

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

Женева, 28–30 сентября 2016 года

Пункт 7 предварительной повестки дня

**Возможные виды использования новых источников  
данных (например, «больших данных»)  
в целях переписей****Три примера инновационных источников данных  
для переписи населения в Испании в 2021 году****Записка Национального статистического института Испании<sup>1</sup>***Резюме*

Испания планирует провести в 2021 году первую в истории перепись населения на основе регистров. В качестве основного источника данных для этой переписи будет использован регистр населения, «Падрон», который ведется с 1996 года.

Для успешного достижения наших целей в последние годы были изучены многие источники данных. Среди прочего, в настоящем докладе будут представлены возможности, возникающие при использовании трех современных источников данных на достаточно подробном географическом уровне: налоговых данных, потребления электроэнергии и информации с мобильных телефонов.

В документе также будет описан каждый из этих источников данных и приведены реальные примеры информации, которую можно получить с их помощью.

**Ключевые слова:** перепись населения на основе регистров; источники данных.

<sup>1</sup> Подготовлено Хорхе Л. Вега, Антонио Аргуэсо и Кармен Тейхейро.



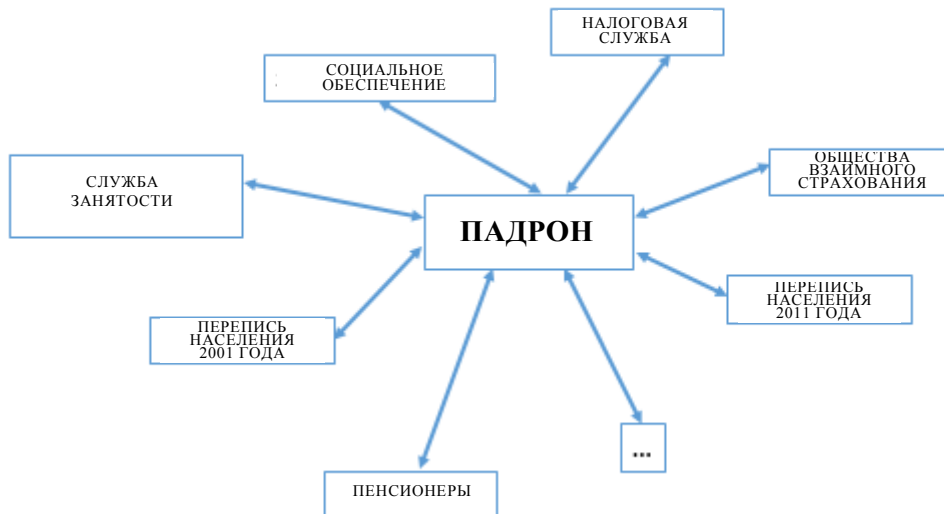
## I. Введение

1. Методология предыдущих переписей все больше основывалась на использовании административных данных, и в 2021 году Испания планирует провести перепись, для которой регистры населения будут служить основным источником<sup>2</sup>.

2. Для успешного осуществления подобного проекта необходим должным образом функционирующий регистр населения, который может обеспечивать взаимосвязь с другими источниками данных. В этом отношении Испания достигла высокого результата, поскольку она располагает одним из наиболее передовых регистров населения («Падрон»), который ведется с 1996 года. Механизм его работы постоянно совершенствуется благодаря деятельности всех специалистов, участвующих в его работе.

3. «Падрон» включает в себя систему цифровых идентификаторов, обеспечивающих взаимосвязь с другими источниками данных, а также списки полных имен и дат рождения, что представляет большую ценность для обеспечения этой взаимосвязи в том случае, когда использование цифровых кодов не дает удовлетворительных результатов, а также для обеспечения прямого доступа к данным переписей (например, для установления родства).

4. Чтобы найти значение различных переменных переписи, Испания планирует связать с регистром населения многочисленные источники данных, а именно: регистры безработных, пенсионеров, договоров, гражданских служащих, иностранцев, регистры социального обеспечения, уровня образования, а также демографическую статистическую информацию, данные предыдущих переписей и т.д.



5. Внутренние исследования об имеющихся источниках данных показали, что в некоторых случаях информация со значительным потенциалом, содержание которой представляет интерес, может также находиться в открытом доступе, однако публикация этих данных на личном уровне невозможна. Причины

<sup>2</sup> В 2001 году всеобщая перепись населения Испании была проведена с использованием данных регистров населения. В ходе переписи 2011 года использование регистров возросло, а охват проведенной полевой операции составил 8% населения.

могут варьироваться от соображений конфиденциальности до трудностей с обеспечением взаимосвязи на уровне микроданных, однако это не лишает эти источники данных значимости для целей переписи населения.

