



**Экономический  
и Социальный Совет**

Distr.: General  
14 March 2012  
Russian  
Original: English

---

**Европейская экономическая комиссия**

**Конференция европейских статистиков**

**Группа экспертов по переписям населения  
и жилищного фонда**

**Четырнадцатое совещание**

Женева, 24–25 мая 2012 года

Пункт 3 предварительной повестки дня

**Доступ к микроданным**

**Планы в области доступа к микроданным переписи  
населения Соединенного Королевства 2011 года**

**Записка Управления национальной статистики Соединенного  
Королевства**

*Резюме*

В настоящем документе рассказывается о планах подготовки предлагаемых материалов микроданных по итогам переписи населения, проходившей в 2011 году в Англии, Уэльсе, Шотландии и Северной Ирландии, а также о тех продуктах, которые предлагались в 1991 и 2001 годах.

В документе рассматриваются последствия изменений в законодательстве Соединенного Королевства и методологии контроля за разглашением статистической информации, которые произошли между 2001 и 2011 годами, а также работа по проведению консультаций с пользователями и подготовке спецификаций микроданных.

Затем в документе описываются текущие планы по подготовке микроданных на основе результатов переписи 2011 года и соответствующие уровни доступа.

## I. Введение

1. Настоящий документ подготовлен Эммой Уайт при содействии Дениз Макгрегор, Кейт Спайсер и Джо Трэйнора из Управления национальной статистики, а также Кейт Дагмор из компании "Демографик десижнс Лтд."
2. Перепись, являющаяся источником самой полной информации о населении Соединенного Королевства, с 1801 года проводится каждые десять лет (за исключением 1941 года и 1966 года, в который перепись проводилась по 10-процентной выборке). Данные переписи отличаются очень широким охватом и высокой степенью подробности, и они используются в процессе планирования и распределения ресурсов, а также в научных, коммерческих и иных целях. Новым аспектом этой деятельности по сбору данных является требование к государствам-членам представить к 31 марта 2014 года набор данных о своем населении в Евростат.
3. Последняя перепись населения в Англии, Уэльсе, Шотландии и Северной Ирландии была проведена по состоянию на 27 марта 2011 года. Ответственность за представление результатов переписи по Англии и Уэльсу лежит на Управлении национальной статистики (УНС), а по Шотландии и Северной Ирландии соответственно на Национальном регистре Шотландии (НРШ) и Агентстве статистики и исследований Северной Ирландии (НИСРА). После того как в июле 2012 года каждое из этих трех учреждений опубликует свои национальные данные, координацию публикации данных по всему Соединенному Королевству будет осуществлять УНС.
4. УНС уверено в том, что национальный целевой показатель доли жителей, ответивших на вопросы переписи, который для Англии и Уэльса в ходе переписи 2011 года был установлен на уровне 94%, будет выполнен, равно как и целевой показатель по каждому муниципальному образованию (МО), который не должен быть меньше 80%. При этом добровольное обследование охвата переписи (ООП) имело показатель ответивших на уровне примерно 90%. Выполнение этих целевых показателей будет означать, что погрешность демографических оценок УНС по Англии и Уэльсу составит  $\pm 0,2\%$ , а по МО  $\pm 3\%$ . Первые результаты внушали оптимизм и в отношении выполнения целевого показателя не ответивших на отдельные вопросы, который не должен превышать 5% применительно к 70% вопросов.
5. Независимый опрос пользователей показал, что методы оценки и корректировки охвата, а также обеспечения качества данных переписи населения Англии и Уэльса 2011 года "позволяют с уверенностью говорить о том, что полученные с их помощью окончательные данные переписи будут лучше данных, полученных с использованием любых других методов, и что они будут пригодны для целей планирования и распределения ресурсов".
6. На сегодняшний день по итогам консультаций с пользователями для переписи 2011 года были подготовлены 100 новых таблиц, что является увеличением на 25% по сравнению с 2001 годом. Одни будут содержать данные, собираемые в 2011 году впервые, а другие восполнять пробелы в результатах переписи 2001 года, на которые указали пользователи. Если удастся обеспечить необходимое качество данных, информация о населении будет представлена и в иных разрезах – по месту работы, продолжительности рабочего дня, деятельности в перерывах между учебой, краткосрочной миграции и основному роду занятий. Эти планы призваны расширить спектр результатов переписи, улучшить

доступ к переписным данным, удовлетворить потребности большего числа пользователей, повысить эффективность использования данных и получить дополнительную отдачу от затраченных на проведение переписи средств.

