

Конференция МАОС: «Пути распространения информации: Стратегии и лучшие практики обеспечения пользователей статистикой для принятия решений»

12-14 сентября 2012 г., Киев, Украина <http://iaos2012.ukrstat.gov.ua/>

Сессия докладов: Расширение доступа к информации

Название: «10 способов повышения эффективности использования микроданных переписей населения в проекте IPUMS-International»¹

Авторы:

Роберт МакКей, Миннесотский центр народонаселения, США

Лара Кливленд, Миннесотский центр народонаселения, США

Альберт Эстив, Центр демографических исследований, Автономный университет Барселоны, Испания

Стивен Раглз, Миннесотский центр народонаселения, США

Мэтью Собек, Миннесотский центр народонаселения, США

«Распространять данные – [значит] открывать их значение всему миру»

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

Аннотация. Многие статистические управления признают необходимость расширения доступа к микроданным переписей населения. Значительным препятствием для осуществления этой задачи «своими силами» являются высокие затраты, неоправданные риски и слишком незначительная отдача. Проект IPUMS-International предлагает экономичное, практически бесплатное решение, получившее поддержку более 90 национальных статистических управлений. В докладе рассматриваются 10 способов, используемых проектом для расширения доступа к микроданным переписей населения и повышения их полезности, которые разделены на четыре категории: 1. Статистическая конфиденциальность (безопасность, защита от косвенного раскрытия, управление доступом); 2. Интеграция (полная исходная документация, интегрированные метаданные, интегрированные объединенные микроданные, переменные, разработанные в рамках проекта IPUMS-I); 3. Распространение (бесплатный трансграничный доступ, выполненные по индивидуальным запросам извлечения); и 4. Этические принципы (статистическая прозрачность, академическая свобода, снижение рисков фальсификации/искажения данных и обмен результатами научных исследований). Предлагаем статистическим управлениям, не участвующим в проекте IPUMS-International, начать с ним сотрудничать. А участников проекта призываем к безотлагательному представлению микроданных переписей раунда 2010 года.

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

ВВЕДЕНИЕ

Представителям науки и политическим деятелям необходим доступ к микроданным переписей населения для анализа сложных социальных, демографических и экономических преобразований (McCa and Esteve 2009). В наше время в большинстве современных статистических служб понимают, насколько велико значение распространения микроданных переписей. И, тем не менее, некоторые из них не проявляют энтузиазма по поводу такой деятельности, поскольку обеспечение открытого доступа - занятие рискованное, дорогостоящее, и часто не оправдывает возлагаемых на него надежд. Передача функции распространения микроданных третьей стороне, например, международному проекту IPUMS-International, устраняет дополнительные затраты, сводит к нулю риски и наилучшим образом удовлетворяет потребности пользователей. Статистическим службам, настаивающим на распространении микроданных «своими силами», следовало бы оценить существенные затраты, а также риски, связанные с потерей профессионального доверия и вероятным недовольством пользователей.

- **Затраты.** Статистическая служба должна составить выборку, обезличить микроданные, подготовить метаданные и обеспечить доступ к информации. Для выполнения этих задач потребуется много усилий, особенно, если в год поступает не более десяти запросов.
- **Риски.** Лишь некоторые статистические службы обладают опытом обезличивания микроданных переписей, да и сложно приобрести такой опыт, если переписи проводятся не чаще одного раза 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. Большинство служб официальной статистики предпочитают не брать на себя ответственность по выполнению трудоемких задач, связанных с проверкой пользователей, контролем доступа к информации и обеспечением пользовательской поддержки. 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 американского обследования общин, переписи населения 2000 г. и текущего обследования населения. В целом, за прошедшие до обнаружения ошибки восемь лет испорченными оказались 144 массива микроданных. Ученые смогли обнаружить ошибки только путем сопоставления обезличенной возрастной структуры населения с полными, не подвергнувшимися обезличиванию микроданными. Этот скандал, случившийся некстати для Бюро переписей незадолго до переписи 2010 года, угодил на первые страницы Нью-Йорк Таймс под броским заголовком «**Можно ли доверять данным переписи?**»². К сожалению, к моменту обнаружения неверно проведенной

² Wolfers, J. Can You Trust Census Data? Freakonomics blog. New York Times, February 2, 2010. <http://freakonomics.blogs.nytimes.com/2010/02/02/can-you-trust-census-data>.

анонимизации ошибка получила широкой распространение. (Cleveland, McCaa, Ruggles, and Sobek 2012). Однако Бюро переписей достойно похвалы за то, что не замалчивало случившегося. Предоставив ученым доступ (под очень строгой защитой) к необезличенным микроданным, эти ошибки удалось, если не исправить до конца, то хотя бы обнаружить.

- **Неудовлетворенность.** Обычно статистические органы допускают ошибку при защите микроданных, прибегая не только к их чрезмерной анонимизации, но и к запрещению доступа всем, кроме самых настойчивых или влиятельных исследователей. Такая стратегия может обеспечить безопасность, но одновременно стимулирует появление контрабандных копий. В этом случае в жертву приносится сама полезность данных, в результате чего неудовлетворенными остаются как статистические ведомства, так и пользующиеся их данными исследователи.

Обратите внимание на жалобы, адресованные одному из ведущих статистических органов с более чем двадцатилетним опытом распространения микроданных переписей:

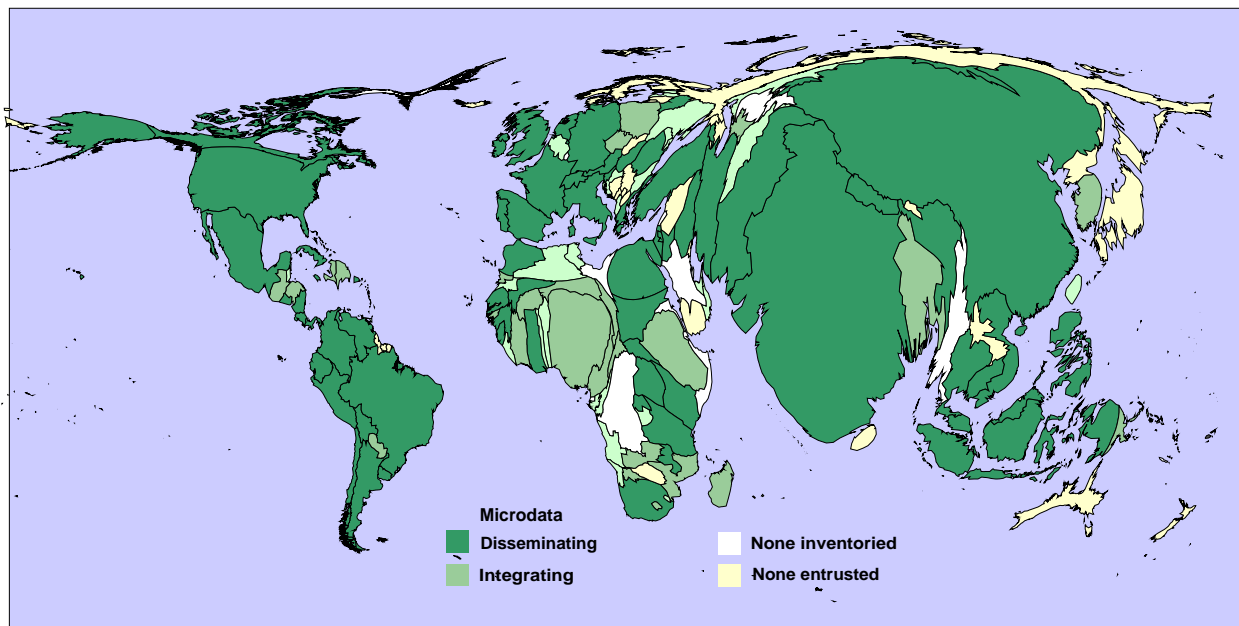
- «Мне не удалось использовать микроданные; длительность прохождения бюрократических процедур сводит на нет практическую пользу получения информации».
- «Невероятно забюрократизированный и сложный доступ отбивает у людей охоту им воспользоваться. Мне для получения данных потребовалось полгода».

