



**Экономический  
и Социальный Совет**

Distr.: General  
30 April 2012  
Russian  
Original: English

---

**Европейская экономическая комиссия**

**Конференция европейских статистиков**

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

**Четырнадцатое совещание**

Женева, 24 и 25 мая 2012 года

Пункт 3 предварительной повестки дня

Доступ к микроданным

**Проекты IPUMS-International (Интегрированные микроданные  
общего пользования) и ИЕСМ (Интегрированные микроданные  
европейских переписей) снижают риск предоставления  
трансграничного доступа и открывают новые возможности<sup>1</sup>**

**Записка Миннесотского центра по народонаселению, Миннеаполис,  
Миннесота, США и Центра демографических исследований Автономного  
университета Барселоны, Барселона, Испания**

**«Распространять данные – значит открывать их ценность всему миру»**

Семинар по новым тенденциям в передаче данных и статистике, Нью-Йорк, 19 февраля  
2010 года. Вальтер Радермахер (Президент Евростата) и Петер Эверарс (Руководитель  
Директората Евростата)

**I. Введение**

1. Представителям науки и правительственных структур необходим доступ к микроданным переписей населения для анализа социальных, демографических и экономических изменений, происходящих в 21 веке (McCaа and Esteve 2009). В наше

---

<sup>1</sup>Средства для данных исследований частично выделены Национальным институтом здравоохранения США, грант HD047283 для осуществления проекта по гармонизации микроданных европейских и азиатских переписей населения (IPUMS-EurAsia).

время в большинстве современных статистических служб понимают, насколько велико значение распространения микроданных переписей. И, тем не менее, многие из них не берут на себя такую ответственность, поскольку делать это в одиночку рискованно, дорого, и часто не оправдывает надежд. Можно передать функцию распространения микроданных третьей стороне, например, международному проекту «Интегрированные микроданные общего пользования» (IPUMS-International) или сотрудничающему с ним проекту «Интегрированные микроданные европейских переписей» (IECM). Это устраним затраты, понизит риск и повысит удовлетворенность пользователей. Статистическим службам, настаивающим на распространении микроданных «своими силами», следовало бы оценить риски, затраты и степень удовлетворенности/неудовлетворенности пользователей.

- **Риски, связанные с распространением микроданных «своими силами».** Лишь некоторые статистические службы обладают опытом обезличивания микроданных переписей, да и сложно приобрести такой опыт при проведении переписи один раз в десять лет. Опыт также теряется с уходом, переходом на другие должности или отсутствием специалистов. Управлять доступом к микроданным чрезвычайно сложно, для этого требуются специальные знания. Лишь немногие службы официальной статистики хотят брать на себя ответственность по выполнению трудоемких задач по проверке пользователей, контролю доступа и обеспечению поддержки пользователям. Статистические службы, настаивающие на самостоятельном управлении доступом к микроданным, должны извлечь урок из скандала, разразившегося по вине Бюро переписей США (Alexander, Davern and Stevenson 2010). Из-за допущенной сотрудником Бюро ошибки в программировании данные по возрасту пожилых людей были безнадежно испорчены в большом числе файлов общего пользования переписи 2000 г., обследования американских сообществ и текущего обследования населения. В целом, были испорчены 144 массива микроданных. Ученые смогли обнаружить ошибки путем сопоставления обезличенной возрастной структуры населения с полными не подвергнутыми обезличиванию микроданными. Этот скандал, разразившийся некстати для Бюро переписей незадолго до переписи 2010 года, угодил на первые страницы Нью-Йорк Таймс. Однако Бюро переписей достойно похвалы за то, что не замалчивало случившегося. Эти ошибки можно обнаружить путем облегчения доступа (под очень строгой защитой) к необезличенным микроданным.
- **Затраты при распространении микроданных «своими силами».** Статистическая служба должна составить выборку, обезличить микроданные, подготовить метаданные и обеспечить доступ к информации. Для выполнения этих задач потребуется много усилий, особенно, если число запросов пользователей не превышает приблизительно десяти в год.
- **Неудовлетворенность от распространения микроданных «своими силами».** Обычно статистические органы допускают ошибку при защите микроданных, прибегая к их чрезмерной анонимизации и запрещению доступа всем, кроме самых настойчивых исследователей. Такая стратегия обеспечивает безопасность, но приносит в жертву пользу. В результате неудовлетворенными

остаются как сами статистические ведомства, так и пользующиеся их данными исследователи. Обратите внимание на жалобы, обращенные к одному из ведущих статистических органов (Управление национальной статистики Великобритании) с более чем двадцатилетним опытом распространения микроданных переписей:

- «Мне не удалось использовать данные; длительность прохождения бюрократических процедур сводит на нет практическую пользу получения информации».
- «Невероятно забюрократизированный и сложный доступ отбивает у людей охоту им воспользоваться. Мне для получения данных потребовалось полгода».

