



**Экономический и
Социальный Совет**

Distr.
GENERAL

ECE/CES/GE.41/2009/23
14 October 2009

RUSSIAN
Original: ENGLISH

ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

Группа экспертов по переписям населения
и жилищного фонда

Двенадцатое совещание
Женева, 28-30 октября 2009 года
Пункт 5 предварительной повестки дня

**МАТЕРИАЛЫ ПЕРЕПИСИ ДЛЯ УДОВЛЕТВОРЕНИЯ ПОТРЕБНОСТЕЙ
ПОЛЬЗОВАТЕЛЕЙ**

Представление микроданных и метаданных переписей населения для своевременной
интеграции и распространения в рамках инициатив IPUMS-Евразия и ИЕСМ,
2010-2014 гг.*

Записка Миннесотского центра по народонаселению, Миннеаполис, США и
Центра демографических исследований, Автономный университет Барселоны

«Без всякого сомнения, проект IPUMS International соответствует четырем Основным принципам,
заявленным на КЕС (Конференция европейских статистиков) 2007 года. Проект был отмечен на
КЕС 2007 как показательный пример передовой практики. Проведенная мною оценка
подтверждает его статус оптимального решения для реализации хранилища данных.
Он действительно является наилучшим практическим решением по созданию хранилища
международных статистических данных».

—Денис Тревин (2007 г.)

www.hist.umn.edu/~rmccaa/ipums-global/trewin_report_2007.pdf

* Средства для проведения исследований выделены: Национальным институтом здравоохранения США, грант HD047283 для осуществления проекта по гармонизации микроданных переписей населения Европы и Азии (IPUMS-Евразия), и Европейским Союзом в рамках разработки научно-исследовательских инфраструктур, FP6- 026033, для осуществления проекта по гармонизации интегрированных микроданных европейских переписей населения (ИЕСМ).

I. КРАТКОЕ ОПИСАНИЕ.

1. В настоящее время на веб-сайте проекта IPUMS-International Миннесотского центра по изучению народонаселения (МЦН) размещены интегрированные, анонимизированные микро- и метаданные переписей населения из 44 стран. На портале проекта по интеграции микроданных европейских переписей населения (IECM) Автономного университета Барселоны можно получить доступ к данным из 12 европейских стран, которые приведены в соответствие с европейскими стандартами. Размер этих банков данных может удвоиться в течение ближайших пяти лет благодаря, с одной стороны, эффективному и плодотворному сотрудничеству национальных статистических служб, а с другой, вследствие постоянной финансовой поддержки проектов американскими и европейскими научными организациями. Несмотря на то, что доступ к данным предоставляется только официально зарегистрированным научным и политическим деятелям, число пользователей составило 3000 человек из 76 стран. В рамках инициативы IECM успешно проведена гармонизация 35 выборок из европейских переписей. Планируется пополнять эту базу данных, в частности, за счет переписей раунда 2010 года.

2. В настоящей работе мы хотим предложить методические рекомендации по приведению результатов переписей в соответствие с потребностями пользователей, в том числе, для своевременного предоставления микро- и метаданных переписей всем заинтересованным исследователям и, в частности, проектам IPUMS/IECM, для эффективной разработки. Следует принять во внимание, что микроданные и документация, полученные проектами IPUMS/IECM за последние десять лет, имели самые разнообразные формы. Безусловно, когда метаданные и микроданные документально оформлены в соответствии с общепринятыми нормами, процесс обработки значительно сокращается, а также уменьшается число возможных ошибок. Статистическим службам настоятельно рекомендуется заполнять краткую форму (см. Приложение А), содержащую характеристики предоставляемых микро- и метаданных переписей населения.

3. Микроданные следует передавать в виде зашифрованных выполняемых файлов, а пароль пересылать координатору проекта отдельно по электронной почте или факсом. Метаданные можно передавать в виде отсканированных изображений, но также в форматах ASCII, CSp, IMP, NESSTAR, SPSS, STATA, SAS, в виде электронных таблиц или файлов документов, гипертекстов DDI (Инициатива по документации данных, обратите внимание, что NESSTAR совместим с DDI) или в виде других новых стандартов. Необходима также документация на национальном языке (языках). Если она переведена на английский язык, следует направлять и переводы. Если нет, МЦН организует перевод и подготовит переведенные тексты в простом формате ASCII. Копии переводов предоставляются соответствующим национальным статистическим органам.

II. ПРОЕКТ IPUMS-INTERNATIONAL: “ПЕРЕДОВАЯ ПРАКТИКА”.

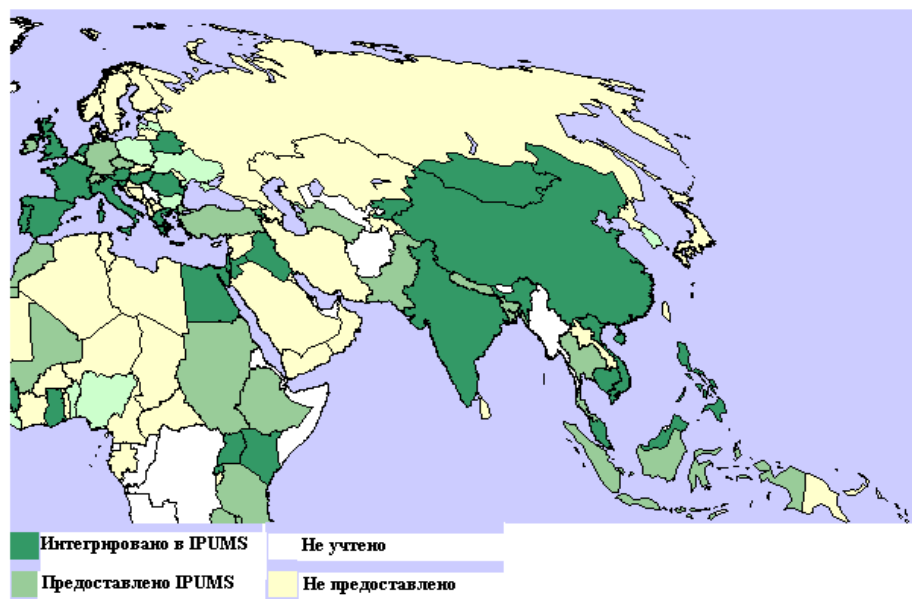
4. В эпиграфе, взятом из отчета Дениса Тревина, в кратком виде представлена блестящая оценка, данная им проекту IPUMS-International, используемым в нем инструментам, стратегиям и процедурам обработки и распространения анонимизированных выборок микроданных переписей населения, которые осуществляются в Миннесотском центре по изучению народонаселения (МЦН). Будучи

председателем Целевой группы ЕЭК ООН по разработке рекомендаций, касающихся передовой практики выпуска микроданных и защиты их конфиденциальности, г-н Тревин является общепризнанным авторитетом в этой области. Благодаря его исключительно положительной оценке предпринимаемых проектом IPUMS-International мер по защите данных, мы уверенно вступили во второе десятилетие развития проекта. Если Вы незнакомы с мерами, предпринимаемыми проектом IPUMS-International по защите данных и соблюдению конфиденциальности, Вы можете обратиться к нашему докладу, подготовленному для рабочей сессии ЕЭК ООН/Евростата по соблюдению конфиденциальности статистических данных:

www.unecce.org/stats/documents/2005.11.confidentiality.htm (см. рабочий документ 5), который позднее был опубликован в Монографиях по официальной статистике (МакКей и Эстив, 2006 г.).