Некоторые официальные статистические органы, несмотря на то, что при разработке своих веб-сайтов и стратегии доступа к микроданным они, наверняка, используют передовые технологии, никогда не отвечают на запросы о предоставлении доступа. 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» (Parkinson, 1970).

Проект IPUMS-International (www.ipums.org/international) предлагает комплексное универсальное решение по распространению выборок обезличенных микроданных. В осуществляемом Миннесотским центром народонаселения (МЦН) проекте IPUMS-International участвуют национальные статистические управления (НСУ), университеты и ведущие научно-исследовательские центры всего мира. Региональными партнерами IPUMS-International являются проекты «Интегрированные микроданные европейских переписей» (Центр демографических исследований Автономного университета Барселоны) и «Интегрированные микроданные африканских переписей» (Африканский центр статистики, Аддис-Абеба). Проект предлагает 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 (McCaa 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.

Рис. 1. Картограмма, взвешенная по численности населения стран, участвующих в проекте IPUMS-International.

Июнь 2012 г.: 3 оттенка зеленого: распространение (темный), интеграция (средний), переговоры (светлый).



Начиная с 1999 года, проект IPUMS-International удовлетворил запросы более 90 НСУ и более 5 000 исследователей из почти 100 стран. 21 европейская страна участвует в программе IPUMS на условиях подписания стандартного соглашения о взаимопонимании (Приложение А и рис. 1). Добросовестные исследователи путем подписания стандартного лицензионного соглашения (Приложение В) получают доступ к запрашиваемым микроданным всех стран, участвующих в проекте. По состоянию на июнь 2012 г. проект завершил анонимизацию и интеграцию микроданных населения по 69 странам, что в целом охватывает 212 выборок и 480 миллионов персональных записей. Благодаря чрезвычайно щедрому вкладу НСУ всего мира 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.

Приглашение к участию и представлению данных переписей раунда 2010 года.

Приглашаем статистические органы, не сотрудничающие с проектом на настоящий момент, связаться с автором доклада. Предлагаем сотрудничающим статистическим органам без лишней задержки представить проекту микроданные переписей населения раунда 2010 года, чтобы обеспечить своевременный доступ к анонимизированным, интегрированным выборкам данных. Это значительно увеличит число заинтересованных сторон и, по словам Президента Евростата, «раскроет значение» микроданных переписей населения (см. эпиграф).

36 европейских стран с численностью населения в один миллион и более можно разделить по степени их участия в проекте IPUMS на 4 группы (в скобках указано количество представленных выборок):

- **17 стран интегрированы:** Армения (1), Австрия (4), Беларусь (1), Франция (7), Германия (8, включая ГДР и ФРГ и 4 микропереписи в стадии интеграции), 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 (6; 2 в стадии восстановления).
- **2 страны на стадии передачи микроданных:** Болгария и Украина.
- **16 стран на стадии вступления в сотрудничество:** Албания, Бельгия (ведутся переговоры), Босния и Герцеговина, Хорватия (ведутся переговоры), Дания, Эстония (ведутся переговоры), Финляндия, Грузия, Латвия (ведутся переговоры), Литва, Македония, Молдова, Норвегия, Россия, Словакия (ведутся переговоры) и Швеция.

В 2012 году проект IPUMS интегрировал выборки Турции (1985, 1990 и 2000 гг.), а также шесть выборок из неевропейских стран: Сальвадора, Индонезии, Мексики, Марокко, Никарагуа и Уругвая. Существует план на ежегодную интеграцию выборок в последующие годы, учитывающий интеграцию выборок переписей раунда 2010 года по мере их поступления.

Проект IPUMS-International делает все возможное для расширения доступа к микроданным переписей представителям науки, а также политическим и общественным деятелям. Проект IPUMS-I – это *не* просто передатчик выборочных данных переписей от НСУ к исследователям. Обычно ведущие сотрудники МЦН работают над обезличиванием, интеграцией и обеспечением доступа к выборке микроданных и сопутствующим метаданным в течение двух или более лет, затрачивая на это десятки тысяч рабочих часов каждый.

В 21 веке для серьезного научного исследования *недостаточно* просто получения копии исходных микроданных и таблицы кодирования. В настоящем докладе представлены 10 дополнительных преимуществ, которые предлагают проекты IPUMS-International, IECM и AICM для расширения научного доступа к микроданным переписей. Их можно разделить на четыре группы: статистическая конфиденциальность, интеграция, распространение и этические принципы.

I. Конфиденциальность и безопасность статистических данных.

1. Безопасность микроданных и статистическая конфиденциальность. 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. В соответствии с Федеральным законом 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.

Проект IPUMS-I, возможно, является первым академическим проектом по распространению микроданных переписей населения 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», «эталонном для лучших статистических служб» и т.д. (Тревин, 2007 г.). Г-н Тревин в течение недели проводил в МЦН проверку, в результате которой сделал следующие заключения (Trewin 2007):

Без всякого сомнения проект IPUMS-International соответствует четырем основным принципам, провозглашенным КЕС [Конференция европейских статистиков] (2007 г.). В материалах КЕС (2007) он приведен в качестве примера передовой практики. Настоящее обследование подтверждает его статус передового хранилища данных. Он на самом деле представляет лучший образец хранилища международных статистических данных [выделено автором].

На совещании Группы экспертов ЕЭК ООН по конфиденциальности статистических данных используемые в проекте IPUMS-International процедуры обеспечения безопасности распространяемых данных были охарактеризованы следующим образом (McCaа and Esteve 2005):

Когда извлечение готово (обычно это вопрос нескольких минут), исследователя уведомляют по электронной почте о необходимости извлечения запрошенных данных в течение 72 часов. Для скачивания конкретного извлечения предоставляется ссылка на защищенный паролем сайт. Во время передачи данные шифруются по стандартному 128-разрядному протоколу SSL (протокол защиты информации), который используется в банковской и иных отраслях, придающих большое значение безопасности и конфиденциальности информации. После успешного скачивания файла исследователь может распаковать его и приступить к анализу, используя предоставленные интегрированные метаданные, содержащие названия и обозначения переменных.

