

**Европейская экономическая комиссия****Конференция европейских статистиков****Группа экспертов по переписи населения и жилищного фонда****Восемнадцатое совещание**

Женева, 28 - 30 сентября 2016 года

Пункт 2 предварительной повестки дня

Оценка издержек и выгод переписей**Жизнь и ценность данных после опубликования
результатов: использование микроданных переписи в
научных и политических целях от партнерства IPUMS
International****Доклад Университета Миннесоты¹***Резюме*

Перепись обеспечивает ценные данные по стране и проживающему в ней населению. Поскольку сейчас возросло наличие выборок микроданных переписи, стало проводиться больше научных исследований с использованием этих данных. Система гармонизации и распространения данных, обеспечиваемая IPUMS, способствует проведению передовых исследований путем минимизации поиска и подготовки данных исследователями и предоставления обширных метаданных в формате, удобном для пользования. ООН рекомендует правительствам обеспечивать доступ к микроданным для целей научных исследований, а в Повестке дня в области устойчивого развития на период до 2030 года подчеркивается необходимость дезагрегирования данных по ряду признаков. Микроданные идеально подходят для этих целей, поскольку они позволяют проводить измерения с учетом таких характеристик как возраст, пол, географические регионы, статус инвалидности и другие характеристики человека или домохозяйства. Микроданные переписи являются чрезвычайно важными для понимания масштабных тенденций, таких как урбанизация, экономическое развитие, динамика в области рождаемости, миграция, старение населения, развитие образования и инвалидность. Предельные издержки на создание выборки микроданных составляют лишь малую часть всего бюджета переписи и невелики по сравнению с получаемыми выгодами. В целях сохранения коллективной памяти и стимулирования научных открытий, статистическим органам необходимо обеспечивать возможность для создания выборок микроданных в ходе подготовки к переписи.

¹ Подготовлено Ларой Кливленд и Патрицией Келли Холл, IPUMS International.

1. Как большой собор, культовое здание или важный момент, перепись обеспечивает письменный след о народе и его культуре в конкретный момент времени. Как те огромные сооружения, перепись является очень дорогостоящим мероприятием, ее проведение обходится в несколько миллионов или даже миллиардов долларов. Даже наименее дорогостоящая перепись может стоить правительству около 1 доллара на человека. По сведениям, США потратили около 42 доллара на душу населения для проведения переписи населения в 2010 году, что является одним из самых высоких показателей. Сметная стоимость ряда последних переписей приведена в Приложении А. Планирование, проведение полевых операций и анализ результатов переписи требует огромных инвестиций и ответственности. В свете последних финансовых событий в мире и сокращения бюджетов при проведении переписи необходимо добиться больших результатов за меньшие деньги.

I. Важность микроданных переписи

2. Можно сказать, что перепись является более ценным отпечатком страны и ее народа, чем какое-либо здание или памятник, особенно, когда информация сохраняется и доступна для научных исследований. При небольших предельных издержках можно получить огромную пользу путем подготовки выборки микроданных переписи для публичного пользования. Для сохранения нашей коллективной памяти и возможности совершения научных открытий мы считаем, что статистические органы должны обеспечивать создание выборок микроданных как важную часть процесса переписи.

3. Доступ к микроданным переписи становится еще более важным, чем когда-либо. Гибкость микроданных для возможности проведения исследований в области социальных наук отмечается в недавнем отчете ООН по укреплению демографической доказательной базы для целей Повестки дня в области устойчивого развития на период до 2030 года. В этом отчете Комиссия по народонаселению и развитию «рекомендует Правительствам принять политику открытого доступа к данным, обеспечивая возможность распространения публичных анонимизированных микроданных с географической привязкой, полученных в ходе переписи, обследований домашних хозяйств, регистрации актов гражданского состояния и содержащихся в регистрах населения, информационных системах здравоохранения и других административных источниках с соблюдением конфиденциальности данных» (<http://www.un.org/en/development/desa/population/pdf/commission/2016/documents/CPD49%20Resolution%202.pdf>). В рекомендациях признается ценность микроданных для дезагрегации и комплексного многомерного анализа.

4. Рекомендации ООН по распространению микроданных согласуются с переходом Повестки дня на период до 2030 года от отдельных задач в рамках Целей развития тысячелетий к измерению изменений и оценки отличий среди подгрупп в рамках показателей Целей устойчивого развития. В новых целях также подчеркивается важность отслеживания прогресса путем направления мониторинговых групп для установления базовой ситуации, в сравнении с которой можно оценивать изменение. Микроданные являются достаточно гибкими для соответствия требованиям по дезагрегации для мониторинга ЦУР ради того, чтобы не оставить никого забытым. В частности, данные переписи обеспечивают достаточно широкий охват, и даже выборка из результатов переписи обеспечивает достаточную статистику для того, чтобы можно было проводить дезагрегацию по возрасту, полу, географическим регионам, наличию инвалидности и т.д. Микроданные переписи также дают богатую информацию о взаимосвязях между характеристиками домохозяйства, семьи и человека.

II. Наличие микроданных в США

5. США имеют более 50 лет опыта распространения выборок микроданных переписи. В США давно придерживаются философии открытости данных. На сбор и создание государством данных тратятся деньги налогоплательщиков. Поэтому данные считаются собственностью народа и должны быть публичными, при надлежащем обеспечении их конфиденциальности. К 1962 г. технологический прогресс позволил Бюро переписей США выпустить первую выборку микроданных переписи для общественного пользования по результатам переписи 1960 г. (Ruggles et. al 2015). Выборка 1 к 1 000 домохозяйств и людей была распространена на 13 лентах Univac или 18 тыс. перфокарт. После переписи США выпустили выборки 1 к 100 с сопоставимыми форматами и схемами кодирования для переписей 1970 и 1960 гг. Наличие двух аналогично форматированных файлов, охватывающих 10-летний период, привело к стремительному росту исследований по динамике изменений. Социолог Отис Дадли Данкан отмечал важность наличия выборок микроданных переписи: «слишком часто усилия по приведению данных в надлежащий формат становятся тщетными вследствие несоответствия публикуемых сводных таблиц поставленным целям. Имея доступ к элементарным записям, социолог может детально указать какие манипуляции необходимо произвести с переменными для того, чтобы получить оптимальные расчеты» (Duncan 1974; 5097).