5. В настоящее время на сайте проекта IPUMS-International www.ipums.org/international можно бесплатно получить доступ к 130 анонимизированным, интегрированным, высокоточным выборкам микроданных переписей населения. На портале Интегрированных микроданных европейских переписей (IECM) (www.iecm-project.org) размещены выборки из 12 европейских стран. Размер этих баз данных может в течение пяти последующих лет увеличиться вдвое благодаря средствам, выделенным Национальным научным фондом и Национальным институтом здравоохранения США на развитие проекта по 2014 год включительно, а также интенсивной и эффективной поддержке со стороны национальных статистических служб. Более 3000 исследователей из 76 стран получили разрешение на доступ к микроданным через указанные выше сайты. Они используют эти данные для проведения временного и пространственного сравнительного анализа. Потенциальные пользователи, которым нужны достоверные сведения по переписям, направляются на веб-сайты национальных статистических органов, Статистического отдела ООН или других официальных источников.

Таблица 1. IPUMS-Евразия: Состояние микроданных по странам



6. Эта массивная инфраструктура данных охватывает 44 страны (см. таблицу 1), включая страны Европы и Азии: Армения, Австрия, Беларусь, Камбоджа, Китай, Франция, Греция, Венгрия, Ирак, Израиль, Италия, Иордан, Кыргызская Республика, Монголия, Нидерланды, Палестина, Филиппины, Португалия, Румыния, Словения, Испания, Соединенное Королевство и Вьетнам. База данных IPUMS-International насчитывает более 279 миллионов анонимизированных интегрированных персональных записей, представляющих 77 миллионов домашних хозяйств. По плану, база данных пополнится в 2010 году выборками из четырех стран Европы и Азии (Непал, Пакистан, Швейцария и Таиланд) и шести стран Африки и Америки (Куба, Мали, Перу, Сент-Люсия, Сенегал и Танзания). В течение последующих пяти лет мы предполагаем ввести в базу выборки домашних хозяйств из переписей раунда 2010 года, а также микроданные из таких стран, как Бангладеш, Бельгия, Болгария, Чешская Республика, острова Фиджи, Германия, Индонезия, Ирландия, Польша, Турция, Туркменистан и Украина.

7. В Америке проведена интеграция данных для приблизительно десятка стран (56 переписей), и в следующей пятилетке планируется интегрировать еще столько же. Что касается Африки, здесь на настоящий момент интегрированы и подготовлены для распространения данные всего 13 переписей (7 стран), но мы ожидаем ускорения этого процесса в скором будущем (МакКей, Эстив, Раглз и Собек, 2006 г.).

III. ПРОЕКТ ИЕСМ – ЕВРОПЕЙСКАЯ ИНТЕГРАЦИЯ.

8. Проект интеграции микроданных европейских переписей (ИЕСМ) стартовал в 2005 году благодаря совместным усилиям Центра демографических исследований, Миннесотского центра народонаселения и 17 европейских статистических служб, направленным на координацию, гармонизацию и распространение интегрированных данных европейских переписей. Проект ИЕСМ, взявший за основу модель проекта IPUMS-Европа, выступил первой региональной инициативой по привлечению научного сообщества и официальных органов Европы к процессу гармонизации микроданных переписей населения и тем самым стал ориентиром для подобных инициатив в других регионах мира. Проект ИЕСМ получил средства на развитие в рамках мероприятий по созданию научно-исследовательских инфраструктур Шестой рамочной программы Европейского Союза. Финансирование выделялось на конкурсной основе, что привело к образованию трех взаимосвязанных проектов: СИЕСМ (Координация интегрированных микроданных европейских переписей), ДИЕСМ (Распространение интегрированных микроданных европейских переписей) и НИЕСМ (Гармонизация интегрированных микроданных европейских переписей), осуществлявшихся с 2005 по 2009 гг. В течение этого периода коллектив проекта ИЕСМ выполнял координационные задачи в Европе по упорядочению и совершенствованию компетенции представителей национальных статистических органов и специалистов по переписям. С этими целями были проведены четыре практических семинара: в 2005 году в Барселоне, в 2006 году в Париже, в 2007 году в Лиссабоне и в 2008 году в Барселоне.

9. Кроме того, коллектив проекта ИЕСМ вел интенсивную работу по гармонизации микроданных европейских переписей населения и разработал собственный веб-сайт (www.iecm-project.org), с которого пользователи получают доступ к микроданным европейских переписей (аналогично тому, как это делается на веб-сайте проекта IPUMS-International, где массив данных включают и европейские показатели, разработанные в

проекте ИЕСМ) и к документации, представленной национальными статистическими службами, а также к дополнительным, характерным для Европы метаданным, которые на веб-сайте IPUMS-International отсутствуют. Со времени открытия веб-сайта ИЕСМ в 2008 году на нем были размещены 35 выборок из 12 стран из переписей раундов 1960-2000 гг. (см. таблицу 2). Эти выборки содержат более 45 миллионов персональных записей, 14 миллионов домашних хозяйств, 166 гармонизированных показателей и 11 внутриевропейских признаков.

Таблица 2. IPUMS-Europe/ИЕСМ: Интегрированные выборки (июнь 2009 г.)