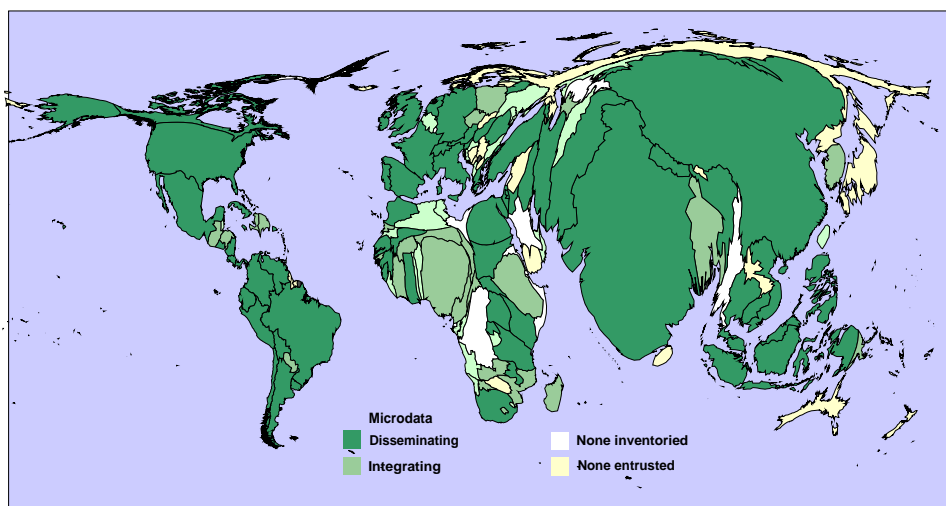
2. Некоторые официальные статистические органы, несмотря на то, что при разработке своих веб-сайтов и развитии стратегии доступа к микроданным они, наверняка, используют передовые технологии, не отвечают на запросы неделями или даже месяцами, а то и не отвечают вовсе. В виртуальном мире такие сайты могут выглядеть грандиозно, но в реальности они приводят к двойному разочарованию, особенно для тех, для кого задержка получения данных является «наихудшей формой отказа».

3. Проекты **IPUMS-I** ([www.ipums.org/international](http://www.ipums.org/international)) и **IECM** ([www.iecm-project.org](http://www.iecm-project.org)) предлагают комплексное универсальное решение по обеспечению доступа к выборкам обезличенных микроданных. В осуществляемом Миннесотским центром по народонаселению (МЦН) проекте IPUMS-International участвуют национальные статистические управления (ЦСУ), университеты и ведущие научно-исследовательские центры всего мира. Центр демографических исследований (ЦДИ) Автономного университета Барселоны ([www.iecm-project.org](http://www.iecm-project.org)) разрабатывает оптимизированные метаданные и развивает сетевую деятельность в Европе. Проекты IPUMS и IECM предлагают современные услуги по представлению микро- и метаданных. Благодаря постоянной финансовой поддержке со стороны Национального научного фонда и Национального института здравоохранения США проект IPUMS-International несет расходы и берет на себя ответственность и риски за восстановление, архивирование, обезличивание, интеграцию и распространение микроданных по всему миру (McCaа and Thomas 2009). Добросовестным исследователям, независимо от страны рождения, места жительства или гражданства, желающим анализировать микроданные, доступ к веб-сайту проекта IPUMS-International обычно предоставляется в течение пяти дней после подачи действительной заявки. Микроданные предоставляются добросовестным ученым и политическим деятелям из любой страны бесплатно.

4. Начиная с 1999 года, проект IPUMS-International удовлетворил запросы более 90 ЦСУ и более 5 000 исследователей из почти 100 стран. 21 европейская страна участвует в проектах IPUMS/IECM на условиях подписания стандартного меморандума о взаимопонимании (Рис. 1). Добросовестные исследователи с помощью стандартного соглашения получают доступ к запрашиваемым микроданным всех стран, участвующих в проекте. На момент написания настоящего доклада проект

завершил анонимизацию и интеграцию микроданных населения по 69 странам, что в целом охватывает 212 выборок и 480 миллионов персональных записей. Благодаря чрезвычайно интенсивному вкладу ЦСУ по всему миру база данных расширяется со скоростью 5-10 новых стран в год. К 2015 г. она, скорее всего, будет включать 85 или более стран, составляющих четыре пятых населения мира.

**Рис. 1. Картограмма IPUMS-International (взвешенная по численности населения). Данные распространяются - темно-зеленый цвет; данные интегрированы – зеленый; на стадии переговоров – светло-зеленый.**



5. **Приглашение к участию и представлению данных переписей раунда 2010 года.** Приглашаем статистические органы, не сотрудничающие с проектом на настоящий момент, связаться с автором доклада для обсуждения возникающих в связи с участием вопросов и сомнений. Предлагаем сотрудничающим службам без лишней задержки представить проекту микроданные переписей населения раунда 2010 года. Это значительно увеличит число заинтересованных сторон и, по словам Президента Евростата «раскроет ценность» микроданных переписей населения (см. эпиграф). 36 наиболее населенных европейских стран распределились по степени участия в проектах IPUMS/IECM следующим образом:

- **17 стран интегрированы:** Армения (1 перепись), Австрия (4), Беларусь (1), Франция (7), Германия (8, включая ГДР и ФРГ и 4 в стадии ввода), Греция (4), Венгрия (4), Ирландия (8), Италия (1), Нидерланды (3), Португалия (3), Румыния (3), Словения (1), Испания (3), Швейцария (4), Турция (3) и Великобритания (6, 4 в стадии восстановления).
- **2 страны в стадии интеграции:** Чешская Республика (2), Польша (4, 2 еще в стадии восстановления).
- **2 страны на стадии передачи микроданных:** Болгария и Украина.
- **15 стран на стадии вступления в сотрудничество:** Албания, Бельгия, Босния и Герцеговина, Хорватия (ведутся переговоры), Дания, Эстония (ведутся

переговоры), Грузия, Латвия, Литва, Македония, Молдова, Норвегия, Россия, Словакия (ведутся переговоры) и Швеция.