6. В 1993 г. Демографический центр Миннесоты получил финансирование на проведение работ по гармонизации кодов, записи макетов и метаданных по девяти ранее несовместимым переписям, охватывающим период с 1880 по 1990 г. Для ученых, которые достаточно стары для того, чтобы помнить, компьютерная лента с 7 дорожками была средством распространения с 60 млн. интегрированными персональными записями, упакованными на одну катушку. Двумя годами позже был запущен первый веб-сайт Интернет для серий открытых комплексных микроданных (IPUMS), и распространение на лентах быстро забыли. Работа, проведенная IPUMS по стандартизации и гармонизации данных, является находкой для исследователей, которым не нужно теперь искать информацию в разных местах и посредством множества документов получать доступ и составлять эти массивы данных. В результате популярности системы документации и доступа к данным было выделено еще несколько грантов для того, чтобы гармонизировать файлы других переписей и обследований.

7. С середины 1990-х гг. семейство проектов IPUMS по интеграции данных имеет новые переписи и обследования США в гармонизированной базе данных и продолжило улучшать систему распространения данных. В число семейства проектов IPUMS входят выборочные данные переписей США, исторические переписи сплошного учета, обследование американских сообществ, Национальное анкетирование по вопросам здоровья, текущий обзор населения и американское обследование бюджета времени. Также в IPUMS были добавлены международные данные переписей и медико-демографических обследований. Сегодня IPUMS в США распространяет специально подбираемые выборки данных по переписям в США, проведенным в период с 1850 по 2010 г. Каждая выборка, вне зависимости от количества запрашиваемых переписей, объединяется в единый массив данных. Ежегодно в течение недели или двух после выпуска Бюро переписей США в базу данных IPUMS США загружаются последние выборочные данные обследования американских общин (ACS). Обновления включают трехгодичные и пятилетние версии ACS, а также годовые версии. В Таблице 1 представлены страны и имеющиеся объемы выборок международных данных переписей.

III. Подход IPUMS к гармонизированным данным, доступным для научных исследований

8. IPUMS предлагает средства распространения микроданных, которые дополняют работу, проводимую национальными статистическими органами по распространению данных. НСО распространяют официальные статистические данные и официальные статистические продукты для широкого круга пользователей – граждан, государственных органов, СМИ, аналитиков и т.д. IPUMS-International распространяет микроданные ограниченного доступа небольшой, но важной категории лиц – исследователям, которым необходимы данные по людям и домохозяйствам для измерения и анализа сложных взаимоотношений, часто проводя сопоставления с течением времени и между разными странами.

9. IPUMS никогда не распространяет первоначальные необработанные массивы данных переписи. Микроданные трансформируются, анонимизируются, гармонизируются и интегрируются таким образом, что любое понятие, такое как оконченное начальное образование, имеет один и тот же код в каждой выборке во всей базе данных. При этом целые массивы данных никогда не распространяются. Вместо этого, каждый исследователь строит специализированную «выборку», подобную таблице с множественными переменными для конкретного вопроса, поднимаемого в его исследовании. Исследователь использует электронное меню инструмента IPUMS по распространению данных для того, чтобы подать заявку на получение выборки данных, которая оптимизирована под такие переменные как страна (страны), год (годы) переписи, подсовокупность (подсовокупности), плотность выборки и др. в соответствии с индивидуальными потребностями исследователя. В результате каждого запроса на предоставление выборки готовится единый набор данных, который регистрируется в целях обеспечения воспроизводимости и защиты от мошенничества. Этот метод обеспечивает мотивацию для того, чтобы пользователи ревностно оберегали свои выборки данных и выполняли условия их использования. Поскольку полные наборы данных не распространяются на DVD-дисках или других носителях, соблазн поделиться микроданными с неуполномоченными лицами сильно сокращается.

10. Классификации в первоначальных исходных файлах микроданных не совпадают в разных годах и странах переписи. Сверка и согласование этих кодов и является основной частью проекта IPUMS. В выборках IPUMS сохраняется вся подробная информация, содержащаяся в исходных выборках за исключением случаев, когда требуется преобразование в целях обеспечения конфиденциальности. В тоже время мы обеспечиваем полностью интегрированную базу данных, в которой идентичным категориям в разных выборках присваиваются идентичные коды. В приложении В перечислены гармонизированные переменные, характерные для большинства выборок в базе данных IPUMS.

11. Нами применяется несколько стратегий для того, чтобы достичь конкурирующих между собой целей: обеспечить максимальную сопоставимость и сохранить детализацию. По простым переменным, таким как возраст и пол, первоначальные переменные являются сопоставимыми и перекодировать их в единую классификацию не представляет трудности. Для более сложных переменных невозможно построить единую последовательную классификацию без потери информации. Некоторые переписи обеспечивают более детальную информацию по сравнению с другими переписями, поэтому при наименьшем общем знаменателе всех выборок неизбежно теряется важная информация. В этих случаях мы строим иерархические или составные схемы кодирования. Первая цифра кода дает информацию, которая имеется во всех выборках. Следующие одна или две цифры

дают дополнительную информацию, имеющуюся в широком множестве выборок. И, наконец, конечные цифры дают информацию, которая имеется в редких случаях (Esteve and Sobek 2003; Ruggles 2006).