Страна	Год переписи	Выборка %	Домохозяйства (кол-во)	Лица (кол-во)
Австрия	1971	10	264,655	749,894
	1981	10	283,693	756,556
	1991	10	310,099	780,512
	2001	10	341,035	803,471
Беларусь	1999	10	385,508	990,706
Франция	1962	5	748,917	2,320,901
	1968	5	815,699	2,487,778
	1975	5	915,624	2,629,456
	1982	5	969,632	2,631,713
	1990	4.2	949,893	2,360,854
Греция	1999	5	1,219,323	2,934,758
	1971	10	249,350	845,483
	1981	10	294,323	923,108
	1991	10	320,387	951,875
	2001	10	367,438	1,028,884
Венгрия	1970	5	172,831	515,119
	1980	5	211,355	536,007
	1990	5	219,389	518,240
	2001	5	227,252	510,502
Италия	2001	5	1,168,044	2,990,739
Нидерланды	1960	1.2	нет данных	143,251
	1971	1.2	нет данных	159,203
	2001	1.2	нет данных	189,725

Португалия	1981	5	179,409	492,289
	1991	5	214,155	491,755
	2001	5	258,843	517,026
Румыния	1977	10	619,904	1,937,021
	1992	10	728,846	2,238,578
	2002	10	732,016	2,137,967
Словения	2002	10	63,637	179,632
Испания	1981	5	нет данных	2,084,221
	1991	5	592,276	1,931,458
	2001	5	714,473	2,039,274
Соединенное Королевство	1991	1	215,761	541,894
	2001	3	нет данных	1,843,525
Всего			14,753,767	45,193,375

10. По мере вступления в инициативу новых стран коллектив проекта ИЕСМ будет продолжать вносить свою лепту в проекты IPUMS, конечно, при условии получения финансирования. Схема выделения средств в Седьмой рамочной программе Европейского Союза аналогична схеме, используемой в Шестой рамочной программе ЕС, и коллектив ИЕСМ будет принимать в ней участие на конкурсной основе. Однако, учитывая задачи научно-исследовательского центра, коим СДИ является, а также упрочение настоящей инфраструктуры, основные усилия будут направлены на расширение возможностей использования микроданных переписей, а также на пропаганду их научного потенциала. С этой целью сейчас в работе находятся несколько предложений, например, по обучению специалистов в области микроданных переписей в Европейской научной школе демографии и по проведению исследовательских разработок на основании микроданных переписей населения (Эстив и др., 2009 г.). Доктор Эстив, научный руководитель проекта ИЕСМ, получил значительные средства от Европейского научного совета с целью полнейшего использования возможностей микроданных переписей населения для изучения определяющих факторов семейной жизни в глобальном масштабе. Мы надеемся, что в ближайшее время, после того, как ученые-аналитики включают микроданные переписей в свои исследовательские программы, последуют и другие новые проекты. В этой связи следует упомянуть о проекте микроданных переписей населения Национального института демографических исследований, которым руководит профессор Патрик Фести. Что касается распространения данных, коллектив ИЕСМ продолжает широко информировать научную общественность о существовании таких данных. Так, была проведена специальная согласованная кампания, приуроченная к Европейской конференции по народонаселению, состоявшейся в июле 2008 года в Барселоне. Проекту были посвящены несколько заседаний, включая заседание по использованию данных, на котором рассматривалась первая научная работа, выполненная на базе интегрированных данных европейских переписей.

IV. ПОТРЕБНОСТЬ В КРАТКОМ ОПИСАНИИ ПЕРЕПИСЕЙ И МИКРОДАННЫХ: ФОРМА “А”.

11. Проекты IPUMS/IECM ставят перед собой очень смелые задачи. На самом деле, совсем недавно они показались бы сверхамбициозными и невыполнимыми. Да и сейчас, чтобы преуспеть на этом поприще, необходимо расширять сотрудничество с национальными статистическими службами. Будучи исследователями, мы хорошо понимаем, что органы официальной статистики перегружены постоянными запросами со стороны государственных учреждений, деловых кругов и общественности, требующими все более срочного исполнения. Поэтому мы готовы работать, как мы и делаем в последнее десятилетие, с метаданными и микроданными в любом виде без какой-либо специальной обработки или особых условий. Хотя нельзя не признать, что документальное оформление мета- и микроданных в соответствии с принятыми требованиями значительно улучшает процесс интеграции и сокращает количество возможных ошибок.

12. Для успешного использования данных необходима краткая и внушающая доверие документация в легкодоступном формате. Для получения краткого описания переписи и ее микроданных следует использовать форму А (см. Приложение А). Она должна заполняться специалистом по проведению переписи соответствующего национального статистического органа для каждого набора представленных данных. В Приложении В даны примеры заполненных форм переписей населения Испании 1981, 1991 и 2001 гг. Дополнительные примеры можно найти на сайте <https://international.ipums.org/international/samples.shtml>, щелкнув по названию страны, или на сайте <http://www.iecm-project.org/> (сначала нужно нажать на “Метаданные”).

13. Форма А имеет четыре раздела: описание переписи, характеристики выборки, выделяемые в микроданных единицы и определения единиц.

- (a) Описание переписи. Запрашиваются следующие сведения: официальное название, орган, проводящий перепись, генеральная совокупность, население де юре или де факто, день переписи, период сбора данных, количество и типы переписных листов, метод (методы) сбора данных, респонденты и охват.
- (b) Характеристики выборки: источник (как правило, национальный статистический орган, национальный архив данных или научно-исследовательская организация), план выборки, доля выборки (как для частных домашних хозяйств, так и для коллективных жилищ, т.к. они могут различаться—см. ниже), размер выборки (количество персональных записей) и краткое описание весов выборки.
- (c) Выделяемые в микроданных единицы (нужно указать «да/нет» и добавить любые необходимы комментарии): жилища, свободные жилые помещения, домашние хозяйства, отдельные лица, коллективные жилища, наименьшие географические единицы (название и код NUTS) и оседлое/неоседлое/особое население.
- (d) Определения единиц: жилища, частные домашние хозяйства, коллективные жилища и оседлое/неоседлое население или особые группы населения.

14. В случае необходимости в форму можно внести дополнительные пункты (например, модули, касающиеся эмиграции, рождаемости, медицинского страхования и т.д.). Если специалист не владеет английским языком, форму следует направлять в МЦН на национальном языке, где она будет переведена профессиональным переводчиком. Если форма А по Вашей стране уже размещена на веб-сайте проекта IPUMS-International, пожалуйста, проверьте правильность сведений по каждой переписи и, если нужно, направьте по электронной почте ipumsi@pop.umn.edu свои предложения, корректировки или замечания.