6. Проект IPUMS-International делает все возможное для передачи того ценного содержания, которым обладают микроданные переписей населения, как научным и академическим кругам, так и политическим и общественным деятелям. Проект IPUMS-I *не* просто распространяет выборочные данные переписей от ЦСУ к исследователям. Обычно ведущие сотрудники МЦН работают над обезличиванием, интеграцией и обеспечением доступа к микро- и метаданным в течение двух или более лет, затрачивая десятки тысяч рабочих часов каждый.

7. В 21 веке для серьезного научного исследования *недостаточно* просто получения копии исходных микроданных и кодовой книги. В настоящем докладе дается характеристика 12 дополнительным преимуществам, которые предлагают проекты IPUMS-International и IECM.

## **II. Конфиденциальность и безопасность статистических данных**

### **A. Безопасность и статистическая конфиденциальность микроданных.**

8. Ни МЦН, ни его сотрудники никогда не распространяют оригинальные исходные данные. Они являются особенно уязвимыми, и поэтому до окончания процесса анонимизации работать с ними разрешено только опытным, хорошо обученным штатным научным сотрудникам. Сотрудники МЦН за нарушение процедур безопасности подвергаются налагаемому в гражданском порядке штрафу (до 250 000 долларов США), уголовному преследованию и лишению работы. Юридическая служба университета несет ответственность за обеспечение общей безопасности и конфиденциальности предоставленных МЦН данных. IPUMS-I, возможно, является первым академическим проектом по распространению микроданных переписей разных стран, который подвергся анализу извне. Рецензент, г-н Денис Тревин, почетный руководитель статистического бюро Австралии и председатель Целевой группы ЕЭК ООН по обеспечению статистической конфиденциальности и доступа к микроданным, назвал организацию рабочего процесса в МЦН «передовой практикой», «эталоном для лучших статистических служб» и т.д. (Тревин, 2007 г.).

### **B. Защита статистических данных от раскрытия.**

9. Микроданные подвергаются строгому единому правовому, административному и техническому контролю, обеспечивающему статистическим данным всех участвующих статистических служб более надежную групповую защиту, чем защита в одной самостоятельно действующей статистической службе (McCaа, Ruggles and Sobek 2010). Наиболее действенной технической защитой от раскрытия статистических данных является блокирование записей за счет использования подвыборки. Все значения в записях за пределами выборки подавляются. Затем осуществляется блокирование имен и низкоуровневая географическая детализация.

10. Все статистические службы стараются найти компромисс между конфиденциальностью и практической пользой и сообщают проекту IPUMS-I свой минимальный порог идентификации географических единиц. Во многих странах этот порог установлен на уровне 20 000 жителей. В других странах он равняется 100 000 жителей (США), а в самых экстремальных случаях (Нидерланды) подавляется все административно-территориальное деление.

11. Дополнительная защита статистических данных от раскрытия достигается с помощью случайного упорядочивания записей и перестановки географических идентификаторов неопределенного числа домашних хозяйств. Перестановка является чрезвычайно эффективным с точки зрения практичности методом обеспечения конфиденциальности при минимальных затратах. Она гарантирует невозможность идентификации отдельного человека или домашнего хозяйства. В переписи Великобритании 2011 года перестановка применяется не только для защиты конфиденциальности микроданных, но и выходных данных в табличной форме (Frend et al, 2012).

12. По согласованию с национальными статистическими управлениями мы применяем к некоторым переменным кодирование с установлением верхнего предела, к другим – глобальное перекодирование, избирательно убираем разряды иерархических переменных (профессия, отрасль, географический признак) или даже блокируем всю переменную. Решения принимаются по согласованию с соответствующими национальными статистическими органами. Веса и коэффициенты расширения, как правило, трудностей не вызывают, поскольку большая часть выборок потенциально стратифицированы, следовательно все записи имеют одинаковый вес.

### **С. Ограниченный доступ к микроданным переписей.**

13. Несмотря на то, что название проекта IPUMS (Интегрированные микроданные общего пользования) содержит словосочетание «общего пользования», доступ к микроданным ограничен. Потенциальный пользователь должен подать [подробную электронную заявку](#) как для подтверждения своей научной добросовестности, так и для обоснования необходимости доступа к микроданным. Основная часть процесса подачи заявки состоит в отдельном подтверждении согласия с каждым из 10 строгих условий использования данных: запрет на дальнейшее распространение, использование только в научных целях, запрет на использование в коммерческих целях, защита конфиденциальности, обеспечение безопасности, соблюдение строгих правил конфиденциальности, разрешение на публикацию данных в научных трудах, правильное цитирование, возможность применения дисциплинарных мер за нарушения и сообщение об ошибках.