12. В целях иллюстрации подхода, используемого IPUMS, рассмотрим схему классификации по статусу занятости EMPSTAT (см. Таблицу 2). Каждая колонка представляет собой одну выборку. Строки отражают понятия, каждая из которых имеет свой иерархический код. В каждой ячейке указано количество случаев для соответствующего понятия и выборки. Составные коды обеспечивают интуитивное средство для понимания взаимосвязей между понятиями. В соответствии с форматом IPUMS первая цифра статуса занятости имеет три категории, которые имеются во всех выборках: 1) экономически активные, 2) незанятые, и 3) экономически неактивные. В каждой из этих трех категорий вторая и третья цифра определяют разные понятия. Например, среди экономически неактивного населения разграничения являются достаточно сложными. На второй цифре неактивного населения (3x) в схеме кодирования IPUMS проводятся разделение между домашней работой, инвалидами, учащимися, пенсионерами, пожилым населением, находящимися в специальных учреждениях, нерегулярными рабочими, получателями других доходов и пр. Третья и последняя цифра обеспечивают еще более подробную информацию, несмотря на то, что конкретное явление может произойти в относительно редких случаях. Пенсионеры (34x) относятся к наиболее сложному контингенту и сильно отличаются в разных выборках. Третья цифра состоит из девяти различных типов, хотя для краткости в Таблице 2 представлены только 4. И, наконец, обратите внимание на то, что схема кодирования IPUMS всегда присваивает «0» «тем, кто не относится к генеральной совокупности» и 9, 99, 999 и т.д., неизвестным, с цифрой 9, определяемой максимальным количеством знаков для конкретной переменной.

13. Основное преимущество интеграционной схемы IPUMS состоит в ее сверке кодов переменных, конкретных для переписи, для выработки наборов данных, которые интегрируют записи за разное время и в разных регионах. Основная цель переменной гармонизации состоит в том, чтобы сделать данные пригодными для сопоставительного анализа путем применения сопоставимых кодов для каждой переменной по всем выборкам в базе данных. Микроданные интегрируются, поэтому одинаковые явления имеют одинаковые коды.

14. Помимо интегрированных кодов, метаданные IPUMS предлагают детальную документацию с помощью «закладок» для общего описания, обсуждений сопоставимости, указаний о совокупности, наличия явлений, детальной формулировки первоначального текста и ссылок на исходные документы на официальном языке и перевод на английский язык.

15. В Приложении С кратко представлены интегрированные переменные IPUMS, которые имеются по каждой выборке домашних хозяйств в базе данных IPUMS. Большинство этих переменных является уникальными для IPUMS и редко имеются в выборках, распространяемых национальными статистическими органами.

16. Интегрированные микроданные и метаданные способствуют проведению сравнительных исследований, но изменения в административных границах представляют собой серьезную проблему для временно-пространственного анализа, необходимого для точного мониторинга демографических и социальных изменений. Для измерения прогресса в достижении целей на субнациональном уровне чрезвычайно важно, чтобы пространство оставалось постоянным; единицы, которые имеют изменившиеся границы, не могут результативно сравниваться с течением времени. Постоянные пространственные географические единицы являются необходимыми для точного измерения изменений с течением времени с привлечением контекстуальных или пространственных элементов (Sarkar et al. 2015).

В IPUMS разработали метод для создания пространственно постоянных единиц в микроданных, начиная с первой и второй административных единиц, определенных в выборках переписи. Также для загрузки имеются файлы с границами в ГИС, соответствующие месту проживания, месту рождения и другим географическим переменным.

IV. Пользователи и применение микроданных IPUMS

17. С того момента, когда началось распространение гармонизированных для исследований массивов данных, использование этих данных для научных исследований резко выросло. На сегодняшний день зарегистрировано свыше 100 тыс. уникальных пользователей (Рис. 1), которые используют один или более источников данных, указанных выше. Эти пользователи загружают 2,6 терабайтов данных в неделю и веб-сайтов IPUMS (Рис. 2), делая более 2 000 научных цитат в год, согласно данным Google Scholar (Рис. 3). По сути, 13% статей в одном из самых авторитетных журналов, «Demography», основывается на данных, распространяемых посредством веб-сайтов проекта IPUMS (Рис. 4).

18. В 1999 г. IPUMS стал международным проектом и в течение пяти лет распространил объединенные выемки анонимизированных, интегрированных микроданных по восьми странам: Бразилия (1960-2000 гг.), Китай (1982-1990 гг.), Колумбия (1962-1993 гг.), Франция (1962-1996 гг.), Кения (1989-1999 гг.), Мексика (1960-2000 гг.), США (1960-2000 гг.) и Вьетнам (1989-1999 гг.). Десять лет спустя сайт предлагает десятикратное увеличение количества микроданных и стран, доступных для исследователей. Степень использования росла еще быстрее, увеличиваясь в два раза каждые два-три года. Сейчас свыше 600 млн. интегрированных записей микроданных, охватывающих четыре пятых населения планеты, предоставляется свыше 10 тыс. исследователей более чем в 100 странах. Согласно соглашениям с национальными статистическими органами для получения доступа к выборкам международных микроданных утверждаются только прошедшие проверку кандидаты, работающие над подтвержденными и авторитетными исследовательскими проектами.