V. ТРЕБОВАНИЯ К МЕТАДАНЫМ

15. Метаданные в проектах IPUMS-International/IECM выполняют целый ряд задач. Большинство общих метаданных используются для точной обработки и оценки микроданных при вводе их в базы данных, а также для облегчения процесса гармонизации конкретных показателей. Для успешной интеграции микроданных и их эффективного использования учеными необходимы разные и полные метаданные (ЦСУ Канады, 2008 г.; а также см. МакКей и Томас, 2009 г.). Метаданные можно передавать в виде отсканированных изображений, но они должны быть также доступны в форматах ASCII, CSpPro, IMPS, NESSTAR, SPSS, STATA, SAS, в электронных таблицах, файлах документов или гипертекстовых файлах или сразу в нескольких из этих форматов. Если представленные документы не существуют в электронном виде, их сканируют, распределяют по странам и годам проведения переписи и размещают на веб-сайтах проектов IPUMS-International и IECM для обеспечения к ним простого доступа. Соответствующие статистические органы, а также национальные и международные научно-исследовательские организации могут получить копии отсканированных в МЦН или АУБ документов переписей на компакт-дисках или цифровых видеодисках.

16. В отношении метаданных мы ставим перед собой три цели. Во-первых, исследователи должны иметь простой доступ к оригиналам документации переписи на национальном языке. Как минимум, необходимы переписные листы, инструкции по проведению переписи или учебные брошюры и кодовые книги. Немаловажное значение для проекта IPUMS-International имеют также дополнительные метаданные, касающиеся организационной и подготовительной работы, а также непосредственно проведения переписи. Они вносятся в каталог и архивируются вместе со всеми остальными полученными документами. Если метаданные опубликованы, предпочтение отдается печатным копиям или документам PDF. Наша цель состоит в создании архива высококачественных файлов PDF для всех видов метаданных, касающихся микроданных переписи.

17. У национальных статистических служб запрашиваются следующие метаданные переписей населения:

- (a) Переписные листы.
- (b) Инструкции по проведению переписи (иногда называются как «Инструкции для счетчиков»).
- (c) «Кодовые книги» или «Словари данных» по каждому набору данных (определения структур записей, размещение показателей в столбцах и обозначения кодов. См. для примера файлы словаря данных «IMPS» Бюро

- цензов США), включающие административно-географическое деление, профессии и т.д.
- (d) Таблицы соответствия между системами кодирования в двух или более переписях или между переписью и международным стандартом (профессии, образование и т.д.). Эти таблицы очень помогают в случаях изменения административно-географического деления и при интеграции признаков уровня образования.
 - (e) Основные таблицы официальных результатов, опубликованных в Интернете, прессе или на компакт-дисках.
 - (f) Технические и методические отчеты по проведению переписи, концепциям, классификациям, сопоставимости, качеству, проведению контрольных обследований после переписи и т.д.
 - (g) Если микроданные представляются в виде выборки, следует дать подробное описание плана выборки. Если используется стандартный план проекта IPUMS-International, т.е. каждое n -ное хозяйство после произвольного старта, то не требуется никаких дополнительных документов (см. ниже характеристики микроданных). В ином случае полезными могут оказаться оценки ошибок выборки в масштабе абсолютных или относительных частот (например, когда частота выборки = 2, 5, 10, 15, 20, 25, 30, 35, 40, и 50 %) или ключевые показатели (или и то, и другое), такие как возраст, родственное отношение к первому лицу, образование и статус занятости. Следует отметить, что на сегодняшний день наиболее комплексную документацию по плану и ошибкам выборки представил Национальный институт статистики Мозамбика (Мегилл, 2007 г.).
 - (h) Файлы данных граничных условий, соответствующие закодированному в микроданных административному делению (NUTS1, NUTS2 и NUTS3) и пригодные для передачи исследователям. Если файлы данных граничных условий не предоставляются, мы создаем неофициальные файлы из имеющихся готовых источников.

18. Во-вторых, мы конструируем динамическую систему метаданных для каждого показателя, как интегрированного, так и негармонизированного, с целью облегчения сопоставления формулировки конкретного вопроса и соответствующих инструкций для счетчиков (на английском языке) для любой комбинации стран и переписей. Например, если исследователь хочет посмотреть, как сформулирован вопрос об экономической активности, ему нужно сначала выбрать нужные страны и годы проведения переписей. Затем на домашней странице нужно щелкнуть по “Variables” (показатели) и затем по “Select samples” (выделить выборки). После того, как нужные выборки выделены с помощью мыши, нужно щелкнуть по “submit sample selections” (подтвердить выбор), затем навести курсор на “person” (лицо), прокрутить вниз и нажать на “work” (работа), затем щелкнуть по “EMPSTAT” (или “Employment Status”) (статус занятости) и, наконец, нажать на “Enumeration text” (текст переписи). На экране появится текст выбранного вопроса и переписей на английском языке, включая формулировку вопроса на переписном листе и в инструкциях для счетчиков. Осуществляя прокрутку вниз, можно очень просто изучить все нюансы исходного текста по каждой из переписей. Для просмотра другого вопроса из этого же набора переписей, нужно вернуться на шаг назад или перейти на экран показателей и выбрать другой показатель, после чего щелкнуть по “Enumeration text” (текст переписи).

19. В-третьих, на основании подлинной исходной документации мы составляем подлинные метаданные, описывающие каждый интегрированный показатель следующим образом:

- (a) краткое определение и описание выбранного показателя,
- (b) наличие (перечень стран и переписей, включающих этот показатель),
- (c) общая сопоставимость (оттенки различных определений),
- (d) совокупность (население, которому адресуется вопрос),
- (e) учетный период (например, для экономической деятельности семь дней, прошлый месяц, год и т.д.)
- (f) различия в определениях конкретных признаков (например «занятый») и
- (g) обсуждение сопоставимости конкретных переписей, проведенных страной.

20. Исследователь вызывает эти страницы для просмотра простым нажатием на название показателя, как описано в предыдущем параграфе. Страницы создаются по запросу динамической системой метаданных. Воспроизводятся только сведения о сопоставимости для выделенных переписей.

21. Хотя мы отдаем предпочтение метаданным в электронном виде, принимаются также публикации или ксерокопии. Электронные файлы можно направлять электронной почтой или на компакт-дисках. Если понадобится перевод документов на английский язык, профессиональные переводчики выполняют перевод в простом текстовом формате. Во избежание потери материалов и экономии усилий лучше собрать их в единый комплект и направить курьерской почтой за счет проекта. Примеры оригинальных исходных документов европейских стран, участвующих в проекте, размещены на веб-сайте проекта ИЕСМ. Там же можно найти почти полный комплект электронных изображений всех видов документов.