7. Консультации продемонстрировали важность составления подробных и точных данных по малым группам населения. В результате по ним будет выпускаться целый ряд статистических продуктов. Кроме того, поскольку этнические и религиозные группы, конкретно не указанные в вопроснике переписи, четко дали понять, что они хотят получить официальное признание наряду с остальными, во второй выпуск таблиц переписи (первый, в котором будут приводиться сведения об этнических и религиозных группах) войдут оценочные данные по всем этническим и религиозным группам, которым УНС присвоило свой классификационный код, вплоть до самого низкого географического уровня переписи – переписного участка (при условии соблюдения требований в отношении конфиденциальности статистической информации). Что касается альтернативных демографических оценок, то для их получения в Соединенном Королевстве отсутствуют источники данных, эквивалентные переписи населения. То же самое касается и микроданных.

## II. Микроданные в Соединенном Королевстве

8. Продольное исследование УНС (ПИ) позволяет собирать сведения об одном проценте населения Англии и Уэльса. Сегодня ПИ содержит данные индивидуального уровня в отношении более чем 500 000 человек, в том числе данные всех переписей начиная с 1971 года. Ранее микроданные получали путем обезличивания данных ПИ (исключения фамилий и адресов). Хотя выборка в рамках этого набора данных носит обезличенный характер, по-прежнему сохраняются данные индивидуального уровня, остающиеся незавуалированными или неагрегированными. Соответственно доступ к ним контролируется.

9. Выборки обезличенных файлов (ВОФ) были впервые составлены по итогам переписи 1991 года. Были выпущены два продукта – каждый отдельно по Великобритании и по Северной Ирландии. С подробностями можно ознакомиться в таблице 1.

Таблица 1  
ВОФ 1991 года

	<i>Выборка</i>	<i>География</i>	<i>Лицензия</i>	<i>Доступность</i>	<i>Стоимость</i>
ВОФ физических лиц	2% физических лиц (>1,1 млн. файлов)	288 районов ВОФ с численностью населения >120 000 человек	Лицензия конечного пользователя (ЛКП). Подача заявок через Центр переписей и выборочных обследований (ЦПВО) им. Кэти Марш	Академические пользователи: данные можно скачать в SPSS, Stata или TAB. Доступны также в Интернете. ЦПВО	Бесплатно (до октября 2011 года с неакадемических пользователей взималась плата)

	<i>Выборка</i>	<i>География</i>	<i>Лицензия</i>	<i>Доступность</i>	<i>Стоимость</i>
				Неакадемиче- ские пользова- тели: компакт- диск	
ВОФ домохо- зяйств	1% домохо- зяйств (216 000 фай- лов, >500 000 че- ловек)	Региональная	Как и для ВОФ физиче- ских лиц	Как и для ВОФ физиче- ских лиц	Как и для ВОФ физиче- ских лиц

*Источник:* Dugmore, Keith. A business case for microdata files from the 2011 Census. Final Report, April 2009. Commissioned by the University of Manchester, and carried out by Demographic Decisions Ltd, London and ONS.

10. Как говорится в докладе компании "Демографик Десижнс Лтд.", ВОФ 1991 года способствовали резкому росту социальных исследований, опирающихся на новые данные, поскольку исследователи в разных областях быстро оценили преимущества новых файлов, в первую очередь в плане доступности.

11. В то же время итогом переписи населения 2011 года стал набор более сложных продуктов, что стало следствием роста озабоченности по поводу разглашения персональных данных. Пять разных продуктов стали предлагаться в рамках трех различных лицензионных соглашений, ограничивающих доступ к данным в зависимости от степени их подробности. См. таблицу 2.

