменные были скорректированы в рамках регистра адресов и зданий, далее личные переменные сформировали регистр населения, привязанный к регистру

адресов и зданий через единый идентификационный номер (идентификатор адреса). Таким образом был построен временный централизованный регистр населения в Германии.

11. Несмотря на то, что использовалось стандартное описание записей данных и переменные имели однородные определения и значения, качество регистров населения иногда было разным, поэтому контроль данных был важным шагом во время их обработки. Одна из основных задач заключалась в урегулировании расхождений между регистрами. Качество регистров далее зависит от поведения регистра, например, после удаления или перемещения. Поэтому может произойти, что человек зарегистрирован в нескольких местах или все еще зарегистрирован, но уже не проживает там.

## **IV. Контроль качества**

12. В целях повышения качества регистров населения перепись в Германии включает два метода исправления ошибок в регистрах населения, особенно в целях определения численности населения. С одной стороны, необходимо скорректировать массив данных регистров. С другой стороны, выборочное обследование выполняет функцию обеспечения качества данных по численности населения путем количественного определения, оценки избыточного и недостаточного охвата в регистрах населения и их статистической корректировки.

### **A. Коррекция массивов данных**

13. Для правильного представления переменной по «месту обычного проживания» необходимо скорректировать данные регистров на множественные или неверные записи, которые не допускаются в соответствии с законом о регистрации населения, но, тем не менее, случаются. Предусматривается, что граждане регистрируются либо с единственным местожительством, либо с основным местожительством и одним или несколькими второстепенными местожительствами. Регистрация с более чем одним единственным или основным местожительством или одним или несколькими второстепенными местожительствами не допускается. В связи с этим центральный массив данных из регистров населения был проанализирован с целью выявления противоречащих законодательству регистраций. Во время проверки на предмет множественной регистрации, двойные счета и регистрации, противоречащие закону о регистрации, должны были быть выявлены и скорректированы в массиве данных. Для привязки потока данных регистров к дате переписи (2011/05/09) регистры населения были запрошены на две базовые даты, которыми являлись базовая дата переписи и три месяца после базовой даты переписи. Из второго сбора данных для последующих шагов переписи использовалась только часть данных. Для дальнейших процессов были приняты только массивы данных с указанной датой до базовой даты переписи, точнее говоря, дата рождения и дата заселения должны были предшествовать дате переписи.

14. Для определения некорректных массивов данных необходимо было сравнить постоянные переменные. В процессе поиска некорректных наборов данных в качестве наиболее постоянных характеристик использовались такие переменные как «фамилия», «имя», «пол», «дата рождения» и «место рождения».

15. Для проверки и корректировки сводного массива данных необходимо было сравнить разные массивы данных с целью выявления данных, повторяющихся дважды, трижды и т.д. На первом этапе записи, повторяющиеся дважды,

определялись как идентичные по основным демографическим характеристикам. Поскольку регистры ведутся децентрализованно, без единого уникального идентификационного номера, необходимо было найти записи, повторяющиеся дважды, имеющие одинаковые значения, особенно по таким переменным как «фамилия», «имя» и «место рождения». В частности, записи регистров по второстепенным местожительствам или по жителям, имеющим опыт миграции, могут отличаться в регистрах населения. Помимо прочего это связано с тем, что значения по зарегистрированным второстепенным местожительствам не используются для официальных документов, таких как удостоверение личности или карт для участия в выборах. В связи с этим содержание данных в регистрах иногда не проверяется так же тщательно, что и содержание по единственному или основному местожительству. По жителям, имеющим опыт миграции и более одной зарегистрированной записи, иногда имеются разные значения в связи с тем, что транскрипция из иностранных языков, например, кириллических или азиатских, была сделана по-разному.

16. После поиска дублирующих и единственных записей данных необходимо проверить тип местожительства. В этом процессе возникли разные комбинации, как указано ниже.

- (1) *единственная запись с основным местожительством*
- (2) *единственная запись с второстепенным местожительством*
- (3) *дублирующие записи с основным/единственным местожительством*
- (4) *дублирующие записи с основным/единственным и второстепенным местожительством*
- (5) *дублирующие записи, в обеих из которых указано второстепенное местожительство*

17. В соответствии с законом о регистрации населения комбинации (1) и (4) являются верными. Остальные комбинации необходимо было очистить.