6. В настоящем документе будет показано, каким образом эти два теоретических недостатка, отнюдь не исключающие возможность использования перечисленных выше источников данных, способствуют достижению результата, представляющего гораздо большую ценность, чем данные традиционных переписей. Мы приведем три конкретных примера источников данных, которые планируется использовать в Испании для проведения предстоящей переписи.

## II. Примеры источников данных

### A. Уровень дохода

7. В Испании в области статистики налажено эффективное сотрудничество между учреждениями. Наглядным примером этого служит обмен многочисленными статистическими данными между Национальным статистическим институтом (НСИ) и Налоговой службой Испании, которые позволяют НСИ иметь доступ к информации о доходах на уровне переписных участков<sup>3</sup>, а также к другим статистическим данным, таким как процентиля уровня дохода и доля населения переписного участка, которое живет за чертой бедности.

8. Из соображений конфиденциальности обеспечить доступ к данным об уровнях дохода населения на индивидуальном уровне невозможно, однако тот факт, что эта информация доступна на уровне переписных участков, ни в коей мере не обесценивает этот источник данных.

9. Информация на уровне переписных участков представляет большую ценность. Если соотнести ее с другими постоянными переписями, появится множество возможностей, несомненно представляющих большой интерес как для исследователей, так и для населения.

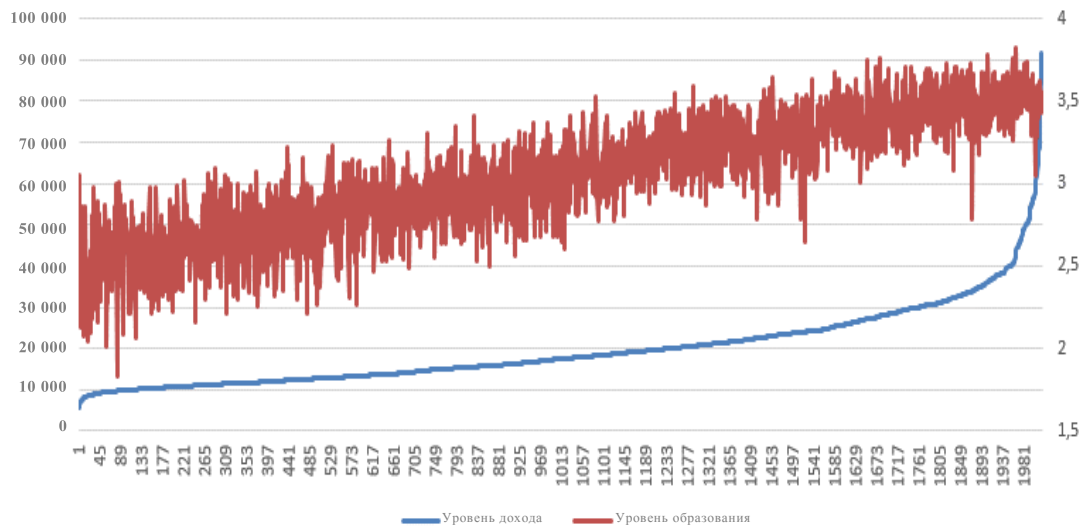
10. Так, можно увидеть, существует ли взаимосвязь между уровнем дохода переписного участка и уровнем образования его населения. Выше ли уровень дохода в секторах с более низким уровнем безработицы? Существует ли связь между уровнем дохода сектора и средним размером домохозяйства? Связан ли уровень дохода со средним возрастом в секторе?

---

<sup>3</sup> Для проведения выборов Испания разделена на 36 000 однородных переписных участков.

Рис. 1

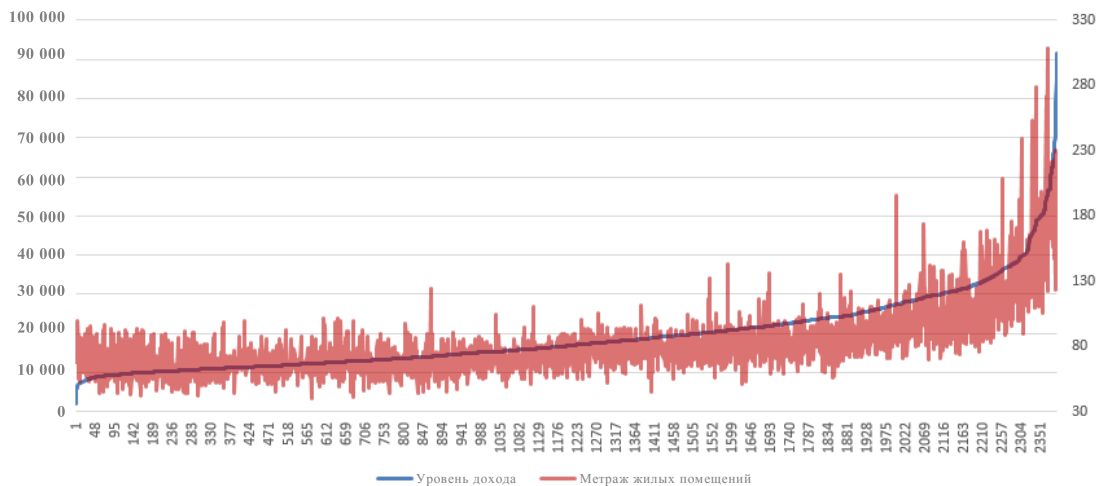
**Корреляция между средним уровнем дохода и средним уровнем образования в 2 000 переписных участков Мадрида**



