

**Европейская экономическая комиссия****Конференция европейских статистиков****Группа экспертов по переписям населения и
жилищного фонда****Тринадцатое совещание**

Женева, 7–9 июля 2010 года

Пункт 2 предварительной повестки дня

Информирование о переписи и распространение данных переписи, включая использование географических информационных систем (ГИС)**Распространение интегрированных на
международном уровне микроданных переписей
населения раунда 2010 года и других раундов:
Проект IPUMS-International*****Записка Миннесотского центра по народонаселению, Миннеаполис,
штат Миннесота, США***Резюме*

Если считать справедливым утверждение, что «статистика становится полезной только при ее использовании», то очевидно, что выборки интегрированных микроданных переписей населения стремительно приобретают статус очень полезной статистики. Всего лишь десятилетие назад микроданные переписей практически не использовались, поскольку доступ к ним предоставлялся очень немногими статистическими органами, и очень небольшое число исследователей знали, как ими пользоваться. В течение прошедшего десятилетия проект IPUMS-International заключил стандартные соглашения с 90 национальными статистическими службами (21 в Европе) на предоставление бесплатного доступа официально зарегистрированным исследователям всего мира к интегрированным микроданным переписей населения. Настоящий проект (www.ipums.org/international) осуществляется Центром по народонаселению Университета Миннесоты (МЦН) при участии (для работы с европейскими переписями) Центра демографических исследований (ЦДИ) Автономного университета Барселоны (www.iecm-project.org).

Для выявления наиболее востребованных показателей и выборок в рамках проекта IPUMS проанализированы 59 170 извлечений с данными. Возможно, самым неожиданным оказался тот факт, что в четырех процентах наиболее часто запрашиваемых показателей оказались четыре, созданные проектом IPUMS: определение положения супруга в домашнем хозяйстве, определение положения матери, определение положения отца и правило определения связей с ними. Такие показатели можно было построить только из выборок домашних хозяйств. Для достоверных исследований на базе микроданных переписей крайне необходимы высокоточные, хорошо

* Средства для данных исследований выделены частично Национальным институтом здравоохранения США, грант HD047283 для осуществления проекта по гармонизации микроданных европейских и азиатских переписей населения (IPUMS-EurAsia) и Европейским Союзом в рамках разработки исследовательских инфраструктур, FP6-026033, для осуществления проекта по гармонизации интегрированных микроданных европейских переписей населения (НЕСМ).

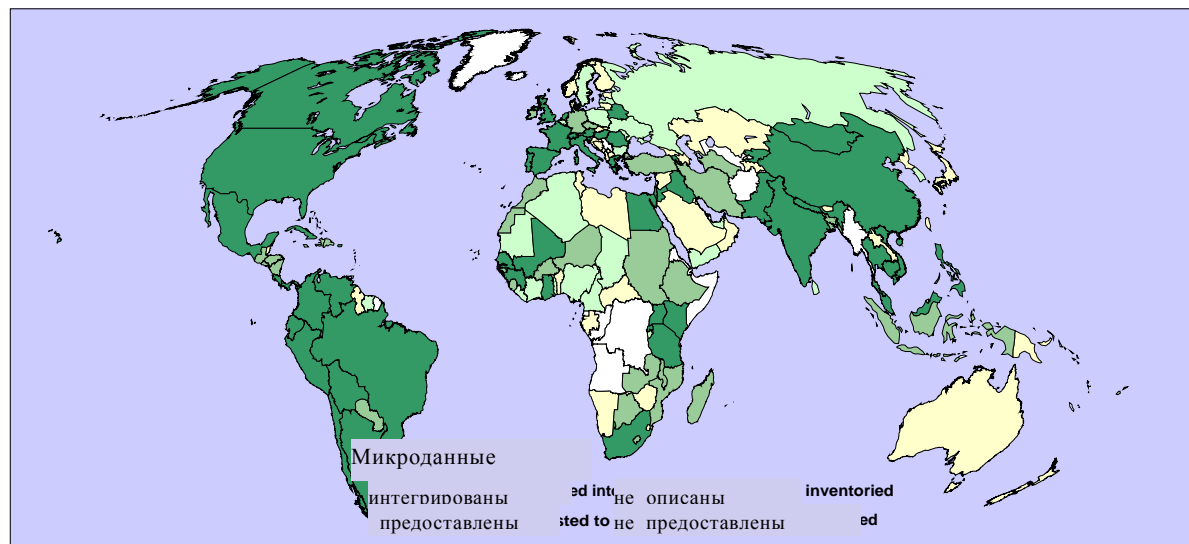
детализированные выборки домашних хозяйств.