18. Для этого для граждан, зарегистрированных с одним или более второстепенными местожительствами – комбинация (2) и (5), – необходимо было выяснить по какому адресу соответствующее лицо следовало учитывать на базовую дату. В этих случаях записи данных с второстепенным местожительством невозможно было соединить с другой записью данных, где указано единственное или основное местожительство. Местожительство необходимо было прояснить при помощи обследования. Такое обследование по выяснению местожительства граждан было проведено для того, чтобы гарантировать, что лица, в отношении которых были выявлены расхождения в регистрах населения, правильно учтены при определении официальной численности населения. С этими людьми связывались, направив им письмо. Они получили анкету, где им было необходимо указать тип местожительства. Такое обследование было организовано Землями, для представления результатами ими применялось центральное приложение. Респондент отвечал, сообщая, что это являлось его основным местожительством или второстепенным местожительством и сообщал основной адрес проживания. Также могло произойти, что анкета не могла быть доставлена. Все различные результаты собирались и анализировались на центральном уровне. В заключение все неверные записи были либо указаны как основные местожительства, либо получали метку к удалению.

19. Комбинация (3) также содержит неверную комбинацию записей. Такие дублирующие записи могут быть разделены на временные множественные записи и постоянно неверные записи. Причиной возникновения первых является запаздывание

по времени между регистрацией и снятием с регистрации. Для этого использовалось второе представление данных из регистров населения спустя три месяца после базовой даты переписи для того, чтобы точно отследить (по отношению к базовой дате) регистрации и снятия с регистрации, которые соответствуют друг другу, но не были сделаны одновременно. Здесь данные из регистра населения, указывающие на то, что человек въехал на определенный адрес за день до базовой даты переписи, были объединены с первым представлением данных из регистров в целях переписи для формирования одной записи. Такие временные множественные записи корректировались автоматически, когда текущая запись данных оставалась активной, а вторая получала метку к удалению. В ходе этого процесса текущая запись данных определялась как запись данных с более актуальной датой въезда, что означает дату, более близкую к базовой дате переписи.

20. Помимо этого, комбинированный массив данных из регистров населения содержал постоянно неверные записи. Такие дублирующие записи необходимо было очистить. Для прояснения неверных записей данных такие записи были разделены на дублирующие записи, относящиеся к адресам, находящимся в населенных пунктах с численностью населения менее 10 тыс. человек, и дублирующие записи, относящиеся к адресам, находящимся в населенных пунктах с численностью населения более 10 тыс. человек. Основанием для такого разделения является то, что регистры населения также корректировались при помощи выборочного обследования. Но вследствие метода выборочного обследования и небольшого размера выборки в малых населенных пунктах, в таких населенных пунктах было неэффективно корректировать регистры населения на основании результатов выборочного обследования. Исправлять такие неверные записи в регистрах населения необходимо было путем опроса, в отличие от корректировки регистров в крупных населенных пунктах, где неверные записи в регистрах очищались автоматически. При таком разделении стало возможным обеспечить, что для статистической корректировки регистров населения, как в крупных, так и малых населенных пунктах использовались первичные данные.

21. Неверные дублирующие записи в крупных населенных пунктах корректировались таким образом, что запись данных с предшествующей датой въезда считалась более старой. Таким образом, в массиве сохранялась запись данных с датой въезда, более близкой к базовой дате переписи; другая запись данных получала метку к удалению. Неверные записи данных в малых населенных пунктах необходимо было корректировать иным образом. В этом случае каждая часть дублирующей записи получала анкету, направляемую Землей. В рамках опроса граждан, зарегистрированных только с второстепенным местожительством, респонденту было необходимо сообщить тип местожительства, по адресу которого он получил анкету. В результате опроса различные ответы были сведены вместе и сравнены. Из ответов на анкеты можно было определить можно ли объединить две записи данных как основное и второстепенное местожительство, или одна анкета не была доставлена, а на вторую анкету был получен ответ, свидетельствующий о том, что данное местожительство является основным. Если оба респондента отвечали, что адрес является их постоянным местожительством, две записи данных необходимо было разъединить. Результаты и комбинации, которые не соответствовали друг другу, или случаи, когда на анкеты не были получены ответы, необходимо было корректировать на центральном уровне таким же образом, как очищались дублирующие записи в крупных населенных пунктах.

22. Данные регистров населения, скорректированные на множественные записи или на записи только с второстепенным местожительством, используются для дальнейшего процесса реализации модели переписи. Результаты обследования или автоматическая очистка данных будут применяться только для определения

официальной численности населения. Данные не будут обратно подаваться обратно в управления регистрации граждан или другие органы.