Таблица 2

**Выборки обезличенных файлов переписи 2001 года**

	<i>Выборка</i>	<i>География</i>	<i>Лицензия</i>	<i>Доступность</i>	<i>Стоимость</i>
ВОФ фи- зических лиц	3% (1,84 млн. файлов)	Файл Соединен- ного Королевст- ва (9 регионов, малый и боль- шой Лондон)	ЛКП через ЦПВО	Академические пользователи: данные можно скачать в SPSS, Stata или TAB. Доступны так- же в Интерне- те. ЦПВО	Бесплатно (до октября 2011 года с не- сколькими пользователей взималась пла- та)
				Неакадемиче- ские пользова- тели: компакт- диск	
Микро- данные по небольшим рай- онам (MHR)	5% (2,96 млн. файлов физи- ческих лиц)	Файл Соединен- ного Королевст- ва (муниципаль- ные образования (МО) или округа Северной Ир- ландии (СИ))	Как и для ВОФ физиче- ских лиц	Как и для ВОФ физических лиц	Как и для ВОФ физических лиц

	<i>Выборка</i>	<i>География</i>	<i>Лицензия</i>	<i>Доступность</i>	<i>Стоимость</i>
ВОФ домохозяйств/специальная лицензия	1% (>200 000 файлов, >500 000 человек)	Только Англия и Уэльс (не по географическому признаку)	Специальная лицензия и допуск	Архив Соединенного Королевства (АСК) и УНС	Как и для ВОФ физических лиц
Выборка микроданных по физическим лицам с контролируемым доступом (ВМКД)	3% (такая же выборка, что ВОФ физических лиц)	Файл Соединенного Королевства (МО или округа СИ)	Специальный допуск	Лаборатория виртуальных микроданных (ЛВМ) УНС	Как и для ВОФ физических лиц
ВМКД по домашним хозяйствам	1%	Как и для ВМКД физических лиц	Как и для ВМКД физических лиц	Как и для ВМКД физических лиц	Как и для ВОФ физических лиц

12. Механизм специального допуска был создан в 2007 году. Необходимо подать специальную заявку на получение доступа к информации в Лабораторию виртуальных микроданных (ЛВМ), которая, прежде чем она будет удовлетворена (как правило, в течение недели), должна быть проверена на предмет опасности разглашения сотрудником УНС. Физический доступ к ЛВМ можно получить лишь через одно из трех бюро УНС в Англии и Уэльсе. Многие пользователи данными переписей были недовольны механизмом доступа к микроданным 2001 года, считая, что доступ к аналогичным данным в 1991 году в ряде случаев был проще. Тем не менее эти продукты широко использовались в исследовательской и аналитической деятельности как самостоятельно, так и совместно с другими данными.

### III. Изменения за период с 2001 года

13. В период с 2001 года по 2011 год произошел ряд изменений, повлиявших на подход УНС к разработке продуктов микроданных:

а) Закон о регистрации статистики и услуг 2007 года наделяет УНС законными полномочиями предоставлять данные получившим соответствующий допуск лицам и определяет критерии предоставления таких данных;

б) растут вычислительные мощности как УНС, так и других структур. В результате этого возможности предоставления данных расширяются и у пользователей появляется потребность в ускорении и упрощении процедуры доступа к статистике в целом и к микроданным в частности, а также в получении более подробных данных по сравнению с теми, которые предлагает УНС;

с) политика повышения прозрачности и открытости органов управления заставляет государственные ведомства предоставлять больше информации бесплатно;

д) как правительство, так и общественность высказывают все большую озабоченность проблемами защиты частной жизни и конфиденциальности информации. По сравнению с 2001 годом, требования к защите данных стали гораздо жестче;

е) появление открытой государственной лицензии (ОГЛ) позволяет пользователям копировать, публиковать, распространять и передавать информацию (с указанием источника); адаптировать информацию; и использовать информацию в коммерческих целях, например объединяя ее с другой информацией или используя ее в собственных продуктах или программах. В то же время пользователи обязаны соблюдать требования Закона о защите данных 1998 года и не должны искажать информацию.