I. Проект IPUMS-International: крупномасштабный, широко используемый всемирный ресурс микроданных переписей населения с ограниченным доступом

1. В рамках проекта IPUMS-International (www.ipums.org/international) архивируются, интегрируются и распространяются высокоточные, хорошо детализированные микроданные из национальных переписей населения и жилищного фонда. Эта массивная инфраструктура данных, включающая более 325 миллионов анонимизированных, интегрированных персональных записей, представляющих почти 90 миллионов домашних хозяйств, построена на базе 159 переписей из 55 стран. Благодаря непрерывному финансированию проекта со стороны Национального научного фонда и Национального института здравоохранения (США), а также чрезвычайно интенсивному сотрудничеству с проектом национальных статистических служб разных стран в базе данных ежегодно прибавляется от 5 до 10 новых стран (см. таблицу 1).

Таблица 1

Микроданные переписей населения стран, сотрудничающих с проектом IPUMS-International, по состоянию на июль 2010 года:
3 оттенка зеленого цвета—данные интегрированы (темно-зеленый), в процессе интеграции (средний) и в стадии переговоров (светло-зеленый)



2. В проекте IPUMS принимает участие 21 европейская страна (в скобках указано число предоставленных массивов данных; * = микроданные интегрированы и используются): *Армения (1), *Австрия (4), *Беларусь (1), Болгария (0), Чешская Республика (2), *Франция (7), Германия (8—включая ГДР, ФРГ и микропереписи), *Греция (4), *Венгрия (4), Ирландия (8), *Италия (1), *Нидерланды (3), Польша (0), *Португалия (3), *Румыния (3), *Словения

(1), *Испания (3), *Швейцария (4), Турция (0), Украина (0), *Великобритания (2—в будущем до 6).

3. Интеграция микроданных переписей населения раунда 2000 года завершена для 14 европейских стран. В 2011 году планируется встроить выборки еще трех европейских государств: Франции (перепись 2006 года), Ирландии и Германии, а также девяти других стран (Камбоджи, Египта, Ямайки, Индонезии, Ирана, Израиля, Никарагуа, Палестины и Сьерра-Леоне). В последующие годы будет проводиться интеграция выборок переписей раунда 2010 года по мере их поступления.

4. Несмотря на бесплатный доступ к микроданным проекта IPUMS-International, он разрешен только добросовестным исследователям, давшим согласие на соблюдение жестких требований использования данных. Проект IPUMS распространяет извлечения данных, выполненные по точным требованиям каждого отдельного пользователя. Среднее извлечение данных в проекте IPUMS включает не более 10 показателей. Такой подход отличается от принятой в большинстве статистических служб практики распространения микроданных переписей населения в виде полных систем, включающих словарь данных и всю выборку. Обычно в таких случаях при выполнении запросов все исследователи получают совершенно одинаковые наборы данных и документации. Учитывая огромные размеры базы данных проекта IPUMS-International, распространение полных наборов показателей и выборок одного размера практически нецелесообразно. Вместо этого в рамках проекта IPUMS исследователь может запросить извлечение из базы данных по следующим параметрам:

- (a) страна (или страны) ;
- (b) год (годы) переписи ;
- (c) показатели (возраст, пол, уровень образования и т.д.);
- (d) подсовокупности (например, главы домашнего хозяйства – женщины в возрасте менее 25 лет вместе со всеми другими лицами, проживающими в отобранном домашнем хозяйстве) ;
- (e) и плотность выборки (в процентах или в абсолютных числах).