22. Что касается структурированных метаданных (словарей данных, перечней кодов, определений, переписных форм и т.д.), можно использовать такие новые стандарты, как Data Documentation Initiative (Инициатива документации данных) (www.icpsr.umich.edu/DDI/codebook/) в формате NESSTAR и Microdata Toolkit (Инструментарий микроданных), разработанные Международной сетью обследований домашних хозяйств (<http://www.surveynetwork.org/home/>) и Всемирным банком, которые облегчают передачу информации в систему обработки проекта IPUMS-International. DDI представляет разметочную структуру, использующую расширяющий язык разметки (XML), который выделяет конкретные элементы, находящиеся в сопровождающей файл данных кодовой книге. Она включает идентифицирующую информацию о файле данных, характеристиках переписи или обследования, характеристиках выборки, определениях единиц, методологии, структурах файлов, содержании и структуре показателей, содержании вопросов и отношений к показателям, перечнях кодов и сопутствующих материалов либо в системе, либо через обращение к внешним документам.

23. Существующие с 2008 года новые версии DDI уже поддерживают сбор и перенаправление информации о комплексном процессе гармонизации, которая используется при построении интегрированных показателей. Мы надеемся, что в недалеком будущем сумеем предложить зарегистрированным исследователям, запрашивающим выборки микроданных, соответствующие их запросам кодовые книги,

построенные из баз метаданных, лежащих в основе интерфейса и системы извлечения IPUMS-International. Эти кодовые книги DDI могли бы работать непосредственно с таким программным обеспечением, как NESSTAR и Microdata Toolkit, которое поддерживает документацию DDI, а также обеспечивает подробную информацию, невозможную в стандартных метаданных статистических пакетов. Мы надеемся, что после решения проблем совместимости DDI станет возможной гибкая, соответствующая принятым стандартам структура для перемещения метаданных в систему IPUMS-International и из нее.

VI. ТРЕБОВАНИЯ К МИКРОДАНЫМ

24. В отношении микроданных у нас две цели: во-первых, непрерывно архивировать первичные данные, предоставляемые национальными статистическими службами, и, во-вторых, обеспечивать доступ зарегистрированным исследователям к высокоточным, анонимизированным, интегрированным и выполненным в соответствии с их требованиями выборкам домашних хозяйств. Мы предпочитаем, чтобы национальные статистические службы представляли нам защищенные копии (с удаленными именами, адресами и идентификационными номерами) полных исходных файлов (т.е. 100% данные), на основании которых мы могли бы последовательно и эффективно строить выборку при минимальной нагрузке на статистический орган. Более того, если в выборке обнаружатся дефектные записи, эту проблему можно очень просто решить не вменением, а заменой. Следует отметить, что все предоставляемые Миннесотскому центру по изучению народонаселения исходные файлы с микроданными архивируются под общей системой защиты ("Icebox") и ни при каких обстоятельствах не копируются никаким лицам или учреждениям. Как отмечено в отчете г-на Тревина, Миннесотский центр по изучению народонаселения стремится к обеспечению совершенной, незапятнанной системы защиты.

25. Следует отметить, что Европейское бюро статистики передало проекту полный комплект исходных файлов с микроданными. Тем не менее, если рассматривать проблему в мировом масштабе, то в течение первого десятилетия сотрудничества проекта IPUMS-International с партнерами обозначились четыре возможности представления микроданных (метками обозначены примеры):

- (a) Задача по архивированию полного комплекта исходных файлов и построению выборок поручается Миннесотскому центру по изучению народонаселения (38 стран).
- (b) Выборки создаются исключительно национальной статистической службой в соответствии с требованиями проекта IPUMS-International при наличии 100% микроданных (25 стран):
 - (i) Федеральное статистическое управление Германии: вся работа выполнена ФСУ Германии, включая переписи 1970 и 1987 гг. Федеративной Республики Германия и переписи 1971 и 1981 гг. Германской Демократической Республики.
 - (ii) ЦСУ Нидерландов (SN): переписи 1960 и 1971 гг. и выборка на базе регистра 2001г.— вся работа выполнена ЦСУ.
 - (iii) Федеральное статистическое управление (ФСУ) Швейцарии: ФСУ подготовило переписи 1971, 1981, 1991 и 2001 гг.

- (c) Выборки микроданных, переданные в открытое или ограниченное пользование исследователям, предоставляются также проекту IPUMS-International с выплатой или без выплаты лицензионной платы (12 стран):
 - (i) Национальное бюро статистики Китая (лицензионная плата за перепись 1982 г; перепись 1990 г. передана бесплатно).
 - (ii) Национальная организация статистических обследований Индии.
 - (iii) ЦСУ Канады (без лицензионной платы).
 - (iv) Офис национальной статистики Великобритании (без лицензионной платы).
- (d) Создание анонимизированных выборок поручается учреждению или отдельному эксперту под руководством национального статистического органа (6 стран):
 - (i) Национальный институт статистики и экономических исследований (INSEE) Франции: переписи 1962, 1968, 1975, 1982, 1990 и 1999 гг. подготовлены индивидуальным исследователем, работающим в INSEE, по контракту с Минессотским центром по изучению народонаселения и под контролем INSEE.
 - (ii) Национальный институт статистики и экономических исследований (INSSE) Румынии: работа по переписям 2002, 1992 и 1977 гг. выполнялась университетским научно-исследовательским институтом по контракту с Минессотским центром по изучению народонаселения и под контролем INSEE.

26. Независимо от способа представления данных проект IPUMS предлагает лицензионную плату в размере 5000 долларов США за набор данных, содержащих 1 миллион или более персональных записей, а также на покрытие расходов по комплектации микроданных и документации. Безусловно, каждый национальный статистический орган вправе устанавливать способ передачи данных, и проект всегда идет навстречу другим предложениям.

27. Чаще всего под «высокоточной» подразумевается 10% выборка домашних хозяйств (70 из 130 интегрированных массивов данных), за ней следует 5% выборка (28 выборок). Из 32 выборок с размером менее 5% восемь являются историческими и включают все сохранившиеся микроданные. Если нельзя получить 100% микроданные, мы предпочитаем систематические случайные выборки, составленные в соответствии со следующими простыми правилами:

- 1) Распределите файлы с микроданными по административным единицам, начиная с крупных, затем мелких вплоть до уровня переписных участков, жилищ, домашних хозяйств, семей и лиц.
- 2) После произвольного старта выберите каждое n -ное частное жилище (в 10% выборке каждое десятое).
- 3) В случае институциональных домашних хозяйств (или крупных частных домашних хозяйств, выделяемых единственно из-за их размера) после произвольного старта выберите каждое n -ное лицо с той же плотностью, что и в случае частных жилищ.