Настоящий метод распространения данных все еще выдерживает испытание временем, и, учитывая его возросшее применение и непрерывное увеличение скорости передачи данных по Интернету, использование протоколов безопасности совершенно оправдано.

2. Защита статистических данных от раскрытия. Микроданные подвергаются строгому единому правовому, административному и техническому контролю, обеспечивающему статистическим данным всех участвующих в проекте статистических органов 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 подробных географических признаков.

Все статистические службы стараются найти компромисс между конфиденциальностью и практической пользой и сообщают проекту IPUMS-I свой минимальный порог идентификации географических единиц. В большинстве стран этот порог установлен на уровне 20 000 жителей. В других он равняется 100 000 жителей (США), а в самых экстремальных случаях (Нидерланды) подавляется все административно-территориальное деление. Нас радует, что в некоторых странах наши партнеры пересмотрели свои прежние решения и представили проекту более точные выборки (выборка Мексики переписи 1999 г. выросла с одного до 10 процентов) с более

мелкой детализацией. В случае Колумбии географический порог, ранее установленный на уровне 100 000 жителей, был после настоятельных обращений колумбийских географов в Управление статистики Колумбии (DANE) снижен до 20 000. Управление статистики не только снизило географический порог до 20 000 жителей, но благодаря гармонизации гео-статистических идентификаторов сделала возможным распространение всего ряда выборок микроданных переписей Колумбии с 1964 по 2005 гг. с помощью единого набора географических кодов.

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

По согласованию с национальными статистическими управлениями мы применяем к одним переменным кодирование с установлением верхнего предела, к другим – глобальное перекодирование, избирательно убираем разряды иерархических переменных (профессия, отрасль, географический признак) или даже блокируем все представляющие значительный риск переменные. 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.

3. Ограниченный доступ. 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, обязательное цитирование, возможность применения дисциплинарных мер за нарушения и сообщение об ошибках. Иными словами, проект IPUMS-I предоставляет допуск правомочным пользователям.

Согласие с условиями использования микроданных налагает обязательства как на исследователя, так и на его учреждение. Юридическая служба университета Миннесоты готова отреагировать на первое проявление ненадлежащего использования. Ответственность за обеспечение безопасности и выполнение условий лицензионного соглашения несут как отдельные исследователи, так и их учреждения. Нарушения могут привести к санкциям как против отдельных исследователей, так и против учреждений, в которых они работают. Нарушение правил одним пользователем может приостановить доступ к микроданным всем другим пользователям этого учреждения. Для того, чтобы нарушившее условия учреждение смогло восстановить свои полномочия для работы с

уязвимыми микроданными, от его сотрудников могут потребовать прохождения дополнительной подготовки по защите респондентов.

Используемые в проекте IPUMS-I методы позволяют решить задачу предоставления максимального доступа к уязвимым микроданным при сохранении защиты их конфиденциальности.

Многие статистические службы с удовольствием предоставляли бы исследователям доступ к своим микроданным, но не могут этого сделать из-за отсутствия значительных материальных и человеческих ресурсов, необходимых для создания и эксплуатации безопасных систем.

То, что проект IPUMS-International предоставляет пользователям по электронным каналам микроданные в виде выполненных по запросу извлечений, 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.

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.

Опытный исследователь, независимо от времени, необходимого для заполнения формы, легко согласится с ее условиями и педантично внесет запрашиваемую информацию, а неопытный, столкнувшись с необходимостью указания службы по защите респондентов своего учреждения, своего руководителя и веб-сайта учреждения, а также предоставления описания научного проекта, для которого ему необходимы микроданные, вовсе откажется от заполнения формы (см. Приложение В). Незаполненные формы автоматически отклоняются системой управления сайтом. Направить для рассмотрения незаполненную форму невозможно.

Из 1 057 указанных выше заполненных форм были отклонены только 46. После подачи заявок заявителей тщательно проверяют на предмет их соответствия требованиям или наличие/отсутствие научных потребностей в микроданных. В большинстве отклоненных запросов имеющиеся для распространения микроданные переписей не соответствовали целям предполагаемых исследований. Необходимые для заполнения заявки подробности отпугивают случайных посетителей, и они самостоятельно отсеиваются. Из этого краткого описания процесса регистрации для получения доступа к сайту IPUMS-International очевидно, что каждый отдельный заявитель, прежде чем получить учетную запись, подвергается, в соответствии с эпитафией, тщательной проверке с целью определения его личности и установления его научной добросовестности.

Проект IPUMS предоставляет извлечения, соответствующие индивидуальным запросам и научным потребностям каждого отдельного пользователя. Среднее извлечение микроданных в прошлом году содержало в целом 35 показателей, учитывая 6 технических переменных, автоматически включаемых в каждое извлечение.

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.

Несмотря на строгие условия и ограничения более пяти тысяч исследователей из более 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.

II. Интеграция

4. Полные исходные данные. 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 таблицы кодирования. 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. Если метаданные были опубликованы, более предпочтительным вариантом для проекта являются непосредственно оригиналы документов или их копии в формате PDF. Наша цель заключается в обеспечении архива высококачественных файлов в формате PDF по всем видам метаданных, относящихся к микроданным переписей. Исходные документы предоставляются исследователям как на английском языке, так и на национальном языке оригинала.

5. Интегрированные метаданные, выполненные в соответствии с международным стандартом DDI—по странам и периодам. В целях облегчения процесса исследования интегрируются микроданные всех переписей и стран, включая подробные описания каждой переписи, каждой выборки и каждой переменной.

а. 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 и метода сбора данных, респондента и оценок неполного учета, если таковой имеет место. Предоставляется доступ к образцам переписных листов и инструкций по проведению переписи на национальном языке и с переводом текста на английский язык. 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.

б. 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, текст переписного листа и использованные в процессе 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 переписного листа» (“Questionnaire text”), откроются исходные вопросы и сопутствующие инструкции на английском языке по каждой выбранной переписи. Дополнительный поиск приведет к изображениям оригинальных документов, которые дают исследователям возможность ознакомиться с макетом документа и его подлинной редакцией на национальном языке.

Исследователи могут перемещаться по метаданным для быстрого просмотра вопросов и инструкций по заполнению переписных листов в любой комбинации стран и дат проведения переписей. Осуществляющие интеграцию сотрудники МЦН применяют к документам переписи XML-теги, связывающие переменные в микроданных с концепциями в тексте, после чего отмеченный тегами материал импортируется в базу данных. По завершении этой операции метаданные можно динамически извлекать для любой комбинации стран и дат проведения переписей по каждой отдельной переменной. Первоначально этот инструмент был разработан для внутреннего использования в целях ускорения процесса интеграции. Но когда его аналитические возможности стали очевидными, мы встроили динамическую систему метаданных в веб-сайт для обеспечения общего доступа к метаданным.