14. По отзывам службы Google Analytics, сама регистрационная форма вызывает у неопытных пользователей значительные затруднения. За последний год 5 593 просмотра страницы регистрации привели к заполнению всего 1 057 заявок. Одна из причин такого большого отсева заключается в том, что заполнение регистрационной формы является сдерживающим фактором для людей, плохо осведомленных в вопросах статистики. Заявитель должен дать согласие на выполнение каждого из 10

строгих условий использования микроданных: запрет на дальнейшее распространение, использование только в научных целях, запрет на использование в коммерческих целях, защита конфиденциальности, обеспечение безопасности, соблюдение строгих правил конфиденциальности, разрешение на публикацию данных в научных трудах, правильное цитирование микроданных, возможность применения дисциплинарных мер за нарушения и сообщение об обнаруженных ошибках.

15. Опытный исследователь, независимо от времени, необходимого для заполнения формы, легко согласится с этими условиями и педантично внесет запрашиваемую информацию, а неопытный, столкнувшись с необходимостью указать службу по защите респондентов своего учреждения, своего руководителя и веб-сайт учреждения, а также дать описание научного проекта, для которого необходимы микроданные, вовсе откажется от заполнения формы. Незаполненные формы автоматически отклоняются системой управления сайтом. Направить для рассмотрения незаполненную форму невозможно.

16. Из 1 057 заполненных форм были отклонены только 46. В большинстве отклоненных запросов имеющиеся для распространения микроданные переписей не соответствовали целям предполагаемых исследований. Пугающие случайных посетителей подробности, необходимые для заполнения формы, приводят к их самоотсеиванию. После подачи заявок заявителей тщательно проверяют на предмет их соответствия или отсутствия научных потребностей в микроданных. Из данного краткого описания процесса регистрации для доступа к сайту IPUMS-International следует, что перед тем, как каждый отдельный заявитель получает свою учетную запись, его, в соответствии с эпитафией, подвергают «тщательной проверке» с целью определения его личности и установления научной добросовестности. Заявка налагает обязательства как на исследователя, так и на его учреждение. Юридическая служба университета Миннесоты готова отреагировать на первое проявление неправильного использования данных.

17. Согласие с условиями использования микроданных налагает обязательства как на исследователя, так и на его учреждение. Юридическая служба университета Миннесоты готова отреагировать на первое проявление неправильного использования. Нарушение правил одним пользователем приостанавливает доступ к микроданным другим пользователям этого учреждения до тех пор, пока они не пройдут дополнительную подготовку по защите респондентов, а учреждение не восстановит своих полномочий для работы с уязвимыми данными. IPUMS-I – единственный академический проект, распространяющий микроданные переписей разных стран, который Целевая группа по обеспечению статистической конфиденциальности и доступа к микроданным назвала на Конференции европейских статистиков «передовой практикой» (2007 г.).

18. То, что проект IPUMS-International предоставляет пользователям по электронным каналам микроданные в виде выполненных по запросу массивов, учитывающих страну (страны), год (годы) переписи, группу (группы) населения и конкретные показатели в соответствии с их индивидуальными потребностями,

дополнительно стимулирует их к ревностной защите таких микроданных. А то, что полные наборы данных не распространяются на компакт-дисках или каких-либо иных носителях, значительно понижает соблазн поделиться ими с не имеющими на это полномочий пользователями. Среднее извлечение микроданных в 2011 году содержало всего 35 показателей, учитывая 6 технических переменных, автоматически включаемых в каждое извлечение.

19. Такая практика отличается от деятельности большинства статистических служб, распространяющих микроданные переписей в одном массиве, содержащем словарь данных и всю выборку со всеми переменными и персональными записями. Обычно при использовании прежних методик все исследователи получали по своим запросам совершенно одинаковые наборы данных и документации. Учитывая огромный размер базы данных IPUMS-International, распространение полных наборов переменных и выборок постоянной величины практически нецелесообразно.

20. Несмотря на строгие условия и ограничения более пяти тысяч исследователей из более чем 100 стран и 900 учреждений получили доступ к базе данных IPUMS-I. Более трети пользователей запросили микроданные по одной стране. В основном, это люди, постоянно проживающие за границей и желающие получить доступ к данным своей страны.

### **III. Интеграция**

#### **D. Полные исходные метаданные**

21. У исследователей должен быть доступ к оригинальной документации переписи на национальном языке. Для эффективного использования микроданных необходим хотя бы минимальный набор таких документов, как переписные листы, инструкции или учебные пособия по проведению переписи и кодовые книги. Дополнительные метаданные, касающиеся организации, подготовки и проведения переписи, тоже представляют интерес для проекта IPUMS-International и поэтому включаются в каталоги и архивы вместе со всей другой полученной документацией. Для публикуемых метаданных предпочтительней оригиналы документов в печатном виде или в формате PDF. Наша цель заключается в обеспечении архива высококачественных файлов в формате PDF по всем видам метаданных, относящихся к микроданным переписей.