28. Систематическая случайная выборка основана на низкоуровневом географическом делении. При обеспечении представительного географического распределения

выборочных объектов они равны самой подробной географической стратификации с пропорциональным взвешиванием. Поскольку многие экономические и демографические характеристики тесно связаны с географическим местоположением, такая скрытая стратификация дает значительно более высокую точность, чем простая случайная выборка домашних хозяйств. В зависимости от степени связи страт, используемых для построения высокоточных выборок, с нужными показателями (например, сиротством, бедностью, безработицей и т.д.), полученные оценки этих показателей будут иметь меньше стандартных ошибок, чем при построении простой выборки записей (Даверн и др., 2009 г.).

29. Одним из основных преимуществ использования микроданных переписей являются географические возможности, позволяющие проводить субнациональный анализ без потери статистической значимости. Благодаря ограничениям, связанным с соблюдением конфиденциальности, из микроданных переписи следует исключать подробные географические признаки даже при условии ограниченного доступа, как в случае проектов IPUMS/IECM. Несмотря на усилия Европейского Союза, предпринимаемые им в вопросе стандартизации административного деления стран для статистических целей (деление NUTS), страны ЕС используют разные критерии определения географических признаков микроданных своих переписей. В сопоставительных целях для всех стран можно было бы считать пригодным уровень NUTS3 (в настоящее время только 5 из 12 представленных в проекте IPUMS стран ЕС дают такую информацию) наряду с дополнительными показателями из местных административных единиц, например, размером единицы или даже индивидуальными кодами крупных городских агломераций. При наличии показателя размера населенного пункта можно было бы проводить в выборках последовательное измерение городского/сельского местожительства. В Германии (предпочтительно) и Франции используются следующие категории размеров географических единиц:

<u>Германия</u>	<u>Франция</u>
1) 1 to 2,499 persons	1 to 4,999
2) 2,500 to 9,999	5 to 9,999
3) 10,000 to 49,999	10 to 19,999
	20 to 49,999
4) 50,000 to 99,999	50 to 99,999
5) 100,000 to 499,999	100,000 to 1,999,999
6) 500,000 or more	2,000,000 or more

30. Анонимизацию может осуществлять либо сам статистический орган, либо по его запросу Миннесотский центр по изучению народонаселения. Выдержки с микроданными предоставляются зарегистрированным исследователем под очень строгим правовым и административным контролем (МакКей и Эстив, 2006 г.; МакКей, Раглз и др., 2006 г.). Хотя мы разделяем мнение Андерсон и Финберга (2001 г.) о том, что построение выборки из наборов данных уже само «вносит дополнительную неточность, необходимую для защиты данных...», мы не намерены на этом останавливаться. Мы применяем шесть слоев технической защиты. Прежде всего, мы в большинстве последних переписей удаляем коды места прохождения переписи, проживания, работы или обучения в географических единицах, находящихся ниже порога размера населения 20000 человек. (Некоторые статистические органы используют более высокий порог, например, в СК он

равен 65000 человек). Во-вторых, в случае категориальных переменных, скорее всего, удаляется любое значение с повторяемостью менее 250 (ФСУ Германии применяет порог 2500). Такие значения записываются как «прочее», «отсутствует» или в случае сложных кодов самая правая цифра является нулем (и процесс повторяется). В-третьих, для непрерывных переменных, например, дохода или размера жилища в квадратных метрах, применяется верхнее и нижнее кодирование, способствующее отсечению шлейфов распределений по мере их «истончения». В-четвертых, удаляются некоторые показатели, особенно уязвимые для установления личностей, например, дата рождения. В-пятых, для создания дополнительной степени неточности небольшая часть домашних хозяйств переносится из географической единицы в соседнюю. И наконец, домашним хозяйствам присваивают уникальный случайный номер и перераспределяют.

VII. ЗАКЛЮЧЕНИЕ

31. Чтобы результаты переписи соответствовали потребностям пользователей, проекты IPUMS/IECM запрашивают огромный объем различных метаданных и микроданных. Однако затраченные усилия в значительной степени окупаются, освобождая статистические службы от многих обременительных задач и обязанностей по распространению микроданных среди исследователей. Более того, мы используем стандартные методы, применяемые в большинстве статистических служб мира, а, как известно, действовать сообща безопаснее. Изолированные статистические органы, нерегулярно распространяющие микроданные, подвергаются серьезному риску и ответственности, а также несут ощутимые кадровые и материальные затраты при сравнительно небольшом числе пользователей их данных. Проекты IPUMS/IECM предлагают значительную экономию за счет роста масштабов при достижении наивысших стандартов безопасности и распространения интегрированных метаданных и микроданных, в большой степени облегчающих проведение основательных научных исследований. Призываем статистические службы, участвующие в данных проектах, как можно скорее направлять мета- и микроданные переписей населения раунда 2010 года. Приглашаем все остальные страны рассмотреть возможность присоединения к инициативе IPUMS-Евразия.

Библиография

- Anderson, Margo and Stephen E. Fienberg. (2001). “*U.S. Census Confidentiality: Perception and Reality*,” International Statistical Institute Biennial Meeting (Seoul). (unpub.)

Марго Андерсон и Стивен Финберг (2001). «*Конфиденциальность переписи США: представление и реальность*», Совещание международного статистического института (Сеул) (не опубликовано).

- CES (2007), “Managing Statistical Confidentiality and Microdata Access: Principles and Guidelines on Good Practice”, published by the Conference of European of Statisticians: http://www.unece.org/stats/publications/Managing_statistical_confidentiality_and_microdata_access.pdf

КЕС (2007), «Управление статистической конфиденциальностью и доступом к микроданным: руководящие принципы добросовестной практики», опубликовано Конференцией европейских статистиков на сайте:

<http://www.unece.org/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf>

- Davern, Michael, Steven Ruggles, Tami Swenson, J. Trent Alexander and J. Michael Oakes. (2009) "*Drawing Statistical Inferences from Historical Census Data, 1850-1950*," Demography, 46(3):589-603.

Майкл Даверн, Стивен Раглз, Тами Свенсон, Трент Александр и Майкл Оукс (2009) «Статистические выводы из данных исторических переписей населения 1850-1950гг.», Демография, 46 (3), стр. 589-603.

- Esteve, A., García, J., Spijker, J., McCaa, R., (2009) "*Integrated European Census Microdata (IECM) Samples: Enhancing the study of ageing with high precision over-samples of the oldest-old*", in Work session on statistical data confidentiality, Luxembourg: Eurostat, pp. 407-416.