6. Интегрированные объединенные микроданные. Проект IPUMS предлагает исследователям и НСУ важное для них преимущество, заключающееся в интеграции полного хронологического ряда выборок микроданных по каждой стране, который обычно начинается с самой первой переписи с существующими или поддающимися восстановлению микроданными и заканчивается переписью раунда 2010 года. Большинство НСУ представляют выборку из самой последней переписи, и только очень немногие из них обращаются к предыдущим переписям с целью гармонизации последующих наборов данных или разработки новой документации, облегчающей сопоставительный анализ двух или более переписей. В лучшем случае, статистические службы просто составляют выборку и снабжают ее словарем данных. Через пять или десять лет при проведении следующей переписи процесс повторяется и сопровождается небольшой инструкцией по сопоставлению новых микроданных с наборами микроданных прежних переписей.

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 и/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.

Следует еще раз напомнить о том, что проект IPUMS не распространяет копии файлов с данными переписей в том виде, в каком они поступили из национальных статистических управлений. Высокоточные выборки переписей проходят анонимизацию (McCaа et. al. 2006), после чего с помощью сложной системы кодирования (Esteve and Sobek, 2003) интегрируются все их переменные. Выборки интегрируются как хронологически, так и по странам. Интегрированные метаданные создаются в результате тщательного изучения исходной оригинальной документации и подробного анализа микроданных. Прежде, чем интегрированные микроданные признаются пригодными для научного использования, на их анализ, обдумывание, обсуждение, планирование, пробное

и повторное тестирование затрачиваются тысячи часов. Этот процесс повторяется ежегодно при вводе в базу данных IPUMS новых выборок переписей.

Главная цель применяемого в проекте IPUMS-I процесса гармонизации состоит в упрощении использования микро- и метаданных без ущерба для их значения. Это трудновыполнимая задача. Для того, чтобы упростить микроданные для проведения сопоставительного анализа во времени и пространстве, необходимо разработать сопоставимые системы кодирования. Микроданные кодируются таким образом, чтобы идентичные концепции (переменные, категории) имели идентичные коды. Во избежание потери важной информации в наиболее подробных выборках проект IPUMS-I использует сложную стратегию кодирования, позволяющую сохранить все первоначальные детали и одновременно обеспечить сопоставительные коды для разных выборок. Пользуясь сложными кодами, исследователи могут очень просто проводить временные и пространственные сопоставления при легко различимых нюансах значения. Первый разряд, который мы называем «общим кодом», дает информацию, характерную для всех выборок (данные наименьшего общего знаменателя). Следующие один или два разряда дают дополнительную информацию, имеющуюся в значительном наборе выборок. Младшие разряды обозначают редко встречающиеся подробности. При отсутствии информации в конкретной выборке в этот разряд ставится ноль.

В качестве примера используемого проектом IPUMS метода интеграции переменных рассмотрим концепцию «уровень образования» как наиболее широко используемую переменную в базе данных IPUMS-International. Большинство микроданных переписей, 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 для отсутствующих данных (9) и нуля («вне совокупности») для малолетних детей, не посещающих школу, и прочих лиц, которым этот вопрос не задавался. Некоторые выборки содержат дополнительную информацию, отмечающую, например, людей, посещавших 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. При успешной международной интеграции такие различия должны подтверждаться документально с тем, чтобы ученые могли свободно получать информацию об этих и тысяче других подробностей.

В Приложении С дается иллюстрация подробной и общей схем кодирования переменной уровня образования по 15 странам, выборки которых были введены в базу данных IPUMS-I в самое последнее время. При нажатии на кнопку «поиск и выбор данных» (“Browse and Select Data”) на домашней странице проекта IPUMS-I можно получить аналогичную таблицу для любой комбинации стран или переписей. Если Вы хотите избежать получения метаданных по всем выборкам, наведите курсор на кнопку «выбрать выборку» (“Select Samples”) и сделайте выбор. Чтобы выбрать переменную уровня образования, наведите мышку на пункт «лицо» (“Person”) и щелкните по рубрике «образование» (“Education”), а также нажмите на кнопку «коды» “Codes” в строке «уровень образования» (“EDATTAN”). Вы получите общие одноразрядные коды переменной «уровень образования, международное перекодирование» по всем выбранным странам и переписям. “X” указывает на наличие соответствующего кода для выбранной

страны или переписи. Нажав на кнопку «подробные коды» (“detailed codes”), Вы получите подробные трехразрядные коды, представленные в Приложении С. Чтобы получить простую невзвешенную встречаемость каждого кода в отобранных выборках, нужно нажать на кнопку «кол-во случаев» (“Case-count view”), как показано в Приложении С. Все метаданные проекта IPUMS-I открыты для всех желающих с ними ознакомиться. Для доступа к ним регистрации не требуется.

7. Переменные IPUMS-I с дополнительными возможностями. 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 можно распределить по трем группам: технические, сводные и переменные-указатели.

а. 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.

б. 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.

в. Переменные-указатели используются для обозначения совместно проживающих супругов (SPLOC) и родителей (MOMLOC, POPLOC) детей: 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). Переменные-указатели значительно облегчают изучение характеристик отдельных людей с учетом совместно проживающих супругов или родителей (см., например, Esteve, Garcia and McCaa, 2011).

III. Распространение данных

8. Трансграничный доступ: 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 в других странах. Почти одна треть китайских демографов живет за пределами Китая. Доля немецких и голландских сотрудников ИЮССП доходит до двух пятых³. Для многих профессиональных

³ Данные предоставлены автору секретариатом Международного союза по научным исследованиям в области народонаселения 14 сентября 2011 г.

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.

Некоторые национальные статистические управления стремятся удовлетворить спрос пользователей на микроданные, предлагая им доступ к интерактивному табулятору. Однако совершенно очевидно, что табуляторы не могут заменить трансграничного доступа к микроданным, поскольку в них нет возможностей для статистического анализа. Лишь в некоторых табуляторах предусмотрены средства для записи или преобразования данных, но ни в одном нет инструментов для построения статистических данных на основании отношений между родителями и детьми, главой домашнего хозяйства и его членами или супругами (см. пункт 7с выше).

9. Извлечение данных по индивидуальным запросам. 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 (см. пункт 6 выше). Чтобы заказать необходимые для построения извлечения данные, исследователь выбирает:

- страну (или страны);
- год (годы) переписи;
- переменные (возраст, пол, образование, т.д.);
- подгруппы населения (например, главы домашних хозяйств женского пола в возрасте до 25 лет. Обратите внимание на существование возможности выбора как отдельных, так и совместно проживающих лиц, учтенных в выбранном домашнем хозяйстве); и
- плотность выборки (в процентном или количественном выражении).

Блок извлечения IPUMS выполняет запрос, генерируя единый объединенный массив данных, содержащий микроданные, сопутствующий им набор соответствующих стандарту DDI метаданных, а также системные файлы SPSS, SAS или STATA, по усмотрению пользователя, вместе с соответствующими таблицами кодирования. С веб-сайта можно скачать копии исходных метаданных, а также получить доступ к интегрированным метаданным в интерактивной форме. Их тоже можно бесплатно скачать с сайта.