5. Механизм извлечения данных в проекте IPUMS позволяет выполнять запрос путем создания такого набора данных, который содержит только запрашиваемые микроданные и соответствующую кодовую книгу (возможны 4 вида: типовой, SPSS, SAS или STATA). Полные метаданные можно получить дополнительно с веб-сайта как в виде документов, так и в диалоговом режиме.

6. На совещании Группы экспертов ЕЭК ООН по конфиденциальности статистических данных в ноябре 2005 года было представлено следующее описание процесса распространения данных в рамках проекта IPUMS-International¹:

¹ Роберт МакКей и Альберт Эстив, "IPUMS-Europe: Меры по защите, используемые в процессе разрешения и предоставления исследователям ограниченного доступа к микроданным переписей населения," Монографии официальной статистики: Рабочее заседание по конфиденциальности статистических данных. Люксембург: Отдел официальных публикаций Европейского Сообщества , р. 40.

7. Как только извлечение данных (обычно это занимает несколько минут) завершено, исследователя извещают по электронной почте о том, что данные следует получить в течение 72 часов. Ему предлагается ссылка на защищенный паролем сайт для скачивания конкретного извлечения. Во время передачи данные шифруются по стандарту шифрования 128-bit SSL (128-разрядный протокол защиты информации), который соответствует уровню, применяемому в банковском деле и других отраслях, требующих безопасности и конфиденциальности данных. Исследователь может, не беспокоясь о защите, скачать файл, разархивировать его и приступить к анализу данных, используя прилагаемые интегрированные метаданные, содержащие названия и обозначения показателей.

8. Данный метод распространения данных выдержал проверку временем, и его применение значительно расширилось благодаря быстрому росту скоростей передачи данных через Интернет.

II. Статистика использования IPUMS-International (по 4 июня 2010 года включительно)

9. С сайта IPUMS-International было произведено 59 170 извлечений данных, что составляет в среднем по 1 000 извлечений на страну из 55 стран, представленных в базе данных (таблица 2). Однако обращение к странам значительно различается. Из 44 стран, выборки которых стали доступными не более одного года, наименьшее количество извлечений, а именно 116, зарегистрировано по переписи населения Кыргызской Республики 1999 года. Больше всего запросов (7 637) было сделано по шести переписям населения Мексики. Наибольшее количество извлечений данных (2 464) было запрошено из выборки Мексики 2000 года.

10. Чаще всего используются выборки Мексики, Бразилии и Колумбии, и не только потому, что они содержат многочисленные показатели и длинные динамические ряды, позволяющие наблюдать полувековую историю населения, но и благодаря тому, что многие латиноамериканские эмигранты проживают в США (или Испании), что дает возможность анализировать эти популяции в единой интегрированной базе данных независимо от того, проживают они в стране рождения или в одной из двух стран со значительной эмиграцией – США или Испании. Кроме того, все латиноамериканские выборки, а также выборки США и Испании, отличаются высокой точностью и хорошей детализацией, дают обширную информацию о миграции, включают экономические, социальные и демографические показатели для характеристики как отдельных лиц, так и домашних хозяйств.