19. Пользователи данных IPUMS International подготовили свыше 1 000 статей и докладов. В числе пользователей - экономисты (около 40%), демографы (20%), социологи (10%) и другие специалисты в области социальных наук. Статистики являются растущей и более новой группой пользователей благодаря росту охвата проекта, привлекая в него членов профессиональных аффилированных групп Международного статистического института. Большинство пользователей запрашивают данные для проведения исследования для написания статьи или диссертации (Таблица 2). Последние статьи на различные темы были опубликованы в журнале прикладной экономики, журнале экономики развития, журнале исследований в области народонаселения, журнале по международной миграции, в исследованиях по миграции, журнале «Народонаселение и развитие» и многих других. Исследователи смоделировали потоки внутренней миграции (García et al 2014); влияние гендера, семейного положения и иммиграции на участие в рабочей силе (Donato et al 2014); сожительство в Латинской Америке (Esteve et al 2012) и разнообразие в составе домашних хозяйств (Demont & Neuveline 2008). В двух из этих исследований данные переписи увязываются с информацией по здоровью для изучения смертности в результате СПИДа и рынка труда (Chicoine 2011) или для анализа квазиэкспериментальной ситуации в результате конкретного случая лечения малярии (Bleakley 2010). Сложно определить ценность исследований, протекающих из доступа к микроданным, но однозначно она значительно

превосходит предельные издержки на подготовку выборки из данных, подготовленных для отчетов о результатах переписи.

V. Заключение

20. Международные микроданные переписи, частично благодаря IPUMS, стали важной частью нашей общей научной инфраструктуры, поскольку они обеспечивают уникальную лабораторию для анализа социально-экономического прогресса и эмпирическое основание для разработки и проверки теоретических моделей. Микроданные являются чрезвычайно важными для понимания масштабных тенденций, таких как урбанизация, экономическое развитие, рождаемость, миграция, старение населения, развитие образования и инвалидность. Данные идеально подходят для оценки последствий социальных, экономических и демографических изменений. Новые методы позволяют оценивать связь между населением, климатом и окружающей средой. Микроданные переписи являются богатыми и уникальными артефактами нашего общего социально-культурного состояния. Сохранение этой богатой истории позволяет непрерывно переоценивать то, что обеспечивает нам рост. Учитывайте уровень инвестиций в проведение переписи. Если публикуемые результаты переписи обеспечивают важное статическое представление в конкретный момент времени, то выборки микроданных являются живой историей, доступной для постоянного исследования. Возможность вернуться к нашей богатой истории и сделать новые научные открытия оправдывает инвестиции в виде предельных издержек на составление выборки микроданных.

VI. Библиография

- Bleakley, Hoyt. 2010. Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure. *American Journal of Applied Economics* 2(2): 1-45.
- Chicoine, Luke. 2011. Aids mortality and its effect on the labor market: Evidence from South Africa. *Journal of Development Economics* 98: 256-269.
- Demont, Floriane and Patrick Heuveline. 2008. Diversity and Change in Cambodian Households, 1998-2006. *Journal of Population Research* 25(3): 287-313.
- Donato, Katherine M., Bhumika Piya, and Anna Jacobs. 2014. The Double Disadvantage Reconsidered: Gender, Immigration, Marital Status, and Global Labor Force Participation in the 21st Century. *International Migration Review* 48: S335-S376. doi: 10.1111/imre.12142
- Duncan, O. D. 1974. Developing social indicators. *Proceedings of the National Academy of Sciences* 71, 5096-5102.
- Esteve, A., R. Lesthaeghe and A. López-Gay. 2012. The Latin American cohabitation boom, 1970-2007. *Population and Development Review* 38: 55-81.
- Esteve, A. and M. Sobek. 2003. Challenges and methods of international census harmonization. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 36 (2): 66-79.
- Garcia, Andres J., Deepa K. Pindolia, Kenneth K. Lopiano, and Andrew Tatem. 2014. Modeling internal migration flows in sub-Saharan Africa using census microdata. *Migration Studies* 2014: 1-22.
- Ruggles, S., 2014. Big microdata for population research. *Demography*, 51(1), pp.287-297.

Ruggles, S., L. Cleveland, R. McCaa and M. Sobek. 2015. The IPUMS Collaboration: Integrating and Disseminating the World's Population Microdata, *Journal of Demographic Economics*, 81(2): 203-216.

Ruggles, Steven, Miriam L. King, Deborah Levison, Robert McCaa, and Matthew Sobek. 2003. "IPUMS-International." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 36 (2): 60–65. doi:10.1080/01615440309601215.

Sarkar, Sula, Lara Cleveland, Majory Silisyene, and Matthew Sobek. 2015. "Harmonized census geography and spatio-temporal analysis: gender equality and empowerment of women in Africa." Paper presented at the Annual Meeting of the Population Association of America, San Diego, CA in April.

Sobek, M., L. Cleveland, S. Flood, M. King, S. Ruggles and M. Schroeder. 2011. Big Data: Large-Scale Historical Infrastructure from the Minnesota Population Center, *Historical Methods* 44: 61-68. doi:10.1080/01615440.2011.564572.

https://www.ons.gov.uk/file?uri=/methodology/classificationsandstandards/sexualidentityguidanceandprojectdocumentation/2011censusassessmentsexualorientation_tcm77-181168.pdf

ONS, 2009a. Measuring sexual identity – a guide for researchers.

<http://www.ons.gov.uk/ons/guide-method/measuring-equality/equality/sexual-identity-project/guidance/measuring-sexual-identity--a-guide-for-researchers.pdf>

ONS, 2009b. 'Trans Data Position Paper'.

<http://www.ons.gov.uk/ons/guide-method/measuring-equality/equality/equality-data-review/trans-data-position-paper.pdf>

ONS, 2016a. 2021 Census topic consultation.

<https://www.ons.gov.uk/census/censustransformationprogramme/consultations/the2021censusinitialviewoncontentforenglandandwales>

Specific reports on sexual identity and gender identity
Sexual identity topic report and the Research plan for sexual identity
Gender identity topic report and the Research plan for gender identity

ONS, 2016b. Results from household surveys.

<https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/sexuality/bulletins/integratedhouseholdsurvey/2015-10-01>.