То, что проект IPUMS-International предоставляет пользователям по электронным каналам микроданные в виде выполненных по запросу массивов, учитывающих страну (страны), год (годы) переписи, подгруппу (подгруппы) 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 неправомочными пользователями.

Блок извлечения IPUMS добавляет каждому массиву еще больше возможностей за счет трех уникальных инструментов:

а. Выбрать случаи (см. «подгруппы населения» выше). Фильтрует выборки для точного отбора случаев, представляющих научный интерес для пользователя. Например, если ученый намеревается исследовать только экономически активных женщин в возрасте от 15 до 19 лет и рожденных не в стране их нынешнего проживания, блок извлечения IPUMS создаст набор данных точно по этим характеристикам. Более того, пользователь одним нажатием на кнопку может заставить блок извлечения включить в этот набор всех лиц, проживающих в домашнем хозяйстве вместе с выбранным лицом.

б. Присоединить характеристики. При простом нажатии на кнопку “change” («изменить») блок извлечения присоединяет переменные матерей, отцов, супругов и глав домашних хозяйств к совместно проживающим с ними лицам. Эта функция облегчает анализ детей по характеристикам их матерей, отцов и/или глав домашних хозяйств. Она полезна при анализе рождаемости собственных детей, супружеской гомогамии, а также для решения множества других проблем, требующих совместных характеристик двух или более членов домашнего хозяйства. Например, на веб-сайте IPUMS-I был очень просто создан набор данных для недавно прошедшего международного семинара по межнациональным бракам (Сеул, Республика Корея, октябрь 2011 г.), включающий 51 страну и представляющий 12 миллионов граждан иностранного происхождения, находящихся в законном браке или добровольном союзе с коренными супругами (Esteve, Garcia and McCaa, 2011; см также Parker, 2011).

с. Адаптировать размер выборки. Система извлечения предоставляет возможность адаптировать размер каждой выборки до любого абсолютного числа или процентного значения лиц или домашних хозяйств. Если исследователю необходима выборка, ограниченная 50 000 домашних хозяйств, он просто вводит “50” в соответствующую таблицу, и блок извлечения создает систематическую выборку нужного размера. Соответствующие веса (коэффициенты расширения) автоматически рассчитываются по ходу операции и включаются в извлечение.

д. Использование. Проект IPUMS-International освобождает микроданные из застенков статистических органов и предоставляет статистикам, демографам, экономистам, социологам и политикам возможность безо всяких затрат использовать заложенные в них сокровища (Meyer, McCaa, and Lam, 2011). Многие исследователи были лишены доступа к микроданным своих стран до тех пор, пока проект IPUMS-International не распахнул им двери. Другим мешали длительные задержки, чрезмерные цены и пр.

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.

В таблице 3 представлены 32 наиболее часто извлекаемые переменные. Первые восемь из них 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). Первая наиболее часто запрашиваемая переменная извлекалась в шесть раз чаще 32^{ой} переменной. Это говорит о том, что ученые не просто запрашивают подряд все переменные выборки, а на самом деле тщательно конструируют свои извлечения. В сущности, им приходится учитывать огромные размеры выборок и требование выбрать из меню необходимые им переменные.

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, переменные-указатели, или, 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. Исследователи используют эти переменные для изучения общих характеристик супругов и характеристик родителей по отношению к их детям. Эти переменные условно рассчитываются на основании отношения к главе домашнего хозяйства, а также возраста, пола, семейного положения, порядка расположения отдельных лиц в домашнем хозяйстве и нескольких других переменных. Столь интенсивное использование переменных «LOC» говорит об их значении для анализа отдельных личностей в зависимости от характеристик их супругов, матерей и отцов. Использование переменных «LOC» значительно расширяет возможности каждой выборки домашних хозяйств в проекте IPUMS-I (Sobek and Kennedy, 2009).

В число 33 наиболее часто использующих данные проекта IPUMS-I учреждений входят многие ведущие университеты и научные организации (см. Приложение D), расположенные 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).

С другой стороны, существуют страны, интегрировавшие свои микроданные в базу данных IPUMS-I, исследователи в которых не проявляют к ним интереса. Это следующие 12 стран: Армения, Беларусь, Гвинея, Ирак, Иордания, Кыргызстан, Мали, Монголия, Непал, Руанда, Сент-Люсия и Словения. Следует обратить внимание на то, что исследователи из этих стран вместо получения доступа к микроданным через веб-сайт IPUMS-I могут приобрести копии интегрированных выборок на компакт-дисках, направляемых Миннесотским центром народонаселения в соответствующие национальные статистические управления. Мы рекомендуем НСУ регистрировать всех таких пользователей и предупреждать их о необходимости соблюдения условий IPUMS-I по использованию данных, хотя они не обязаны это делать. Ответственность за такое распространение данных берет на себя НСУ.

IV. Расширение этических рамок доступа к микроданным переписей

10. Расширение этических рамок доступа к микроданным через статистическую прозрачность, академическую свободу, снижение фальсификации научных данных и облегчение доступа к результатам исследований.

10а. Статистическая прозрачность. НСУ, предоставляя проекту IPUMS-International микроданные своих переписей, тем самым демонстрирует всему миру отсутствие желания что-либо скрывать и согласие подвергнуть свою самую ценную информацию еще более строгому изучению со стороны.

10б. Академическая свобода. Микроданные, распространяемые проектом IPUMS-International, не подвергаются никакому предварительному запрету на опубликование или распространение результатов, кроме требований соблюдения статистической конфиденциальности. НСУ, ограничивающие доступ к данным, не должны удивляться тому, что у исследователей, желающих снискать благосклонность и получить привилегированный доступ, может появиться желание несколько «подправить» полученные результаты. Принципиальные ученые сторонятся таких фальсифицированных данных не только из-за боязни оказаться в числе «прирученных», но, что более важно, в целях сохранения чистоты науки и снижения фальсификаций.

10с. Снижение фальсификаций или соблазна преувеличения результатов исследования. Поскольку все исследователи имеют равный доступ к распространяемым проектом IPUMS-International данным, результаты легко поддаются восстановлению для сравнения с результатами, полученными другими исследователями. В проекте IPUMS ведется реестр, поэтому в случае возникновения сомнений в достоверности полученных результатов извлечение можно восстановить и отдать на проверку третьей стороне. Такая возможность является хорошим стимулом для поддержания высокого уровня проводимого анализа и, если не полного устранения, то хотя бы снижения фальсификации результатов исследований. Эти проблемы вызывают серьезную озабоченность Национального научного фонда и Национального института здравоохранения США.

10д. Результаты исследований. Исследователи размещают ссылки на полученные ими результаты в библиографическом указателе IPUMS с открытым доступом: <http://bibliography.ipums.org/>. Указатель в настоящее время насчитывает более 500 единиц с меткой IPUMS-International.