Таблица 2

Первая пятерка стран, а также все европейские страны, плюс Канада и США по количеству извлечений данных

Место	Страна	Выборка % *	Показатели (число)*	Годы переписей	Извлечения
1	Мексика	10	120	1960р, 70, 80, 95, 2000, 05	7 637
2	Бразилия	5	106	1960, 70, 80, 91, 2000	5 191
3	США	5	92	1960, 70, 80, 90, 2000, 05	4 559
4	Колумбия	10	120	1964р, 72, 85, 93, 2005	3 428
5	Франция	5	99	1962, 68, 75, 82, 90, 99	2 795
10	Канада	2.5	59	1971р, 81р, 91р, 2001р	1 614
12	Испания	5	99	1981, 91, 2001	1 514
13	Греция	10	89	1971, 81, 91, 2001	1 496
19	Венгрия	5	74	1970, 80, 90, 2001	1 132
21	Австрия	10	75	1971, 81, 91, 2001	1 087
22	Португалия	5	96	1981, 91, 2001	1 028
23	Румыния	10	97	1976, 92, 2002	1 012
29	Великобритания	3	47	1991, 2001р	657
30	Нидерланды	1	33	1960р, 71р, 2001р	570
32	Беларусь	10	84	1999	333
38	Италия	5	81	2001	209
43	Словения	10	80	2002	133
46	Швейцария	5	79	1970, 80, 90, 2000	40**
Всего извлечений из базы данных IPUMS-International по 55 странам (159 выборок) на 4 июня 2010 г.					59 170
*перепись раунда 2000 г.; относится ко всем интегрированным показателям, включая созданные проектом IPUMS.					
**выборка, выпущенная 1 июня 2010 г. (т.е. с доступом менее одной недели)					
"р" = выборка лиц; все другие выборки построены на домашних хозяйствах					

11. 3 645 исследователей из 84 стран получили право на доступ к базе данных IPUMS-International. На среднего пользователя приходится по 16 запросов. В общей сложности были скопированы более полумиллиона (586 643) показателей. Тот факт, что среднее извлечение включает не более 10 показателей, подтверждает, что пользователи базы данных IPUMS-International не просто собирают данные, а по-настоящему исследуют их. Запросы всегда минимальны, ограничены четко сформулированными потребностями конкретного исследования.

12. В таблице 3 перечислены 32 наиболее часто запрашиваемых показателя из 736 интегрированных. Первые 8 включают 4 демографических показателя (семейное положение, родственное отношение к главе, возраст и пол), 2 экономических (статус занятости и класс рабочих), а также один социальный (уровень образования) и один технический (вес лица).

Таблица 3

32 наиболее используемых показателя в IPUMS-International

Место	Показатель	Извлечения	Обозначение	Комментарии
1	Уровень образования	19 307	EDATTAN	
2	Возраст (годы до 85+)	19 009	AGE	Возраст по группам n=3 838
3	Статус занятости	18 490	EMPSTAT	
4	Семейное положение	18 214	MARST	
5	Вес лица	17 511	WTPER	Техническая переменная
6	Родственное отношение к главе	15 783	RELATE	
7	Пол	14 595	SEX	
8	Класс рабочих	12 583	CLASSWK	
9	Владение жилищем	8 050	OWNRSH	
10	Профессия по МККП с новым кодом	8 004	OCCISCO	
11	Посещение учебного заведения	7 919	SCHOOL	
12	Годы учебы	7 576	YRSCHL	
13	Грамотность	7 290	LIT	
14	Городской/сельский	7 098	URBAN	
15	Отрасль - общий код	7 044	INDGEN	
16	Вес домашнего хозяйства	6 656	WTHH	Техническая переменная
17	Рожденные дети	6 363	CHBORN	
18	Статус рождения (в стране/за границей)	6 332	NATIVITY	
19	Профессия	6 246	OCC	
20	Страна рождения	6 153	BPLCTRY	
21	Религия	6 075	RELIG	
22	Отрасль	5 670	IND	
23	Определение положения супруга в домашнем хозяйстве	5 007	SPLOC	Созданная (домашнее хозяйство)
24	Правило определения положения супруга	4 171	SPRULE	Созданная (домашнее хозяйство)
25	Определение положения матери в домашнем хозяйстве	4 153	MOMLOC	Созданная (домашнее хозяйство)
26	Количество живых детей	4 074	CHSURV	
27	Постоянное местожительство 5 лет назад	4 064	MGRATE5	
28	Определение положения отца в домашнем хозяйстве	3 983	POPLOC	Созданная (домашнее хозяйство)
29	Общий доход домашнего хозяйства	3 965	INCTOT	Переменная домашнего хозяйства
30	Трудовой доход	3 655	INCEARN	
31	Количество комнат	3 465	ROOMS	
32	Гражданский брак	3 443	CONSENS	