Приложение А. Затраты на проведение переписи

Страна	Год	Затраты	На душу населения
Австралия ₁	2011	\$445,000,000 австр. долл.	19.75 долл. на 1 чел.
Камбоджа ₂	2018	11.8млн.	
Канада ₁	2011	\$630,373,000 кан. долл.	18.25 на 1 чел.
Китай ₃	2011	\$1.34 млрд.	1 долл. на 1 чел.
Финляндия ₃	2011	Около 1 млн. Евро	
Германия ₄	2011	710 млн. евро	
Индия ₅	2010	\$446 млн. долл.	
Ирландия ₁	2011	63,519,000 евро	14.12 евро на 1 чел.
Кения ₆	2009	80 млн. долл.	
Мозамбик ₇	2017	75 млн. долл.	
Непал ₈	2011	15 млн. долл.	
Нидерланды ₉		1.4 млн. евро	
Новая Зеландия ₁	2011	90,332,087 новозеланд. долл.	20.41 долл. на 1 чел.
Польша ₁₀	2010	138,184,507 долл.	
Шотландия ₁	2011	57,035,027 фунтов стерлингов	10.84 фунтов стерлингов на 1 чел.
Уганда ₁₁	2014	Около 76 млн. долл.	
Соединенное Королевство ₁	2011	482 млн. фунтов стерлингов	8.66 фунтов стерлингов на 1 чел.
США ₃	2010	13 млрд. долл.	42.11 долл. на 1 чел.

Источники

- 1- Statistics New Zealand (2012). *Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy*. Wellington: Statistics New Zealand.
- 2- Deervew Marc GL (2016) The Cambodia Population Census: Leaving no one behind. *The Phnom Pen Post*. July 6.
<http://www.phnompenpost.com/analysis-and-op-ed/cambodia-population-census-leaving-no-one-behind>
- 3- *The Economist*. (2011). "Costing the count." June 2. <http://www.economist.com/node/18772674>
- 4- Der Spiegel Online. (2011). "Germany launches first census since reunification." May 9.
<http://www.spiegel.de/international/germany/the-count-germany-launches-first-census-since-reunification-a-761497.html>
- 5- Chandramouli, D. (2015). Census of India: Lessons learnt and the way ahead. Presentation at the UN EGM on Strengthening the Demographic Evidence Base for the Post-2015 Development Agenda. October 5-6.
<http://www.un.org/en/development/desa/population/events/pdf/expert/23/Presentations/EGM-S2-Chandramouli%20presentation.pdf>
- 6- The Coast Week. "Kenyan government kicks off preparations for census." September 13, 2016.
<http://www.coastweek.com/3933-Kenyan-government-kicks-off-preparations-for-census.htm>
- 7- 2017 Census to cost US75 million.
<http://clubofmozambique.com/news/mozambique-2017-population-census-to-cost-us75-million/>
- 8- Malla, Uttam Narayan (2011). "Managing population census 2011 of Nepal." Nepal Constitution Foundation.
http://www.ncf.org.np/upload/files/658_en_2441.pdf
- 9- Statistics Netherlands (2014). *Dutch Census 2011: Analysis and Methodology*. The Hague: Statistics Netherlands.
<https://www.cbs.nl/NR/rdonlyres/5FDCE1B4-0654-45DA-8D7E-807A0213DE66/0/2014b57pub.pdf>
- 10- Dygaszewicz, Janusz (2014). "Modern Census Polish Case Study." GUS e-census presentation. Washington, D.C. July 31-August 1.
http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_088794.pdf
- 11- New Vision staff reporter (2014). "National census to cost sh190b." May 19.
http://www.newvision.co.ug/new_vision/news/1340887/national-census-cost-sh190b

Приложение В. Выборочный охват гармонизированных переменных IPUMS

ХАРАКТЕРИСТИКИ ДОМОХОЗЯЙСТВ

География

- Первый административный уровень
- Второй административный уровень
- Город/село

Жилье

- Количество комнат
- Доступ к туалету
- Строительные материалы
- Возраст постройки
- Жилая площадь

Коммунальные услуги

- Электричество
- Вода
- Канализация
- Топливо
- Отопление

Удобства

- Автомобили
- Стиральная машина
- Телевидение
- Компьютер
- Телефон

Прочее

- Владение домом или землей
- Количество смертей
- Количество международных мигрантов
- Состав семьи и домохозяйства

ЛИЧНЫЕ ХАРАКТЕРИСТИКИ

Основные демографические

- Возраст/Год рождения
- Пол
- Семейное положение
- Возраст вступления в брак
- Степень родства с главой домохозяйства

ИНДИВИДУАЛЬНЫЕ ХАРАКТЕРИСТИКИ

Миграция

- Место прежнего проживания
- Количество лет проживания в нынешнем месте жительства

Рождаемость/смертность

- Количество родившихся детей
- Количество живых детей
- Смертность родителей

Этническая принадлежность

- Место рождения
- Страна рождения
- Гражданство
- Год иммиграции
- Религия
- Раса
- Этническая группа
- Разговорный язык
- Родной язык

Образование

- Посещение школы
- Грамотность
- Уровень образования
- Количество лет обучения

Рабочая сила

- Статус занятости
- Род деятельности
- Отрасль
- Класс рабочего
- Количество часов работы
- Общий доход
- Доход от заработной платы
- Источник средств к существованию

Инвалидность

- Наличие инвалидности
- Вид инвалидности
- Причина инвалидности

Приложение С. Ценные переменные IPUMS-International www.ipums.org/international

Команда IPUMS, имея десятки лет опыта работы с микроданными, разработала более 30 ценных переменных для дополнения каждой выборки. Такие переменные можно отнести к четырем видам: технические, географические, сводные и указатели.

Технические переменные: вид записи, страна, год, идентификатор выборки IPUMS, серийный номер домохозяйства, количество индивидуальных записей в домохозяйстве, вес домохозяйства, количество подвыборок, статус кварталов групп, континент, регион страны, проживание на первом и втором административном уровне и факторы расширения (веса выборки – для домохозяйств и людей).