Заключение

Проекту IPUMS-International для удовлетворения потребностей научного сообщества и политических деятелей на мировом уровне необходимо огромное количество разнообразных мета- и микроданных. Помимо значительной отдачи от инвестиций проект освобождает статистические органы от многочисленных обременительных проблем и обязательств по распространению микроданных. Более того, передавая ответственность по предоставлению доступа к данным проекту IPUMS-International, национальные статистические управления обеспечивают им защиту. С проектом IPUMS-International сейчас сотрудничает большинство статистических служб 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 роста масштабов при соблюдении самых высоких норм безопасности.

Когда мы десять лет назад приступили к осуществлению проекта, мы мечтали интегрировать в течение последующих десяти лет выборки из 21 страны. Благодаря плодотворному сотрудничеству национальных статистических управлений и появлению ранее невообразимых технических инноваций количество стран с интегрированными микроданными увеличилось более чем втрое и достигло 69. И работа по интеграции все еще продолжается. Количество пользователей и масштаб использования микроданных тоже далеко превзошли наши ожидания.

В последующее десятилетие мы мечтаем удвоить число пользователей и количество выборок. Для успешного решения этой задачи первостепенное значение имеют высокоточные выборки переписей раунда 2010 года. Микроданные переписей раунда 2010 года проходят интеграцию безо всякой задержки. Микроданные, полученные до 1 сентября, обрабатываются в течение девяти месяцев, т.е. к 1 июня следующего года. На настоящий момент все участники проекта IPUMS-International стремятся представлять микроданные переписи раунда 2010 года по мере их готовности. Выборки переписей Индонезии и Мексики 2010 года были интегрированы в июне 2012 года, т.е. меньше чем через два года после проведения переписи.

Предлагаем статистическим органам, сотрудничающим с проектом, направлять свои метаданные и микроданные переписей раунда 2010 года, как только они будут обработаны. Приглашаем статистические органы, не участвующие в проекте IPUMS, рассмотреть возможность присоединения к нему. Ученым, обращающимся к микроданным проекта IPUMS, советуем внимательно изучать метаданные и использовать микроданные в соответствии с их научными потребностями.

Список литературы.

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

Cleveland, L; McCaa, R.; Ruggles, S.; and Sobek, M. 2012. "When Excessive Perturbation Goes Wrong and Why IPUMS-International Relies Instead on Sampling, Suppression, Swapping, and Other Minimally Harmful Methods to Protect Privacy of Census Microdata," in J. Domingo-Ferrer and E. Magkos (Eds.): *Privacy in Statistical Databases 2012*. Springer, Heidelberg.

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. 2005. "[IPUMS-Europe: Confidentiality measures for licensing and disseminating restricted access census microdata extracts to academic users](#)," *Joint UNECE/Eurostat Work Session on Statistical Confidentiality*, Geneva, Nov. 9-11.

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: United Nations Economic Commission for Europe, 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 Databases 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>

Parkinson, C. Northcote. 1970 *The Law of Delay: Interviews and Outerviews*. London: John Murray.

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 .

Trewin, D. 2007. "A Review of IPUMS-International." Unpub. http://www.hist.umn.edu/~rmccaa/IPUMSI/trewin_ipums_report.pdf

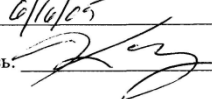
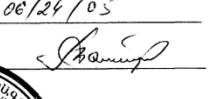

Таблица 1. Распределение стран (5 первых, все европейские, плюс Канада и США) по числу извлечений из переписей раунда 2000 года (на 2011^й календарный год)

№ п/п	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 2011 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					

Таблица 2.
32 наиболее используемые переменные в проекте IPUMS-International (85 505 извлечений)

№ п/п	Название	Выборки	Мнемокод	Примечания
1	Уровень образования	19 307	EDATTAN	
2	Возраст (в годах до 85+)	19 009	AGE	Возрастные группы =3 838
3	Статус занятости	18 490	EMPSTAT	
4	Семейное положение	18 214	MARST	
5	Вес лица	17 511	WTPER	Техническая переменная
6	Отношение к главе домохозяйства	15 783	RELATE	
7	Пол	14 595	SEX	
8	Категория работника	12 583	CLASSWK	
9	Собственность на жилье	8 050	OWNRSHIP	
10	Профессия по МСКП	8 004	OCCISCO	
11	Посещение школы	7 919	SCHOOL	
12	Продолжительность обучения	7 576	YRSCHL	
13	Грамотность	7 290	LIT	
14	Городское/сельское население	7 098	URBAN	
15	Отрасль – общий код	7 044	INDGEN	
16	Вес домашнего хозяйства	6 656	WTHH	Техническая переменная
17	Количество живорожденных детей	6 363	CHBORN	
18	Место рождения (в своей стране/за границей)	6 332	NATIVTY	
19	Занятие	6 246	OCC	
20	Страна рождения	6 153	BPLCTRY	
21	Вероисповедание	6 075	RELIG	
22	Отрасль	5 670	IND	
23	Положение супруга в домашнем хозяйстве	5 007	SPLOC	Уникальная переменная IPUMS
24	Правило определения положения супруга	4 171	SPRULE	Уникальная переменная IPUMS
25	Положение матери в домашнем хозяйстве	4 153	MOMLOC	Уникальная переменная IPUMS
26	Количество живых детей	4 074	CHSURV	
27	Местожителство 5 лет назад	4 064	MGRATE5	
28	Положение отца в домашнем хозяйстве	3 983	POPLOC	Уникальная переменная IPUMS
29	Общий доход домашнего хозяйства	3 965	INCTOT	Переменная домохозяйства
30	Доход от основной деятельности	3 655	INCEARN	
31	Количество комнат	3465	ROOMS	
32	Гражданский брак	3 443	CONSENS	

Приложение А. Пример стандартного Соглашения о взаимопонимании между Университетом Миннесоты и национальным статистическим управлением (Армения, 2004 г.)