#### **E. Интегрированные метаданные, соответствующие DDI**

22. Интегрированные соответствующие DDI метаданные всех стран и периодов. В целях облегчения процесса исследования интегрируются микроданные всех переписей и стран, включая подробные описания каждой переписи, каждой выборки и каждой переменной. Интегрированные метаданные на английском языке составлены на основании полной оригинальной исходной документации после подробного анализа микроданных. Прежде чем интегрированные микроданные признаются пригодными для представления их исследователям, сотрудники МЦН затрачивают тысячи часов на их анализ, обсуждение, планирование, первоначальные и повторные



проверки. Этот процесс повторяется ежегодно при вводе в базу данных IPUMS новых выборок переписей.

- a) **Переписи и выборки.** В метаданных IPUMS содержатся краткие описания каждой находящейся в базе данных переписи с указанием ее названия, года, генеральной совокупности, населения де-юре/де-факто, единицы учета, официального дня проведения переписи, переписных листов, периода сбора данных, типа, респондента и оценок неполного учета, если таковой имеет место. Описание каждой выборки включает источник, план выборки, единицу выборки, долю выборки, количество персональных записей, веса выборки, жилища или жилые единицы, свободные жилые помещения, домашние хозяйства, коллективные жилища и особые группы населения.
- b) **Описания переменных, исходные тексты и коды.** В метаданных IPUMS дается определение каждой интегрированной переменной и предлагаются основные характеристики: наличие в переписи, совокупность переменной или вопроса, коды, исходный текст (текст переписи) и использованные в интеграции негармонизированные переменные. Доступ к этой информации осуществляется через активируемый мышью гипертекст на веб-сайте IPUMS. При проведении сопоставлений выделяются наиболее важные сходства и расхождения в определениях каждой переменной, включая информацию по стране или переписи, указывающую на отклонения от стандартной практики. Цель таких сопоставлений заключается в выделении важных различий. Если щелкнуть мышью по кнопке «Текст переписи», откроются исходные вопросы и сопутствующие инструкции на английском языке по каждой переписи. Дополнительный поиск приведет к изображениям оригинальных документов, дающим исследователям возможность ознакомиться с макетом документа и его подлинной редакцией на национальном языке.
- c) **Пример: «Уровень образования».** В качестве примера используемого проектом IPUMS метода интеграции переменных рассмотрим концепцию «уровень образования» как наиболее широко используемую переменную в базе данных IPUMS-International. В таблице 1 приведена подробная трехразрядная схема кодирования признака уровня образования по 12 европейским странам (код ИСО 3166) и показан год проведения переписи (век опущен). Большинство микроданных переписи, содержащих информацию по данной переменной, указывают, получил ли респондент начальное, среднее или высшее образование или вообще не посещал школу. Таким образом, первый разряд сложного кода IPUMS-International состоит из четырех категорий (1-4), плюс кода отсутствующих данных (9) и нуля («за пределами совокупности») для малолетних детей, не посещающих школу, и прочих лиц, которым этот вопрос не задавался. Некоторые выборки содержат дополнительную информацию, отмечающую, например, людей, посетивших начальную, среднюю или даже высшую школу, но не прошедших полного курса обучения. Эта информация содержится во втором разряде. В третьем разряде делается различие между

техническим и общим образованием или другими параметрами, характерными для двух или более стран.

#### **Ф. Интегрированные объединенные микроданные**

23. Проект IPUMS предлагает исследователям и ЦСУ важное преимущество, заключающееся в интеграции полного ряда выборок микроданных по каждой стране, обычно начинающегося с самой первой переписи с существующими или поддающимися восстановлению микроданными. Проект IPUMS объединяет микроданные по каждому индивидуальному запросу в единый файл. Метаданные составляются с учетом требований исследователя и предоставляются в виде единого, доступного для поиска файла в формате DDI или текстовом формате. Благодаря этому исследователь получает возможность анализировать переписи разных лет и даже разных стран в одном файле данных, что упрощает проведение сопоставительного анализа во времени и пространстве.

#### **Г. Интегрированные файлы границ**

24. Проект IPUMS-International предоставляет файлы границ (шейп-файлы), облегчающие отображение национальных и международных данных. Предлагаются файлы границ ГИС двух уровней: файл мира с границами стран и файлы стран с границами крупных административных единиц (штаты, провинции и т.д.). Пользователи могут на основании данных IPUMS-International создавать карты с помощью статистического программного обеспечения и картографической программы ГИС (ArcMap). Более подробную информацию о подключении данных IPUMS-International к картам мира и конкретных стран можно найти на: <https://international.ipums.org/international/gis.shtml>.

#### **Н. Переменные IPUMS-I с дополнительными возможностями**

25. Коллектив проекта IPUMS обладает многолетним опытом работы с микроданными переписей. Пользуясь приобретенными навыками, мы разработали для удобства исследователей более тридцати переменных, дающих им дополнительные возможности, которые распределили по трем группам: технические, сводные и переменные-указатели.

- a) **Технические переменные:** тип записи, страна, год, идентификатор выборки IPUMS, порядковый номер домашнего хозяйства, количество персональных записей в домашнем хозяйстве, вес домашнего хозяйства, номер подвыборки, статус коллективного жилища, континент, регион страны, местожительство на первом административном уровне, коэффициенты расширения (веса выборки для домашних хозяйств и лиц) и т.д.
- b) **Сводные переменные домашнего хозяйства и семьи:** классификация домашних хозяйств, количество семей в домашнем хозяйстве, количество супружеских пар в домашнем хозяйстве, количество матерей в домашнем хозяйстве, количество отцов в домашнем хозяйстве, положение главы в домашнем хозяйстве, количество лиц, не связанных родством, принадлежность к

семейной ячейке, количество членов собственной семьи в домашнем хозяйстве, количество собственных детей в домашнем хозяйстве, количество собственных детей до 5 лет в домашнем хозяйстве, возраст своего старшего ребенка в домашнем хозяйстве, возраст своего младшего ребенка в домашнем хозяйстве, т.д.