#### **IV. Статистическая методология борьбы с разглашением данных (МБРД) переписи 2011 года**

14. По итогам анализа применения в 2001 году методологии случайной перестановки данных одновременно с корректировкой малых ячеек (КМЯ) к данным переписи 2011 года в масштабах всего Соединенного Королевства будет применяться метод целевой перестановки создающих риск идентификации данных по домохозяйствам. Руководство статистических служб согласилось с тем, что малые наблюдаемые значения (0,1 и 2) могут включаться в публикуемые переписные таблицы в следующих случаях:

а) в случае систематического возникновения достаточно серьезной неопределенности в отношении достоверности малой ячейки; и

б) в случае, когда такая неопределенность не причиняет серьезного вреда данным.

15. Оценить, насколько "серьезной" является неопределенность, можно по двум факторам: проценту данных, описывающих реальные признаки, которые удастся защитить путем перестановки и импутации, и проценту созданных очевидных случаев идентифицируемых признаков. Масштабы перестановки данных на той или иной территории зависят от уровня непредоставления ответов и импутации, причем уровень перестановок обратно пропорционален количеству случаев импутации.

16. Применение согласованной методологии МБРД будет давать больший эффект на уровне малых ячеек, который при переходе на более высокие географические уровни будет ослабевать. Отсутствие КМЯ обеспечивает отсутствие несогласованности между таблицами. Таблица 3 описывает главные различия между методологиями, применявшимися УНС в ходе переписей 2001 и 2011 годов, и эффект их применения.

Таблица 3  
Различия в МБРД 2001 и 2011 годов

	2001	2011
Перестановка данных	Случайная	Ориентированная на сопряженные с риском идентификации данных
Корректировка таблиц	КМЯ в Англии, Уэльсе и Северной Ирландии	Отсутствует
Аддитивность	Не во всех таблицах	Да
Согласованность	Нет	Да
Гармонизация МБРД	Нет	Да – хотя параметры могут различаться между странами
Влияние на малые наблюдаемые значения	Сильное	Некоторое

Источник: ONS SDC.

## V. МБРД и микроданные 2011 года

17. Планы подготовки микроданных 2011 года направлены на обеспечение оптимального баланса между достоверностью и полезностью с учетом следующих факторов:

- компромисса между перекодированием и искажением;
- придания выборкам взаимоисключающего характера;
- использования новых переменных и категорий; и
- лицензирования и допуска.

## VI. Консультации

18. Консультации по важнейшим статистическим материалам переписи населения 2011 года проходили в два отдельных этапа: с 14 декабря 2009 года по 26 марта 2010 года и с 7 февраля 2011 года по 28 апреля 2011 года.

19. Эти консультации, участникам которых было предложено заполнить и вернуть официальные анкеты со своими соображениями, дополнялись национальными консультациями с общественностью, а также прямыми контактами с основными пользователями и группами пользователей.

20. За консультациями последовали целевые рабочие совещания, посвященные специализированным продуктам, в том числе микроданным. После прошедшего 11 ноября 2011 года рабочего совещания по микроданным информационные записки, материалы, анкеты с вопросами и ответами были предоставлены пользователям переписной информации, которым было также предложено представить свои замечания до 19 декабря 2011 года. Документация и материалы консультаций были обобщены, и в январе 2012 года был нанят внешний эксперт, которому было поручено проанализировать всю информацию и разработать подробные спецификации микроданных для переписи 2011 года.

На момент написания настоящей записки работа над спецификациями продолжалась. Пакет спецификаций самых важных и востребованных микроданных, равно как и окончательный доклад должны быть готовы 30 марта 2012 года. Свой вклад в эту деятельность внесла также Рабочая группа по стратегии микроданных переписи (РГСМП) 2011 года, которая была учреждена в мае 2007 года и в состав которой вошли представители УНС, НРШ, НИСРА и Центра переписей и выборочных обследований (ЦПВО) Манчестерского университета.