<p style="text-align: center;">Соглашение между Международным центром общего открытого доступа к коллекциям микроданных и Национальной статистической службой Республики Армения</p> <p>Цель: Целью данного документа является определение условий и правил, согласно которым метаданные и 10% выборка микроданных переписи населения, разработанные <i>Национальной статистической службой Республики Армения</i>, будут распространяться <i>Международным центром общего открытого доступа к коллекциям микроданных</i> Университета Миннесоты (США).</p> <ol style="list-style-type: none"> Права собственности. <i>Национальная статистическая служба Республики Армения</i> является обладателем всех прав на интеллектуальную собственность (включая копирайт) в отношении метаданных и микроданных, приобретаемых Университетом Миннесоты для дальнейшего распространения <i>Международным центром общего открытого доступа к коллекциям микроданных</i>. Использование. Эти данные предоставляются для использования исключительно в образовательных целях, для научных исследований и публикаций и не могут быть использованы в других целях без заранее полученного четко сформулированного согласия (в письменном виде) <i>Национальной статистической службы Республики Армения</i>. Получение доступа. Чтобы получить доступ к данным или приобрести копии интегрированных микроданных по Армении через <i>Международный центр общего открытого доступа к коллекциям микроданных</i>, потенциальный пользователь (т.е. главный исполнитель исследования) сначала должен заполнить электронную идентификационную форму на право получения доступа, содержащую имя, электронный адрес и название учреждения. Исследователь должен охарактеризовать задачи предлагаемого проекта и выразить свое согласие с правилами, изложенными в данном документе. После того, как проект получит одобрение, исследователю будет предоставлен пароль и он сможет получать данные с серверов или других электронных средств, управляемых <i>Международным центром общего открытого доступа к коллекциям микроданных, Национальной статистической службой Республики Армения</i> или другими уполномоченными учреждениями-распространителями. После одобрения проекта пользователю предоставляется лицензия на получение интегрированных мета- и микроданных от <i>Международного центра общего открытого доступа к коллекциям микроданных</i> или других уполномоченных учреждений-распространителей. Никакие права собственности или другие права пользователю не передаются. Ограничения. Пользователям запрещается любое использование данных, полученных от <i>Международного центра общего открытого доступа к коллекциям микроданных</i> или других уполномоченных учреждений-распространителей, в коммерческих целях и любых других целях, связанных с получением прибыли. Конфиденциальность. Пользователи обязаны сохранять абсолютную конфиденциальность в отношении людей и домохозяйств. Запрещается любая попытка установить личность отдельного человека или провести идентификацию домохозяйств на основе полученных микроданных. 	<p>Запрещаются также утверждения, что отдельный человек или домохозяйство были идентифицированы по этим данным.</p> <ol style="list-style-type: none"> Безопасность. Пользователи обязаны соблюдать меры безопасности с целью предотвращения несанкционированного доступа к микроданным по Армении, полученным от <i>Международного центра общего открытого доступа к коллекциям микроданных</i> или его партнеров. Публикации. Разрешается публикация данных и результатов анализа, выполненного на основании использования мета- и микроданных по Армении, в научных докладах, научных журналах и других аналогичных изданиях. Авторы опубликованных сообщений обязаны ссылаться на <i>Национальную статистическую службу Республики Армения</i> и <i>Международный центр общего открытого доступа к коллекциям микроданных</i> как на источник данных по Армении и указывать, что все полученные в ходе исследования результаты и высказанные точки зрения являются исключительно авторскими. Нарушения. Нарушения данной лицензии на право пользования могут привести к негативной реакции профессионального сообщества, потере работы и / или к судебному разбирательству в соответствии с гражданским кодексом. Университет Миннесоты, национальные и международные научные организации и <i>Национальная статистическая служба Республики Армения</i> будут содействовать реализации на практике условий данного соглашения. Совместное использование. <i>Международный центр общего открытого доступа к коллекциям микроданных</i> будет предоставлять <i>Национальной статистической службе Республики Армения</i> электронные копии документов и данных, имеющих отношение к интегрированным микроданным <i>Национальной статистической службы Республики Армения</i>, а также своевременные отчеты получивших право доступа пользователей. Разрешение спорных ситуаций. Возможные разногласия подлежат урегулированию путем согласительных процедур и компромиссов между сторонами. В случае невозможности разрешения возникших разногласий подобными способами, последует обращение в международный арбитражный суд, который займется решением спора в соответствии с правовыми нормами. Третейский судья выбирается международным арбитражным судом МТП. Данное Соглашение строится на основе общепринятых норм международного права. Порядок старшинства. В случае возникновения противоречий между условиями данного Соглашения и условиями какого-либо контракта на профессиональные услуги, к которому прилагается данное Соглашение, преимущественную силу имеют условия данного Соглашения.
<p>Дата: 06/16/05</p> <p>Подпись: </p> <p>Члены правления Университета Миннесоты Кевин МакКоски, Администратор финансируемых проектов</p>	<p>Дата: 06/24/05</p> <p>Подпись: </p> <p>Национальная статистическая служба Республики Армения Руководитель службы</p> 

Приложение В. Фрагменты бланка заявки на использование микроданных ограниченного доступа, распространяемых проектом IPUMS-International. См: <https://international.ipums.org/international-action/register/0>

<p>IPUMS International Page 1 of 1</p> <h3>Application to Use Restricted Microdata</h3> <p>IPUMS-International microdata are available free of charge, but their use imposes responsibilities upon the user. To access the data, a prospective user must submit an electronic authorization form (this form) identifying the user by name, electronic address, and institutional affiliation.</p> <p>The investigator must state the purpose of the proposed project and agree to abide by the regulations specified below. If multiple investigators are involved in a project, all must register separately.</p> <p>Once a user is approved, a message will be sent by email granting access to the system. The notification licenses the user to acquire microdata from Integrated Public Use Microdata Series International or other authorized distributors. No titles or other rights are conveyed to the user.</p> <p>Legal notice: Submission of this application constitutes a legally binding agreement between the applicant, the applicant's institution, the University of Minnesota, and the relevant official statistical authorities. Submitting false, misleading or fraudulent information constitutes a violation of this agreement. Misusing the data by violating any of the conditions detailed below also constitutes a violation of this agreement and may lead to professional censure, loss of employment, or civil prosecution under relevant national and international laws, and to sanctions against your institution, at the discretion of the University of Minnesota and the official statistical authorities.</p> <p>Information provided on this form will be kept confidential. All information on this form is required for registration unless otherwise indicated by an asterisk.</p> <p style="text-align: center;">PERSONAL INFORMATION</p> <p>.....</p> <p style="text-align: center;">INSTITUTIONAL AFFILIATION</p> <p>IPUMS-International staff must confirm the identity of prospective users. To speed the processing of your application, please provide as much of the following</p>	<h3 style="text-align: center;">USAGE LICENSE</h3> <p>Please check all of the following boxes to indicate that you have read about the limitations of the IPUMS-International data and you agree to abide by the conditions of use. The purpose of this license is to specify the terms and condition under which integrated microdata samples distributed by Integrated Public Use Microdata Series International of the University of Minnesota may be used. Note: The license is valid for one year and may be renewed.</p> <p style="text-align: center;">Data must not be redistributed without authorization.</p> <p>All data extracted from the IPUMS-International database are intended solely for the use of the licensee. Under IPUMS-International agreements with collaborating agencies, redistribution of the data to third parties is prohibited. Each member of a research team using the data must apply for access and be licensed individually.</p> <p>The microdata are intended only for scholarly research and educational purposes.</p> <ul style="list-style-type: none"> <input type="checkbox"/> These microdata are provided for the exclusive purposes of teaching and scholarly research, and may not be used for any other purposes without explicit written approval from the relevant official statistical authority. <p>Commercial use and redistribution of the microdata is strictly prohibited.</p> <ul style="list-style-type: none"> <input type="checkbox"/> Users are prohibited from using microdata acquired from the Integrated Public Use Microdata Series International or other authorized distributors in the pursuit of any commercial or income-generating venture either privately, or otherwise. <p>Use of the microdata must follow strict rules of confidentiality.</p> <ul style="list-style-type: none"> <input type="checkbox"/> Users will maintain the confidentiality of persons and households. Any attempt to ascertain the identity of persons or households from the microdata is prohibited. Alleging that a person or household has been identified in these data is also prohibited. Statistical results that might reveal the identity of persons or entities may not be reported or published in any form. <p>The microdata must always be safely secured.</p> <ul style="list-style-type: none"> <input type="checkbox"/> Users will implement security measures to prevent unauthorized access to microdata acquired from Integrated Public Use Microdata Series International, its partners or authorized distributors. Upon the completion of this research, data may be retained only if they can be safely secured. If security cannot be guaranteed, the microdata must be destroyed.
--	---