- с) **Переменные-указатели** для обозначения совместно проживающих супругов, детей и их родителей: положение матери, отца и супруга в домашнем хозяйстве, правила определения связи родителя (родителей) и супруга (супругов), предполагаемая неродная мать, предполагаемый неродной отец, мужчина с 2+ женами, вторая и последующая жена, т.д. (Sobek and Kennedy, 2009).

#### IV. Управление трансграничным доступом к микроданным переписей

##### I. Трансграничный доступ

26. В современном мире с легко переезжающими с места на место учеными трансграничный доступ к микроданным крайне необходим. Возьмем для примера демографию, где одна пятая мирового профессионального сообщества, т.е. Международный союз по научным исследованиям в области народонаселения (ИЮССП), проживает за пределами своей страны. Из 506 сотрудников ИЮССП, проживающих в США, 30% родились не в США. Почти одна треть китайских демографов живет за пределами Китая. Доля немецких и голландских сотрудников ИЮССП доходит до двух пятых<sup>2</sup>. Для многих профессиональных демографов (а также социологов и политологов), если они намерены проводить сопоставительные или несопоставительные исследования на микроданных переписей своей страны рождения, трансграничный доступ имеет большое значение. Портал IPUMS-International предлагает одинаковые возможности для доступа к выборкам всех стран и переписей, независимо от страны рождения, места проживания или национальности исследователя.

##### J. Извлечение данных по индивидуальным запросам

27. В проекте IPUMS-I не бывает двух одинаковых извлеченных массивов данных. Каждое извлечение выполняется в соответствии с требованиями исследователя. Отбор параметров осуществляется на экране методом «укажи и выбери». Для облегчения процесса выбора можно пользоваться метаданными, просматривая документы в любой нужной последовательности и не прерывая процесса извлечения. Чтобы заказать необходимые данные, исследователь выбирает: страну (или страны), год (годы) переписи, переменные (возраст, пол, образование, т.д.), подвыборку (например, главы домашнего хозяйства женского пола в возрасте до 25 лет) и плотность выборки (в процентном или количественном выражении).

28. Блок извлечения IPUMS выполняет запрос, генерируя единый объединенный массив данных, содержащий микроданные, набор сопутствующих

<sup>2</sup> Данные предоставлены автору секретариатом Международного союза по научным исследованиям в области народонаселения 14 сентября 2011 г.

соответствующих DDI метаданных, а также кодовые книги для системных файлов SPSS, SAS или STATA. На веб-сайте можно получить доступ к копиям исходных метаданных, а также к интегрированным метаданным в интерактивной форме. Их можно загрузить с сайта бесплатно.

29. То, что проект IPUMS-International предоставляет пользователям по электронным каналам микроданные в виде выполненных по запросу массивов, учитывающих страну (страны), год (годы) переписи, группу (группы) населения и конкретные переменные в соответствии с их индивидуальными потребностями, дополнительно стимулирует их к ревностной защите таких извлечений микроданных. А то, что полные наборы данных не распространяются на компакт-дисках или каких-либо иных носителях, значительно снижает, если не устраняет полностью, соблазн поделиться ими с не имеющими полномочий пользователями.

30. Блок извлечения IPUMS добавляет каждому массиву еще больше возможностей за счет трех уникальных инструментов: выбрать случаи (фильтрует массив для точного отбора конкретных случаев), присоединить характеристики (присоединяет любую переменную отца, матери, супруга или главы домашнего хозяйства к записи соответствующего лица), адаптировать размер выборки (уменьшить выборку до любого процентного значения или абсолютного числа домашних хозяйств или лиц).

## **К. Использование**

31. Проект IPUMS-International освобождает микроданные из застенков статистических органов и предоставляет статистикам, демографам, экономистам, социологам и политикам возможность безо всяких затрат использовать необыкновенную ценность микроданных (Meurer, McСаа, and Lam, 2011). Многие исследователи были лишены доступа к микроданным своих стран до тех пор, пока проект IPUMS-International не распахнул им двери.

32. В 2011 году с веб-сайта IPUMS-International было сделано 8 048 извлечений, в целом охватывающих 40 142 выборки и 281 640 переменных. Среднее число извлечений на страну составило почти 150 выборок по 55 странам, представленным в базе данных за весь год (таблица 2). Правда, интерес к данным разных стран сильно колеблется. Так, по переписи Палестины 1997 г. было сделано наименьшее число запросов, а именно 127. Больше всего извлечений (712) пришлось на выборку из переписи Бразилии 2000 года.

33. Из 32 самых запрашиваемых переменных первые восемь включают четыре демографических признака (семейное положение, отношение к главе домашнего хозяйства, возраст и пол), два экономических (статус занятости и категория работника), один социальный (уровень образования) и один технический (весовой коэффициент лица).