Географические переменные: город/село, регион мира, географический субнациональный уровень 1, географический субнациональный уровень 2 и файлы с границами ГИС, которые должны соответствовать географическим единицей, гармонизированные по названию или физическому пространству, в соответствии с предпочтениями исследователя. См.: <https://international.ipums.org/international/gis.shtml>

Сводные переменные домохозяйства и семьи: классификация домохозяйства, количество семей в домохозяйстве, количество женатых пар в домохозяйстве, количество матерей в домохозяйстве, количество отцов в домохозяйстве, место нахождения главы домохозяйства, количество лиц, не являющихся родственниками, состав членов семьи, количество своих членов семьи в домохозяйстве, количество собственных детей в домохозяйстве, количество собственных детей в возрасте до 5 лет в домохозяйстве, возраст собственного старшего ребенка в домохозяйстве и возраст собственного младшего ребенка в домохозяйстве.

Переменные - указатели для определения совместно проживающих супругов, детей и их родителей: место матери, отца и супруга в домохозяйстве, правило по определению связи родителей и супругов, возможная мачеха, возможный отчим, мужчина с двумя или более женами и вторая жена или жена более высокого порядка и т. д. Наличие переменных-указателей является одним из наиболее ценных дополнений в интегрированную базу микроданных IPUMS, поскольку такие переменные –указатели способствуют проведению анализа детей по характеристикам их матерей или отцов, а также мужей по характеристикам их жен и наоборот (Sobek and Kennedy 2009). Легко проводить анализ рождаемость по собственным детям, поскольку каждый массив данных IPUMS по домохозяйствам уже привязывает матерей к их детям, проживающим совместно с ними (MOMLOC) и для того, чтобы поместить характеристики матери к записям каждого ребенка можно использовать функцию «Attach Characteristics» в системе IPUMS. Для биологических матерей важно отфильтровать при помощи переменной STEPMOM=0, также предоставляемой IPUMS. https://international.ipums.org/international-action/variables/MOMLOC#description_section.

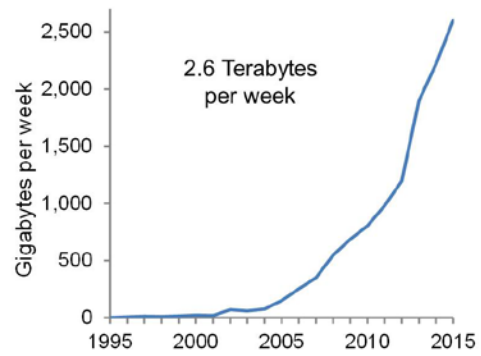
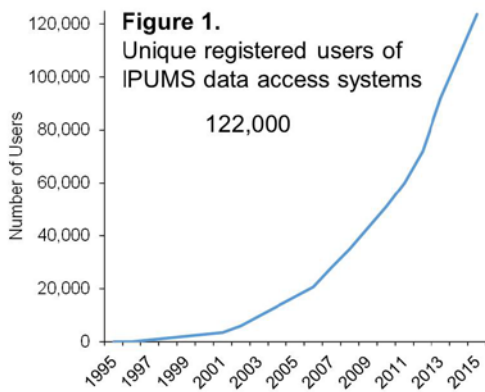


Figure 2. IPUMS Data Dissemination

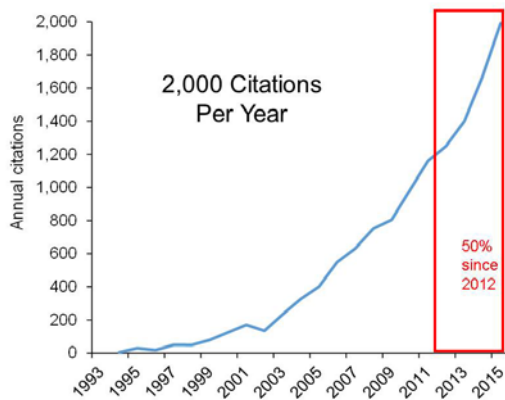


Figure 3. Citations of IPUMS Data in Google Scholar, 1994-2015

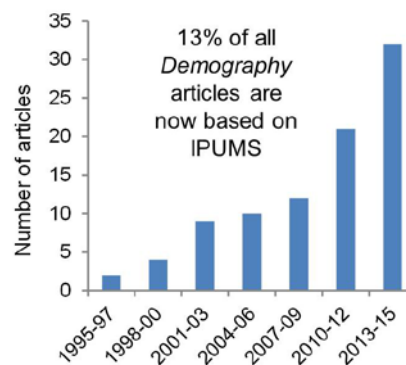


Figure 4. Demography articles based on IPUMS Data

Table 1. IPUMS Integrated Microdata Samples <https://international.ipums.org> - 86 countries 675 million person records (+ = 2016 launch)