А. Эстив, Гарсиа, Спийкер, Р. МакКей (2009) «Выборки Интегрированных микроданных европейских переписей (IECM): повышение качества изучения старения за счет высокоточных дополнительных выборок самого пожилого населения», Рабочее заседание по конфиденциальности статистических данных, Люксембург, Евростат, стр. 407-416.

- McCaa, Robert and Albert Esteve. (2006). "*IPUMS-Europe: Confidentiality measures for licensing and disseminating restricted access census microdata extracts to academic users*," Monographs of official statistics: Work session on statistical data confidentiality. Luxembourg: Office for Official Publications of the European Communities, pp. 37-46.

Роберт МакКей и Альберт Эстив (2006) «IPUMS-Европа: меры по соблюдению конфиденциальности для лицензирования и предоставления выборок микроданных переписей ограниченного доступа научным сотрудникам высшей школы», Монографии официальной статистики. Рабочее заседание по конфиденциальности статистических данных, Люксембург, Бюро официальных публикаций ЕС, стр. 37-46.

- McCaa, Robert; Albert Esteve, Steven Ruggles and Matt Sobek. (2006) "*Using integrated census microdata for evidence-based policy making: the IPUMS-International global initiative*," African Statistical Journal, 2:83-100.

Роберт МакКей, Альберт Эстив, Стивен Раглз и Мэтт Собек (2006) «Использование интегрированных данных переписей населения для выработки политического курса, основанного на фактическом материале: всемирная инициатива IPUMS-International», Африканский статистический журнал, 2:83-100.

- McCaa, Robert; Steven Ruggles, Michael Davern, Tami Swenson, and Krishna Mohan, Palipudi. (2006) "*IPUMS-International High Precision Population Census Microdata Samples: Balancing the Privacy-Quality Tradeoff by Means of Restricted Access Extracts*," Privacy in Statistical Databases (New York: Springer), 375-382.

Роберт МакКей, Стивен Раглз, Майкл Даверн, Тами Свенсон и Кришна Мохан (2006) «Высокоточные выборки микроданных переписей населения в проекте IPUMS-International: Сохранение равновесия между конфиденциальностью и качеством за

счет использования ограниченного доступа», Конфиденциальность в статистических базах данных (Нью-Йорк: Шпрингер), 375-382.

- McCaa, Robert and Wendy Thomas. (2009) “*IPUMS-International: lessons from 10 years of archiving and disseminating census microdata*,” 57th Session International Statistical Institute, Durban, South Africa (unpub.) www.hist.umn.edu/~rmccaa/ipums-global (scroll to “IPM 100 Microdata Session” and click “IPUMS”).

Роберт МакКей и Венди Томас (2009) «*IPUMS-International: итоги десятилетнего опыта архивирования и распространения микроданных переписей*», 57-я сессия Международного статистического института, Дурбан, Южная Африка (не опубликовано) www.hist.umn.edu/~rmccaa/ipums-global (найдите «Заседание по микроданным IPM 100» и щелкните по “IPUMS”).

- Megill, David. (2007). “*Technical Documentationl for Public Use Microdata Samples Files for the 1997 Mozambique Census of Population and Housing*,” Instituto Nacional de Estatística, Maputo (unpub.). www.hist.umn.edu/~rmccaa/ipums-africa (click “electronic holdings,” scroll to Mozambique, and click “sample design” for the 1997 census.

Дэвид Мегилл (2007) «*Техническая документация к файлам с выборками микроданных открытого доступа для Переписи населения и жилищ Мозамбика 1997 года*», Национальный институт статистики, Мапуту (не опубликовано) www.hist.umn.edu/~rmccaa/ipums-africa (нажмите на «электронный фонд», найдите Мозамбик и щелкните по «плану выборки» переписи 1997 года.

- Statistics Canada. (2008) “*Metadata requirements for archiving structured data*,” Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS), Luxembourg.

ЦСУ Канады (2008) «*Требования к метаданным для архивирования структурированных данных*», Совместное рабочее совещание ЕЭК ООН/Евростат/ОЭСР по статистическим метаданным (METIS), Люксембург.

- Trewin, Dennis (2007). *A review of IPUMS-International*. (unpub.).

Денис Тревин (2007) «*Оценка проекта IPUMS-International*» (не опубликовано).

Приложение А. Форма для записи характеристик переписи и выборки

Инструкции: Дайте краткое описание каждой переписи и выборки микроданных.

Можно писать на национальном или английском языке. Не требуется никакого форматирования.

ФИО: _____ email: _____ дата: _____

Если другие переписи Вашей страны уже интегрированы, пожалуйста, проверьте их характеристики на сайте:

<https://international.ipums.org/international/samples.shtml>

Направляйте вопросы Роберту МакКею по адресу: rmccaa@umn.edu

Характеристики переписи (страна):

Название	
Учреждение, проводящее перепись	
Генеральная совокупность	
Население де юре или де факто	
Единица учета	
День переписи	
Период сбора информации	
Используемые переписные листы	
Метод сбора информации	
Респондент	
Охват	

Характеристики выборки микроданных

Источник микроданных	
План выборки	
Единица выборки	
Доля выборки	
Размер выборки (кол-во персональных записей)	
Веса выборки (дать описание)	

Выделяемые единицы ("да" = единица идентифицируется, в ином случае ставьте "нет")	
Жилища	
Свободные единицы	
Домашние хозяйства	
Отдельные лица	
Коллективные жилища	
Оседлое/неоседлое население	
Особые группы населения	
Наименьшие географические единицы в микроданных	
Определения единиц	
Жилища	
Частные домашние хозяйства	
Коллективные жилища	
Неоседлое население	
Особые группы населения	
Представленные метаданные (перечень файлов или названий бумажных копий документов)	
Переписные листы	
Инструкции/руководства по заполнению форм	
Словарь данных	
Кодовые книги (образование, профессии, отрасли, географические признаки т.д.)	
Таблицы соответствия (образование)	
Официальные результаты	
Технические и методические отчеты	
Отчет по контрольному обследованию после проведения переписи	
План выборки, ошибки выборки	
Файлы данных граничных условий (если есть)	

ПРИЛОЖЕНИЕ В. ХАРАКТЕРИСТИКИ ПЕРЕПИСИ И ВЫБОРКИ НА ПРИМЕРЕ ИСПАНИИ

https://international.ipums.org/international/sample_designs/sample_designs_es.shtml