34. Наиболее удивительным оказался тот факт, что 60% из запрашиваемых показателей составляют 40 переменных, среди которых 16 созданы в рамках проекта

IPUMS, а четыре из них, так называемые, переменные «LOC», являются уникальными. Они выводятся из сведений о домашнем хозяйстве: положение супруга в домашнем хозяйстве, положение матери, положение отца и правила определения таких связей. Ученые используют эти переменные для изучения общих характеристик супругов и характеристик родителей по отношению к их детям. Столь интенсивное использование переменных «LOC» говорит об их значении для анализа отдельных личностей в зависимости от характеристик их супругов, матерей и отцов, что значительно повышает возможности каждой выборки домашних хозяйств (Sobek and Kennedy, 2009).

35. В число 40 наиболее часто использующих данные проекта IPUMS-I учреждений входят многие ведущие университеты и научные организации, сконцентрированные в 14 странах мира. В 46 странах мы обнаружили в общей сложности 501 учреждение, сотрудники которых запросили по десять или более извлечений (при этом в США 295 таких учреждений). На удивление большое число извлечений было сделано исследователями из стран, не участвующих в проекте IPUMS-I. Десять первых из них: Сингапур (494 извлечения), Бельгия (250), Австралия (229), Япония (170), Российская Федерация (58), Республика Корея (45), Чешская Республика (42), Швеция (41), САР Гонконг (40) и Новая Зеландия (40).

#### **L. Выборки переписей раунда 2010 года**

37. Обработку микроданных переписей населения раунда 2010 года проект IPUMS-International осуществляет в кратчайшие сроки. Микроданные, полученные до 1 сентября, обрабатываются в течение девяти месяцев и интегрируются в базу к 1 июня следующего года. Выборки переписей населения Индонезии и Мексики 2010 года были обработаны к 1 июня 2012 года, т.е. всего лишь через два года после проведения переписи. На сегодняшний день все участники проекта IPUMS-International стараются незамедлительно направлять проекту свои микроданные по мере их готовности.

#### **V. Заключение**

38. Проект IPUMS предлагает значительные преимущества при минимальном риске. Статистические органы освобождаются от многочисленных обременительных проблем и обязательств по распространению микроданных. С проектом IPUMS-International сейчас сотрудничает большинство статистических органов всего мира. 96 из них представляют страны, входящие в 157 государств с более чем миллионным населением (плюс две страны с меньшим населением – Острова Фиджи и Сент-Люсия). Изолированные статистические органы, эпизодически распространяющие свои микроданные, подвергаются существенному риску, а также несут значительные трудовые затраты. А в награду получают лишь ограниченное число пользователей. Проект IPUMS предлагает значительную экономию за счет эффекта масштаба при соблюдении самых высоких норм безопасности.

39. Приглашаем статистические органы, не участвующие в проекте IPUMS, рассмотреть возможность присоединения к нему. Статистические органы,

сотрудничающие с проектом, просим направлять свои метаданные и микроданные переписей раунда 2010 года, как только представится возможность.

## Справочная литература

Alexander, J.T.; Davern, M.; and Stevenson, B. 2010. "Inaccurate Age and Sex Data in the [United States] Census PUMS Files: Evidence and Implications," *Public Opinion Quarterly*, 10 (Aug 10), pp. 1-10. doi: 10.1093/poq/nfq033

Conference of European Statisticians. 2007. "Annex 1.23 Case study: Access to anonymized census microdata samples via the IPUMS-International and the Integrated European Census Microdata websites," *Managing Statistical Confidentiality and Microdata Access: Principles and Guidelines on Good Practice*. Geneva: United Nations Economic Commission for Europe. See online edition: <http://www.unece.org/stats/publications/> pp. 98-104.

Esteve, A., J. Garcia and R. McCaa. 2011. "Comparative perspectives on Marriage and International Migration, 1970-2000: findings from IPUMS-International census microdata samples," *Seminar on Global Perspectives on Marriage and International Migration*, Seoul, South Korea: IUSSP Scientific Panel, Oct. 20-21.

Frend, J., C. Abrahams, A. Forbes, P. Groom, K. Spicer, C. Tudor and P. Youens. 2012. "Statistical Disclosure Control in the 2011 UK Census: Swapping Certainty for Safety." *ESSnet Workshop on Statistical Disclosure Control of Census Data*, Luxembourg, Apr. 19-20.

McCaa, R. and A. Esteve. 2009. "[Entrusting census microdata and metadata for timely integration and dissemination via the IPUMS-EurAsia and IECM initiatives, 2010-2014](#)," *Census Outputs to Meet User Needs*. Geneva: UNECE, Oct. 28-30.

McCaa, R., S. Ruggles and M. Sobek. 2010. "[IPUMS-International statistical disclosure controls: 159 census microdata samples in dissemination, 100+ in preparation](#)," in J. Domingo-Ferrer and E. Magkos (Eds.): *Privacy in Statistical Data 2010*, LNCS 6344. Springer, Heidelberg, pp.74-84.

McCaa, R. and W. Thomas. 2009. "[IPUMS-International: lessons from 10 years of archiving and disseminating census microdata](#)," *International Statistical Institute IPM100*. Durban, South Africa.