Census	%	Persons	Census	%	Persons	Census	%	Persons	Census	%	Persons
1970 Argentina	2	466,892	1960 Chile	1	88,184	1962 France	5	2,320,901	2005 Indonesia-cont.	0.5	1,090,892
1980	10	2,667,714	1970	10	890,481	1968	5	2,487,778	2010	10	23,603,049
1991	10	4,286,447	1982	10	1,133,062	1975	5	2,629,456	2006 Iran	2	1,299,825
2001	10	3,626,103	1992	10	1,335,055	1982	5	2,631,713	2011+	2	1,482,000
2010	10	3,966,245	2002	10	1,513,914	1990	4.2	2,360,854	1997 Iraq	10	1,944,278
2001 Armenia	10	326,560	1982 China	1	10,039,191	1999	5	2,934,758	1971 Ireland	10	296,878
2011	10	301,831	1990	1	11,835,947	2006 (RR)	33	19,973,287	1979	10	337,686
1971 Austria	10	749,894	2000+	1	11,804,000	2011	33	20,541,337	1981	10	344,291
1981	10	756,556	1964 Colombia	2	349,652	1970 Germany	5	3,094,845	1986	10	355,020
1991	10	780,512	1973	10	1,988,831	1971 DR	25	4,089,856	1991	10	353,149
2001	10	803,471	1985	10	2,643,125	1981 DR	25	4,278,563	1996	10	365,323
2011	10	839,501	1993	10	3,213,657	1987	5	3,160,224	2002	10	410,688
1991 Bangladesh	10	10,580,904	2005	10	4,006,168	1984 Ghana	15	1,309,352	2006	10	440,314
2001	10	12,442,115	1963 Costa Rica	6	82,345	2000	10	1,894,133	2011	10	474,535
2011	5	7,205,720	1973	10	186,762	2010	10	2,466,289	1972 Israel	10	315,608
1999 Belarus	10	990,706	1984	10	241,220	1971 Greece	10	845,483	1983	10	403,474
2009+	10	941,000	2000	10	381,500	1981	10	923,108	1995	10	556,365
1976 Bolivia	10	461,699	2011	10	430,082	1991	10	951,875	2001 Italy	5	2,990,739
1992	10	642,368	2002 Cuba	10	1,118,767	2001	10	1,028,884	1982 Jamaica	10	223,667
2001	10	827,692	1960 Dominican Rep	6.6	201,556	2011+	10	1,057,000	1991	10	232,625
1981+ Botswana	10	97,000	1970	6.8	272,090	1983 Guinea	10	457,837	2001	10	205,179
1991+	10	133,000	1981	8.5	475,829	1996	10	729,071	2004 Jordan	10	510,646
2001+	10	169,000	2002	10	857,606	1971 Haiti	10	434,869	1969 Kenya	3.3	659,310
2011+	10	202,000	2010	10	943,784	1982	2.5	128,770	1979	5	1,033,769
1960 Brazil	5	3,001,439	1962 Ecuador	3	136,443	2003	10	838,045	1989	5	1,074,098
1970	5	4,953,759	1974	10	648,678	1970 Hungary	5	515,119	1999	5	1,407,547
1980	5	5,870,467	1982	10	806,834	1980	5	536,007	2009	10	3,841,935
1991	5.8	8,522,740	1990	10	966,234	1990	5	518,240	1999 Kyrgyzstan	10	476,886
2000	6	10,136,022	2001	10	1,213,725	2001	5	510,502	2009	10	564,986
2010	5	9,693,058	2010	10	1,448,233	2011+	5	497,000	1974 Liberia	10	150,256
1985 Burkina Faso	10	884,797	1986+ Egypt	15	6,799,000	1983 India - NSSO	0.1	623,494	2008	10	348,057
1996	10	1,081,046	1996	10	5,902,243	1987	0.1	667,848	1987 Malawi	10	798,669
2006	10	1,417,824	2006	10	7,282,434	1993	0.1	564,740	1988	10	991,393
1998 Cambodia	10	1,141,254	1992 El Salvador	10	510,760	1999	0.1	596,688	2008	10	1,341,977
2008	10	1,340,121	2007	10	574,364	2004	0.1	602,833	1970 Malaysia	2	175,997
1976 Cameroon	10	736,514	1984 Ethiopia	10	3,404,306	2010+	0.1	460,000	1980	2	182,601
1987	10	897,211	1994	10	5,044,598	1971 Indonesia	0.5	634,642	1991	2	347,892
2005	10	1,772,359	2007	10	7,434,086	1976	0.2	281,170	2000	2	435,300
1971 Canada	1	214,019	1966 Fiji Islands	10	47,579	1980	5	7,234,577	1987 Mali	10	785,384
1981	2	486,875	1976	10	57,214	1985	0.3	605,858	1998	10	991,330
1991	3	809,654	1986	10	72,158	1990	0.5	912,544	2009	10	1,451,856
2001	2.5	801,055	1996	10	77,382	1995	0.3	718,837			
2011+	3	926,000	2007	10	84,323	2000	10	20,112,539	(continued)		

Table 1 (continued) IPUMS Integrated Microdata Samples <https://international.ipums.org> - 86 countries 675 million person records (+ = 2016 launch)