Характеристики переписи			
Год переписи	1981	1991	2001
Название	Перепись населения и жилищного фонда Испании 1981 года	Перепись населения и жилищного фонда Испании 1991 года	Перепись населения и жилищного фонда Испании 2001 года
Учреждение, проводящее перепись	Национальный институт статистики (INE)	Национальный институт статистики (INE)	Национальный институт статистики (INE)
Генеральная совокупность населения	Перепись включает лиц с постоянным местом жительства на территории страны, а также лиц, находящихся на территории страны в момент переписи, без пропусков и дублирования. Лица с постоянным местом жительства в Испании составляют постоянное население [обладающее правами]. Лица, находящиеся на территории страны в момент переписи, составляют фактическое население [de facto].	Перепись включает лиц с постоянным местом жительства на территории страны, а также лиц, находящихся на территории страны в момент переписи, без пропусков и дублирования. Лица с постоянным местом жительства в Испании составляют постоянное население [обладающее правами]. Лица, находящиеся на территории страны в момент переписи, составляют фактическое население [de facto].	Резиденты: лица, на момент переписи постоянно проживающие в Испании. Сюда не входят лица, находившиеся в Испании на момент переписи, но не проживающие там постоянно. Исключение составляют граждане и резиденты, временно находящиеся за границей.
Население де юре или де факто	Де факто	Де факто	Де юре
День переписи	1 марта 1981 года	1 марта 1991 года	1 ноября 2001 года
Период сбора информации	—	С 1 марта по 1 апреля 1991 года	Два месяца
Переписные листы	Два вида форм: одна для населения, проживающего в частных жилищах, и другая – для населения, проживающего в коллективных жилищах.	Использовано пять видов форм: общий вопросник, индивидуальный вопросник, вопросник для проживающих в коллективном жилище, вопросник по второму жилью без зарегистрированных жильцов и вопросник по строениям 1990 года.	Четыре вида форм: жилой фонд, переписные данные, данные по домашнему хозяйству и персональные данные.

Метод сбора информации	Переписные листы предназначены для самостоятельного заполнения. Задача счетчика, в основном, сводится к раздаче пустых и сбору заполненных форм, а также проверке правильности заполнения.	Переписные листы предназначены для самостоятельного заполнения. Задача счетчика, в основном, сводится к раздаче пустых и сбору заполненных форм, а также проверке правильности заполнения.	Данные собирает счетчик, но их можно передать по телефону или через Интернет. Формы разработаны таким образом, чтобы респонденты могли их заполнить без посторонней помощи.
Контроль за процессом переписи	Руководители групп по-очереди проверяют все переписные листы. В случае необходимости счетчик должен лично навестить соответствующее жилище для сбора недостающей или отсутствующей информации.	—	Переписчики собирают информацию из разных источников и проверяют ответы, указанные в формах.
Характеристики выборки микроданных			
Источник микроданных	Интегрированные микроданные европейских переписей	Интегрированные микроданные европейских переписей	Интегрированные микроданные европейских переписей
План выборки	Систематическая стратифицированная выборка. Доля выборки провинций Алава, Гипускоа, Наварра и Бискайя в пять раз превышала долю других провинций.	Систематическая стратифицированная выборка	Систематическая стратифицированная выборка
Единица выборки	Жилище	Жилище	Жилище
Доля выборки	5%	5%	5%
Размер выборки (персональные записи)	2084221	1931458	2039274
Веса выборки	Рассчитаны учреждением, проводившим перепись. Настоятельно рекомендуется использовать веса из-за повышения доли выборки в некоторых провинциях.	Рассчитаны учреждением, проводившим перепись, и должны применяться при выполнении всех видов анализа.	Коэффициент расширения = 20.

Выделяемые единицы			
Жилища	да	да	да
Свободные единицы	да	да	да
Домашние хозяйства	да	да	да
Отдельные лица	да	да	да
Коллективные жилища	да	да	да
Пристанища [Alojamientos]	да	да	да
Наименьшая географическая единица	Населенные пункты с населением 20000 и более человек	Населенные пункты с населением 20000 и более человек	Населенные пункты с населением 20000 и более человек, сформированные в МЦН
Определения единиц			
Жилища	Жилище – это отдельное в плане конструкции и независимое помещение, которое, учитывая способ его строительства, переделки, изменения или приспособления, считается пригодным для жилья, или даже если это не так, в момент переписи фактически заселено людьми.	Жилище – это отдельное в плане конструкции и независимое помещение, которое, учитывая способ его строительства, переделки, изменения или приспособления, считается пригодным для жилья, или даже если это не так, в момент переписи фактически заселено людьми.	Жилище – это отдельное в плане конструкции и независимое помещение, которое, учитывая способ его строительства, переделки, изменения или приспособления, считается пригодным для жилья, или даже если это не так, в момент переписи фактически заселено людьми.
Частные домашние хозяйства	Семья [домашнее хозяйство] – это группа лиц, как правило, связанных родственными отношениями, которые проживают вместе, обычно занимая все жилище. В домашнее хозяйство включаются домашняя прислуга, ночующая в жилище, и гости, принадлежащие семье.	Домашнее хозяйство – это группа людей, проживающих в одном жилище, которые несут совместные затраты по его использованию и/или питанию. Учитываются как домашние хозяйства, включающие одно лицо, так и хозяйства, состоящие из нескольких человек.	Группа людей, проживающих в одном семейном жилище и необязательно имеющих общие расходы. Семейные жилища – это жилища, в которых проживает одно или несколько лиц, обычно, но необязательно связанных родством, которые не являются коллективными жилищами [см. определение ниже].

Коллективные жилища	Коллективные жилища – это жилища или строения, предназначенные для проживания людей, не являющихся семьей, которые подчинены общему режиму или связаны общими целями или интересами.	Коллективные жилища – это жилища или строения, предназначенные для проживания людей, не являющихся семьей, которые подчинены общему режиму или связаны общими целями или интересами.	Коллективные жилища – это жилища или строения, предназначенные для проживания людей, не являющихся семьей, которые подчинены общему режиму или связаны общими целями или интересами.
Пристанища [Alojamientos]	[Alojamiento] "Пристанищем" называют единицы, которые не соответствуют жилищам, потому что являются мобильными, полупостоянными или временными, или потому что они не предназначены для проживания, но являются заселенными на момент переписи.	[Alojamiento] "Пристанищем" называют единицы, которые не соответствуют жилищам, потому что являются мобильными, полупостоянными или временными, или потому что они не предназначены для проживания, но являются заселенными на момент переписи.	Пристанище – это семейное жилище, которое является мобильным, полупостоянным или временным или даже не предназначенным для проживания, но заселенным на момент переписи.