## **VII. 2011 год: по сравнению с 2001 годом набор продуктов стал проще и доступнее**

21. Для переписи 2011 года запланированы три уровня микроданных:

### **A. Открытый файл**

22. Для УНС такой файл является новшеством. Выпуск этого не содержащего конфиденциальных сведений файла по физическим лицам согласуется с политикой правительства Соединенного Королевства, стремящегося сделать свою деятельность максимально прозрачной. Этот файл можно будет загрузить с веб-сайта УНС. Он будет не столь подробным, как ВОФ физических лиц, доступ к которому можно было получить, приобретя лицензию конечного пользователя (ЛКП) в 2001 году. Если же он выйдет более подробным, то в него не будет включена географическая информация. Данные файла будут легкодоступными и пригодны для использования в качестве методического и учебного материала, а также для высокопрофессионального изучения связей между переписными переменными. Этот файл потенциально можно будет использовать, хотя и в ограниченной мере, для проверки синтаксической структуры на ошибки до посещения ЛВМ. Файл ни одного домохозяйства не будет находиться в открытом доступе.

### **B. Защищенный(ые) файл(ы)**

23. Этот файл также будет содержать информацию о физических лицах, и доступ к нему пользователи смогут получить через свои персональные компьютеры. Планируется сделать его столь же подробным, как и файлы микроданных по малым районам (ММР) и ВОФ физических лиц, возможно, на условиях ЛКП 2001 года. Этот файл можно рассматривать как возврат к спецификации ВОФ физических лиц 1991 года, но при большей выборке. Рассматривается и возможность составления файла домохозяйств. Вместе с правовыми службами УНС ведется работа над определением понятия "защищенный" и составлением соответствующих лицензионных соглашений.

### **C. Файлы ЛВМ**

24. Эти анонимные, но поддающиеся идентификации данные будут храниться в защищенном формате и в целом будут аналогичны выборке микроданных с контролируемым доступом (ВМКД) 2001 года на уровне как физических лиц, так и домохозяйств. В то же время, размер выборки этих данных будет составлять 3–5%.

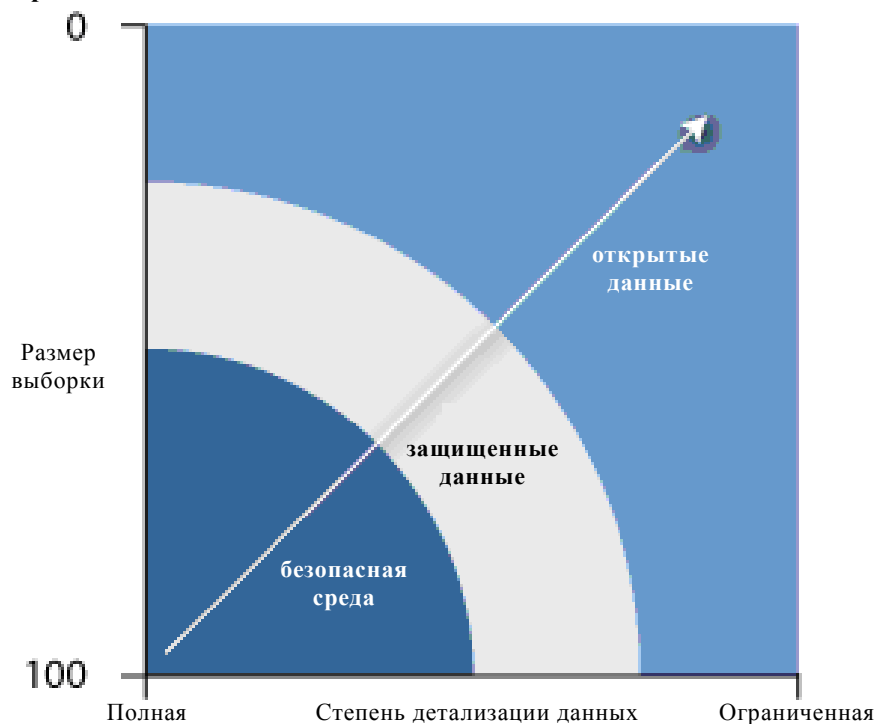


На диаграмме 1 изображены уровни доступа к данным 2011 года.

В таблице 4 описываются предлагаемые микроданные 2011 года.