## **В. Корректировка посредством выборочного обследования**

23. После корректировки центрального массива данных регистров населения, используя информацию, имеющуюся в регистрах, обследование домохозяйств обеспечило качество данных по численности населения. Данное выборочное обследование охватывало переменные и группы населения, которые не содержались в регистрах. Например, посредством регистров занятости было невозможно охватить переменные по занятости самозанятых или лиц, имеющих временную работу, кроме того, ни в одном из административных регистров не было переменных по образованию. Для восполнения этого пробела было проведено выборочное обследование с целью охвата переменных и групп населения, не содержащихся в регистрах. Помимо этой задачи обследование домохозяйств охватывало ряд жителей, проживающих в элементе выборки. Данная информация обеспечила качество данных по численности населения посредством количественного определения и статистической корректировки избыточного и недостаточного охвата регистров населения. Для этого интервьюеру необходимо было определить всех лиц, проживающих по адресу. Для подкрепления интервью интервьюер получал список лиц, зарегистрированных по соответствующему адресу. Таким образом можно было соотнести лиц, проживающих по адресу и определенных в ходе выборочного обследования домохозяйств, с лицами, зарегистрированными по этому адресу в соответствии с информацией в регистре населения. Это позволило определить степень избыточного и недостаточного охвата регистров населения и сделать необходимые статистические поправки. В населенных пунктах с численностью населения более 10 тыс. человек информация выборочного обследования использовалась для корректировки численности населения путем экстраполяции результатов, полученных в части избыточного и недостаточного охвата. В населенных пунктах с численностью населения менее 10 тыс. человек объем выборки был недостаточно большим для использования результатов для экстраполяции. В этих населенных пунктах для корректировки регистров только по единице выборки использовались результаты анкет, примененных в целях внесения поправок в центральный массив данных.

## **V. Регистры населения – базовый массив данных**

24. Помимо использования регистров населения в качестве источников для переменных переписи записи применялись для построения базового массива данных переписи. Этот массив данных выполнял руководящую функцию для взаимодействия между различными компонентами переписи, поскольку позволяет консолидировать информацию из разных источников данных. Для этого консолидация была организована в два этапа. Сначала была сделана привязка адресов в разных источниках, используя регистр адресов и зданий. Затем были выявленные граждане, правильно зарегистрированные по адресу. Это означает, что вся информация, основывающаяся на одном из компонентов переписи (регистр, выборочное обследование домохозяйств и сплошной пересчет) необходимо было слить с наборами данных из регистров населения при помощи регистра адресов и зданий. Таким образом, массив данных на основе регистров населения называется базовым массивом данных. Каждый набор данных из базового массива данных привязан к адресу в регистре адресов и зданий. Во время процесса консолидации информация и данные из разных компонентов, основывающиеся на отдельных переменных, были

скорректированы с наборами данных из базового массива данных в целях дополнения или корректирования необходимых переменных. Это означает, что с одной стороны, были добавлены переменные или целые наборы данных, такие как информация, основывающаяся на выборочном обследовании, а с другой стороны, переменные были скорректированы, например, информация о типе местожительства. Задача базового массива данных в модели переписи заключается в том, чтобы гарантировать объединение всей информации в один массив данных. Для этого каждый компонент модели переписи должен был объединить свою информацию и должен был использовать информацию, содержащуюся в базовом массиве данных. Поскольку в Германии в каждой Земле имеются разные статистические органы, важно, чтобы все Земли могли корректировать и использовать базовую базу данных. Поэтому было необходимо разработать веб-приложение.

## VI. Заключение

25. Новая перепись в Германии является сложной процедурой, требующей координации и взаимодействия между отдельными компонентами и источниками данных. Несмотря на определенные преимущества, использование регистров также приводит к различным проблемам, которые необходимо учитывать и решать на этапе планирования и реализации. Различия в уровне качества данных в разных регистрах делают объединение информации особенно сложным. Кроме того, сложность объединения данных обусловлена отсутствием в Германии стандартных идентификационных номеров граждан или жилищ и децентрализованным хранением данных во многих регистрах. Тем не менее, регистры населения обеспечивают сведения обо всех жителях, которые позволяют проводить детальную интерпретацию данных. В частности, для регистров населения необходимо, прежде всего, создать массив данных на федеральном уровне, который затем должен координироваться с остальными источниками данных. Такой центральный базовый массив данных необходим для реализации сложной модели переписи и консолидации разных источников данных.

---