<p>Name of institution or employer <input type="text"/></p> <p>Your email address at institution (*) <input type="text"/></p> <p>Web link showing your affiliation with institution (*) <input type="text"/></p> <p>Email address of employer, supervisor, or instructor (*) <input type="text"/></p> <p>Phone number of institution (*) <input type="text"/></p> <p>Does your institution have an Institutional Review Board (IRB), or Office for Human Subject Protections, Professional Conduct or similar committee?</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Yes; Name of board or office <input type="text"/></p> <p>RESEARCH PROJECT</p> <p>Please provide at least 75 words <i>in English</i> describing your research project or educational use for the data. This description will be used to evaluate your application.</p> <p>.....</p> <p>If your research is funded by someone other than your employer, indicate the name of the granting institution, title of grant, and other pertinent information. (*)</p> <input type="text"/>	<p>Scholarly publications are permitted, and must be cited appropriately.</p> <ul style="list-style-type: none"> <input type="checkbox"/> The publishing of research results based on IPUMS-International microdata is permitted in communications such as scholarly papers, journals and the like. The authors of these communications are required to cite Integrated Public Use Microdata Series-International and the relevant official statistical authority as the source of the microdata, and to indicate that the results and views expressed are those of the author. Users are requested to provide the IPUMS-International staff with a full citation for any publications resulting from their work with these data. <p>Any violation of this license agreement will result in disciplinary action, including possible loss of employment.</p> <p>Violation of this agreement will lead to revocation of this license, recall of all microdata acquired, a motion of censure to the relevant professional organization(s) and civil prosecution under national or international statutes, at the discretion of the Regents of the University of Minnesota and the official statistical agencies. Sanctions likewise may be taken against the institution with which the violator is affiliated.</p> <ul style="list-style-type: none"> <input type="checkbox"/> User agrees to notify ipums@pop.umn.edu regarding errors in the data.
---	---

**Приложение С. Уровень образования: интегрированные коды IPUMS для последних выборок переписей 15 стран.
По каждой стране показано абсолютное число лиц в выборке с соответствующим кодом.**

Appendix C. Educational attainment: IPUMS integrated codes for most recent census sample: 15 countries																
Case counts indicate the absolute number of persons in the sample with the corresponding code																
Code	Label	Belarus 1999	Canada 2001	France 2006	Germany 1987	Greece 2001	Hungary 2001	Ireland 2006	Italy 2001	Portugal 2001	Romania 2002	Slovenia 2002	Spain 2001	Switzer 2000	Turkey 2000	USA 2000
0	NIU (not in universe)	195,221	155,094	3,500,985	889,475	60,222	35,708	104,603	160,897	.	223,986	26,186	9,429	61,858	398,619	570,821
100	LESS THAN PRIMARY COMPLETED	2,888	14,056	.	.	34,748	29,783	.	317,651	136,635	.	10,175	.	.	1,029,257	.
110	No schooling	.	.	327,422	78,667	.	473,420	12,776	.	655,973
120	Some primary	.	.	2,223,735	.	124,318	64,846	.	.	.	1,456	1,243,973
130	Primary (4 years)	143,511	384,871
PRIMARY COMPLETED, LESS THAN SECONDARY																
Primary completed																
211	Primary (5 years)	288,939	719,220	.	.	.	502,742	.	1,129,667	.
212	Primary (6 years)	.	49,390	2,425,484	1,198,197	300,134	145,691	57,791	.	142,932	85,730	1,429,549
Lower secondary completed																
221	General and unspecified track	.	138,894	1,577,008	332,591	108,576	.	71,586	850,317	.	528,362	39,227	462,072	.	212,962	1,488,337
222	Technical track	.	.	3,504,160	7,389	.
SECONDARY COMPLETED																
General or unspecified track																
311	General track completed	207,853	91,053	2,687,840	159,502	213,311	108,812	95,777	600,717	57,048	122,104	41,637	199,858	77,527	312,107	3,534,504
312	Some college/university	.	28,393	19,065	3,391	7,717	2,955,657
320	Technical track	177,645
321	Secondary technical degree	.	22,605	.	123,152	19,591	78,940	.	129,727	.	576,739	41,066	159,802	134,987	99,849	.
322	Post-secondary technical education	.	163,114	.	211,963	46,434	.	33,809	10,513	.	56,970	7,732	114,012	27,674	.	.
400	UNIVERSITY COMPLETED	112,692	138,456	3,726,653	128,093	121,550	46,722	60,668	182,632	33,509	128,632	11,822	117,939	23,033	168,195	2,202,652
999	UNKNOWN/MISSING	5,468	.	.	117,251	.	.	16,080	.	.	28,463	1,787	.	26,231	681	.
Source: https://international.ipums.org/international-action/variables/EDATTAN#codes_section																

Приложение D. 33 университета/научных учреждения, наиболее активно использовавших микроданные IPUMS-I в 2011 году

<u>Учреждение</u>	<u>Извлечения</u>
Колумбийский университет	558
Университет Гонконга (Гонконг, САР)	309
Университет Мичигана	296
Университет штата Аризона	256
Институт медико-санитарных измерений и оценки (Сиэтл)	245
Национальный университет Ла-Плата (Аргентина)	205
Гарвардский университет	205
Национальный университет Сингапура (Сингапур)	198
Центр демографических исследований (Барселона, Испания)	185
Дартмутский колледж	177
Колледж Барух	152
Университет штата Виржиния	141
Венский институт демографии (Австрия)	138
ООН-Хабитат (Найроби, Кения)	134
Университет Стерлинга (Шотландия, Великобритания)	110
Университет Пенсильвании	108
Институт политических исследований (Париж, Франция)	107
Университет Вандербильта	105
Университет Колорадо в Боулдере	90
Университет Кейптауна (Южная Африка)	88
Университет Чикаго	86
Университет Гуэлфа (Канада)	85
Всемирный банк	81
Университет Калифорнии в Беркли	77
Университет Бригама Янга	76
Университет Помпеу Фабра (Барселона, Испания)	75
Исследовательский центр Пью	74
Межамериканский банк развития	73
Университет Кронингена (Нидерланды)	72
Колледж Кеньон	70
Лондонская школа экономики (Великобритания)	62
Университет Боккони (Милан, Италия)	59
Университет Индианы	57

Источник: База данных пользователей проекта IPUMS-International по состоянию на 1 января 2012 г. (Университет Миннесоты не включен)