Диаграмма 1

**Предлагаемые уровни доступа: размер выборки (в %) и степень подробности данных**



Источник: ONS Data Visualisation.

Таблица 4

**Предлагаемые микроданные 2011 года**

Файл	Подробность	География	Лицензия	Доступность
Открытые ВОФ	3%	По 9 регионам с меньшими подробностями или без географической привязки с большими подробностями	ОГЛ	Загрузка с вебсайта
ВОФ физических лиц	5%	МО – тот же уровень подробности, что и ВОФ физических лиц 2001 года	Защищенные данные	Загрузка с вебсайта при условии заключения специального соглашения
ВОФ домохозяйств	1%	По 9 регионам или без географической привязки	Как и для ВОФ физических лиц	Как и для ВОФ физических лиц

Файл	Подробность	География	Лицензия	Доступность
ВКМД по физическим лицам	Больше подробностей, чем в 2001 году? Можно увеличить размер выборки	Не меньшая по степени подробности, чем в 2001 году	Требуется доступ	ЛВМ УНС (защищенный формат)
ВКМД по домохозяйствам	Как и ВКМД физических лиц	Как и ВКМД физических лиц	Как и ВКМД физических лиц	Как и ВКМД физических лиц

Источник: ONS.

## VIII. Сроки выхода данных

25. Согласно нынешним планам, файлы микроданных переписи 2011 года станут доступны в 2013 году. По мнению пользователей, приоритетность выхода данных должна быть следующей: ВОФ физических лиц и домохозяйств (защищенные данные), ВКМД (ЛВМ) по физическим лицам и домохозяйствам и, наконец, открытые ВОФ. Пользователи просят также информировать их о ходе подготовки этих продуктов и времени их публикации.

## IX. Динамичная Интернет-сводка

26. Сводка итогов переписи 2011 года задумывалась УНС как динамичный онлайн-документ, который впервые выйдет в свет в марте 2012 года. Подписавшимся на нее пользователям новые данные будут направляться по электронной почте сразу же после их появления. Поскольку файлы микроданных являются одним из важнейших продуктов переписи, в сводке им уделяется соответствующее внимание.

## X. Консультирование пользователей, инструктаж и поддержка продуктов

27. Планируется сделать все необходимое для подготовки сотрудников, которые могли бы оказывать консультационную помощь, вести инструктаж и осуществлять поддержку продуктов переписи 2011 года, в частности микроданных. УНС обсудит вопрос о том, как лучше сделать, с РГСМП и Советом экономических и социальных исследований (СЭСИ), поскольку этот совет, который раньше финансировал участвовавшие в проведении переписей службы, в том числе ЦПВО, после завершения нынешнего цикла финансирования переписей в июле 2012 года приступит к созданию нового национального сервиса данных.

## XI. Выводы

28. УНС понимает важность ВОФ для пользователей: к достоинствам этого продукта относятся больший размер выборки, возможность анализировать богатые и гибкие наборы данных, возможность проверить итоговые данные преж-

де, чем они будут сведены в заказанные таблицы, а также возможность провести статистический анализ с использованием файлов, сопоставимых со стандартными результатами переписей, поскольку они имеют одни и те же источники.

29. Соединенное Королевство способно сегодня провести серьезный анализ временных рядов микроданных, поскольку в его распоряжении имеются файлы переписей 1991 и 2001 годов, а файлы за 2011 год оно получит в 2013 году. Кроме того, рассматривается вопрос о подготовке файлов микроданных переписей, предшествовавших переписи 1991 года.

*Источник:* ONS Data Visualisation.

30. Подготовка продуктов переписи 2011 года велась с акцентом на потребности пользователей. Это касается и микроданных, если принимать во внимание сложность повышения подробности данных с точки зрения МКРД, лицензирования и требований доступа. Тем не менее, УНС уверено в том, что отвечающие требованиям, доступные файлы микроданных по Соединенному Королевству будут готовы в 2013 году.

## **Справочная литература**

Dugmore, Keith. A business case for microdata files from the 2011 Census. Final Report, April 2009. Commissioned by the University of Manchester, and carried out by Demographic Decisions Ltd, London.

---