Census	%	Persons	Census	%	Persons	Census	%	Persons	Census	%	Persons
1960 Mexico	1.5	502,800	1993 Peru	10	2,206,424	2008 Sudan	17	5,066,530	1963 Uruguay	10	256,171
1970	1	483,405	2007	10	2,745,895	1970 Switzerland	5	312,538	1975	10	279,994
1990	10	8,118,242	1990 Philippines	10	6,013,913	1980	5	317,803	1985	10	295,915
1995	0.4	332,061	1995	10	6,864,758	1990	5	342,797	1996	10	315,920
2000	11	10,099,182	2000	10	7,417,810	2000	5	364,086	2006	6	256,866
2005	10	10,284,550	1978+ Poland	10	3,577,000	1988 Tanzania	10	2,310,424	2011	10	328,425
2010	10	11,938,402	1988+	10	3,894,000	2002	10	3,732,735	1971 Venezuela	10	1,158,527
2015+	10	11,292,000	2002+	10	3,824,000	2011+	10	4,497,000	1981	10	1,441,266
1989 Mongolia	10	190,631	2011+	5	2,000,000	1970 Thailand	2	772,169	1990	10	1,803,953
2000	10	243,725	1981 Portugal	5	492,289	1980	1	388,141	2001	10	2,306,489
1982 Morocco	5	1,012,873	1991	5	491,755	1990	1	485,100	1989 Vietnam	5	2,626,985
1994	5	1,294,026	2001	5	517,026	2000	1	604,519	1999	3	2,368,167
2004	5	1,482,720	2011	5	528,870	1970+ Trinidad & T	10	69,000	2009	15	14,177,590
1997 Mozambique	10	1,551,517	1970 Puerto Rico	1	27,212	1980+	10	105,000	1990 Zambia	10	787,461
2007	10	2,047,048	1980	5	160,219	1990+	10	113,000	2000	10	996,117
2001 Nepal	11	2,583,245	1990	5	177,655	2000+	10	112,000	2010	10	1,321,973
1960 Netherlands	1.2	143,251	2000	5	189,828	2011+	10	117,000	Candidates for 2017 and beyond:		
1971	1.2	159,203	2005 (PRCS)	1	35,416	1985 Turkey	5	2,554,364	2010 round censuses		
2001	1.2	189,725	2010	1	36,032	1990	5	2,864,207	More countries/places		
1971 Nicaragua	10	189,469	1977 Romania	10	1,937,021	2000	5	3,444,456	Angola		
1995	10	435,728	1992	10	2,238,578	1991 Uganda	10	1,548,460	Australia		
2005	10	515,485	2002	10	2,137,967	2002	10	2,497,449	Belgium		
2006 Nigeria - GHS	0.1	83,700	2011+	10	1,990,000	2001 Ukraine	10	4,889,288	Benin		
2007	0.1	85,183	1991 Rwanda	10	742,918	1991 UK	1	541,894	Bulgaria		
2008	0.1	107,425	2002	10	843,392	2001	3	1,843,525	Cote d' Ivoire		
2009	0.1	77,896	2012+	10	1,038,369	1960 USA	1	1,799,888	Finland		
2010	0.1	72,191	1980 Saint Lucia	10	11,451	1970	1	2,029,666	Guatemala		
1973 Pakistan	2	1,453,332	1991	10	13,382	1980	5	11,343,120	Guinea Bissau		
1981	10	8,433,058	1988 Senegal	10	700,199	1990	5	12,501,046	Honduras		
1998	10	13,102,024	2002	10	994,562	2000	5	14,081,466	Japan		
1997 Palestine	10	259,191	2004 Sierra Leone	10	494,298	2005 (ACS)	1	2,878,380	Korea, Republic of		
2007	10	227,067	2002 Slovenia	10	179,632	2010	1	3,061,692	Madagascar		
1960 Panama	5	53,553	1996 South Africa	10	3,621,164	1963 Uruguay	10	256,171	Mauritius		
1970	10	150,473	2001	10	3,725,655	1975	10	279,994	Myanmar		
1980	10	195,577	2007	2	1,047,657	1985	10	295,915	Namibia		
1990	10	232,737	2011	10	4,418,594	1996	10	315,920	Niger		
2000	10	284,081	2008 South Sudan	7	542,765	2006	6	256,866	Nigeria PES		
2010	10	341,118	1981 Spain	5	2,084,221	2011	10	328,425	Papua New Guinea		
1962 Paraguay	5	90,236	1991	5	1,931,458	1971 Venezuela	10	1,158,527	Russia		
1972	10	233,669	2001	5	2,039,274	1981	10	1,441,266	Tunisia		
1982	10	301,582	2011	10	4,107,465	1990	10	1,803,953	Turkmenistan		
1992	10	415,401				2001	10	2,306,489	Yemen		
2002	10	516,083							Zimbabwe, etc.		

Table 2. IPUMS Integrated variable: Employment Status (EMPSTAT) - Detailed codes, case count view for 8 selected samples										
	ISO Country Code	BD	KH	CN	IN	ID	IR	KG	VN	
	Year	2011	2008	1990	2004	2010	2006	2009	2009	
Code	Label									
0	NIU (not in universe)	1,117,354		3,264,675	59,373	4,587,786	214,837	173,245	3,746,089	
	ACTIVE (in labor force)									
100	EMPLOYED , not specified	2,123,058	692,961	6,775,149				222,282		
110	At work				226,309	10,272,827	339,173		7,890,417	
111	At work, and 'student'									
112	At work, and 'housework'									
113	At work, and 'seeking work'									
114	At work, and 'retired'									
115	At work, and 'no work'									
116	At work, and other situation									
117	At work, family holding, not specified									
118	At work, family holding, not agricultural									
119	At work, family holding, agricultural									
120	Have job, not at work in reference period				2,901	257,792	21,262		45,629	
130	Armed forces									
131	Armed forces, at work									
132	Armed forces, not at work in reference									
133	Military trainee									
140	Marginally employed									
200	UNEMPLOYED , not specified	56,753		61,553	10,672	287,300	62,495	25,880	56,105	
201	Unemployed 6 or more months		2,780							
202	Worked fewer than 6 months, permanent job									
203	Worked fewer than 6 months, temporary job									
210	Unemployed, experienced worker									
220	Unemployed, new worker		9,090							
230	No work available								61,460	
240	Inactive unemployed				1,138	1,026,714				
300	INACTIVE (not in labor force)	2,057,472				7,170,630		143,579		
310	Housework	1,851,082	47,041	729,891	105,647		305,437		557,280	
320	Unable to work/disabled			295,175					151,033	
321	Permanent disability				4,876					
322	Temporary illness				539					
323	Disabled or imprisoned									
330	In school		330,375	436,317	155,130		222,874		816,753	
340	Retirees and living on rent		4,523		6,662					
341	Living on rents									
342	Living on rents or pension						55,332			
343	Retirees/pensioners									
344	Retired			230,000						
345	Pensioner									
346	Non-retirement pension									
347	Disability pension									
348	Retired without benefits									
350	Elderly									
351	Elderly or disabled									
360	Institutionalized									
361	Prisoner									
370	Intermittent worker								20,418	
371	Not working, seasonal worker									
372	Not working, occasional worker									
380	Other income recipient									
390	Inactive, other reasons		1,937	43,187	29,586		65,781		800,359	
391	Too young to work									
392	Dependent		251,414							
999	UNKNOWN/MISSING						12,634		32,047	