13. К нашему большому удивлению четыре созданных проектом IPUMS показателя оказались среди 30 наиболее часто запрашиваемых, это: определение положения в домашнем хозяйстве супруга, матери и отца и правило установления положения супруга (для опытных пользователей IPUMS показатели “LOC”). Исследователи используют эти показатели для изучения общих характеристик супругов и характеристик отношений между родителями и их детьми. Эти показатели создаются на основании родственных связей с главой хозяйства, возраста, пола, семейного положения, порядка перечисления лиц в домашнем хозяйстве и нескольких других показателей. Интенсивное использование показателей “LOC” указывает на их большое значение для анализа лиц в соответствии с характеристиками их супругов, матерей и отцов.

14. Вследствие повышения спроса на интегрированные микроданные переписей, увеличения объемов работы по интеграции микро- и метаданных переписей, а также расширения взаимодействия со многими статистическими службами сотрудники проекта IPUMS-International приняли следующие минимальные стандарты по интеграции и распространению микроданных²:

(а) Выборки домашних хозяйств (выборки отдельных лиц используются только в случаях отсутствия других микроданных, как в переписях раунда 1970 года и более ранних) ;

(б) Высокая точность—минимум 5%, предпочтительно 10% ;

(с) Обширный набор показателей—обычно все имеющиеся, за исключением требующих соблюдения статистической конфиденциальности ;

(д) Детализированные коды— для возраста (точные годы до возраста 85 лет), профессия (3 знака МСКП), отрасль (3 знака МСОК), страна рождения (подробно) и т.д.— скрываются только коды, способные повредить статистической конфиденциальности, например, коды с низкой частотностью.

15. ИНСЭЕ Франции, член-учредитель настоящей инициативы, уже прокладывает путь к раунду 2010 года своим представлением проекту выборки высокой плотности, состоящей из почти 20 миллионов персональных записей, содержащих 94 показателя из последней переписи 2004-2008 гг.

III. Заключение

16. Когда мы десять лет назад начинали этот проект, нашей заветной мечтой было интегрировать в течение каждого десятилетия данные из 21 страны. Благодаря интенсивной помощи со стороны национальных статистических служб и невообразимым техническим достижениям эта цифра выросла в более чем два раза. Число пользователей и размах использования данных тоже сильно превысили наши ожидания. Во втором десятилетии мы хотели бы удвоить число пользователей и еще раз удвоить число выборок. Для успеха этого предприятия огромное значение будут иметь высокоточные выборки переписей раунда 2010 года. Просим участвующие в проекте статистические органы как можно скорее представить мета- и микроданные своих переписей раунда 2010

² Роберт МакКей, Венди Томас, Альберт Эстив и Антонио Лопез Гей “Предоставление микроданных и метаданных переписей для своевременной интеграции и распространения через проекты IPUMS-EurAsia и IECSM, 2010-2014 гг.”, Совместное совещание ЕЭК ООН/Евростата по переписям населения и жилищного фонда (28-30 октября 2009 г., Женева, Швейцария)

года. А еще не присоединившиеся к проектам IPUMS/IECM страны приглашаем к рассмотрению возможности вступления.

**Приложение А. Проект IPUMS-Europe, 17 стран: 1
301 интегрированный показатель выборок переписей
населения раунда 2000 года (июнь 2010 г.)**

См. прилагаемую таблицу.