Minnesota Population Center. 2011. *Integrated Public Use Microdata Series – International: Version 6.0*. Minneapolis: University of Minnesota: <https://www.ipums.org/international>.

Meier, A., R. McCaa and D. Lam. 2011. "[Creating statistically literate global citizens: The use of IPUMS-International integrated census microdata in teaching](#)". *Statistical Journal of the IAOS* 27(3):145-156.

[Parker, J.] 2011. "Herr and Madame, Señor and Mrs. Research at last begins to cast some light on the extent, causes and consequences of cross-border marriages," *Economist*, Nov. 12. <http://www.economist.com/node/21538103>

Sobek, M and S. Kennedy. 2009 The development of family interrelationship variables for international census data, Minnesota Population Center. [https://international.ipums.org/international/resources/misc\\_docs/pointer\\_working\\_paper\\_2009.pdf](https://international.ipums.org/international/resources/misc_docs/pointer_working_paper_2009.pdf).

Trewin, D. 2007. "A Review of IPUMS-International." Unpub. [http://www.hist.umn.edu/~rmccaa/IPUMSI/trewin\\_ipums\\_report.pdf](http://www.hist.umn.edu/~rmccaa/IPUMSI/trewin_ipums_report.pdf)

**Таблица 1. Уровень образования: гармонизированные коды IPUMS-I  
последних выборов 12 европейских стран**

“X” означает наличие кода в соответствующей выборке

Код	Значение	Страна	BY	FR	DE	GR	HU	IE	IT	PT	RO	SI	ES	CH
		Перепись	99	06	87	01	01	06	01	01	02	02	01	00
0	NIU (за пределами совокупности)		X	X	X	X	X	X	X	·	X	X	X	X
100	НИЖЕ ЗАКОНЧЕННОГО НАЧАЛЬНОГО		X	·	·	X	X	·	X	X	·	X	·	·
110	Не посещал школу			X	·	·	·	·	·	·	X	·	X	X
120	Незаконченное начальное			X	·	X	X	·	·	·	X	·	·	·
130	Начальное (4 года)			·	·	·	·	·	·	X	X	·	·	·
ЗАКОНЧЕННОЕ НАЧАЛЬНОЕ, НИЖЕ СРЕДНЕГО														
Законченное начальное														
211	Начальное (5 лет)		X	·	·	·	·	·	X	·	·	·	X	·
212	Начальное (6 лет)			X	X	X	X	X	·	X	·	·	·	·
Закончена первая ступень среднего образования														
221	Общее и с неустановленным профилем			X	X	X	·	X	X	·	X	X	X	·
222	С техническим профилем			X	·	·	·	·	·	·	·	·	·	·
ЗАКОНЧЕННОЕ СРЕДНЕЕ														
Общее или с неустановленным профилем														
311	Законченное общее		X	X	X	X	X	X	X	X	X	X	X	X
312	Незаконченное высшее			·	·	·	·	·	X	X	X	·	·	·
320	Технический профиль		X	·	·	·	·	·	·	·	·	·	·	·
321	Среднее техническое			·	X	X	X	·	X	·	X	X	X	X
322	Послесреднее техническое			·	X	X	·	X	X	·	X	X	X	X
400	ЗАКОНЧЕННОЕ ВЫСШЕЕ		X	X	X	X	X	X	X	X	X	X	X	X
999	НЕЗВЕСТНО/ДААННЫЕ ОТСУТСТВУЮТ		X	·	X	·	·	X	·	·	X	X	·	X

<b>Таблица 2. Распределение стран (5 первых, все европейские, плюс Канада и США) по числу извлечений из переписей раунда 2000 года (календарный год 2011)</b>					
Место	Страна	Выборка %*	Переменные (к-во)*	Год переписи	Извлечения
1	Бразилия	5	106	1960, 70, 80, 91, 2000	712
2	Мексика	10	120	1960р, 70, 90, 95, 2000, 05	626
3	США	5	92	1960, 70, 80, 90, 2000, 05	554
4	Колумбия	10	120	1964р, 72, 85, 93, 2005	516
5	Южная Африка	10	108	1996, 2001, 2007	428
7	Канада	2.5	59	1971р, 81р, 91р, 2001р	409
9	Франция	33	94	1962, 68, 75, 82, 90, 99, 06	380
10	Испания	5	99	1981, 91, 2001	366
13	Греция	10	89	1971, 81, 91, 2001	327
18	Австрия	10	75	1971, 81, 91, 2001	310
25	Италия	5	81	2001	285
26	Португалия	5	96	1981, 91, 2001	283
29	Румыния	10	97	1976, 92, 2002	272
30	Швейцария	5	79	1970, 80, 90, 2000	266
32	Великобритания	3	47	1991, 2001р	263
38	Венгрия	5	74	1970, 80, 90, 2001	222
42	Нидерланды	1	33	1960р, 71р, 2001р	211
45	Словения	10	80	2002	185
48	Беларусь	10	84	1999	179
Всего извлечений по 55 странам (162 выборки) на 1 января 2001 г.					8 048
*перепись раунда 2000 г.; относится ко всем интегрированным переменным, включая переменные, созданные в рамках проекта IPUMS.					
“р” = выборка физических лиц; все другие выборки касаются домашних хозяйств					