11. Корреляция между средним уровнем дохода (левая шкала) и средним уровнем образования (правая шкала) в 2 000 секторов в районе города Мадрида. Переписные участки расположены в порядке возрастания их среднего уровня дохода.

Рис. 2

**Корреляция между средним уровнем дохода и средним метражом жилых помещений в 2 000 переписных участков Мадрида**



12. Корреляция между средним уровнем дохода (левая шкала) и средним метражом жилых помещений (правая шкала) в 2 000 секторов в районе города Мадрида. Переписные участки расположены в порядке возрастания их среднего уровня дохода.

13. Мы убеждены в том, что обеспечение широкого доступа к этой информации впервые в одном месте будет положительно встречено всеми пользователями – как для проведения исследований, так и для обеспечения большей степени детализации данных (на географическом уровне и в плане числа показателей).

## **В. Потребление электроэнергии**

14. Одна из наиболее характерных особенностей переписи заключается в том, что она предоставляет информацию о категориях жилых помещений (заселенных, сезонных и пустующих), расположенных на территории всей страны. Чтобы получить эту информацию, счетчики проводят кропотливую работу на местах, которая, несомненно, имеет субъективный элемент.

15. Из различных источников данных, изучаемых сейчас в рамках подготовки к проведению следующей переписи, можно выделить показатели потребления электроэнергии. Имеющаяся информация<sup>4</sup>, обработанная соответствующим образом, позволит составить новую классификацию домохозяйств в зависимости от уровня потребления электроэнергии.

16. Первый этап применяемой методики – координация территориальной информации из кадастра, предыдущих переписей и адресов из регистра населения. Именно таким образом можно создать жилищную базу данных и знать общее количество жилых помещений (VT) в регионе и их расположение.

17. Регистр населения позволит определить, какие жилые помещения в секторе мы относим к категории заселенных жилых помещений (VP). Информация о потреблении электроэнергии<sup>5</sup> будет использоваться для определения того, какие жилые помещения были заняты на протяжении всего года или хотя бы какого-то промежутка времени (VP+VS). И наконец, имея значения слагаемых, можно вычислить количество пустующих жилых помещений (VV), поскольку должно выполняться равенство  $VT = VP + VS + VV$ .

---

<sup>4</sup> Компании электроснабжения обязаны предоставлять в налоговые органы информацию о количестве потребленной электроэнергии по определенным адресам за один год.

<sup>5</sup> На самом деле потребление электроэнергии также позволяет получить информацию о некоторых пустующих жилых помещениях (тех, в которых установлены счетчики электроэнергии), но их можно отфильтровать, установив порог потребления.

18. Эта информация обобщена в представленной ниже схеме:



19. Хотя на первый взгляд эта концепция может показаться простой и понятной, привязка данных к территориальной информации всегда сложна, а результаты кажутся менее удовлетворительными по сравнению с теми, что были получены с помощью сопоставления информации о людях из различных источников.

20. К настоящему времени было проведено несколько проверок, а сравнение между информацией, полученной после переписи в 2011 году, и данными потребления электроэнергии показало, что использование данных о потреблении электроэнергии оказалось вполне удовлетворительным и даже более достоверным.

21. Использование информации о потреблении электроэнергии на индивидуальном уровне осложняется трудностями, появляющимися в ходе привязки данных, однако это не исключает возможности использования такого рода информации для составления некоторых показателей в переписных участках.

22. В общем, задача заключается в расчете новых показателей в подрайонах муниципальных образований, например процент жилых помещений с потреблением на уровне 10-го перцентиля или ниже, процент домов с потреблением от 10-го до 25-го перцентиля и т.д.

### **С. Мобильные телефоны**

23. Информация, связанная с показателями мобильности (место работы/учебы, связанное население и т.д.), была наиболее востребованной в ходе проведения последней переписи. Есть несколько способов использования этой информации: организация коммуникаций, существующих между различными районами, знание реального веса населения в пределах одной географической зоны, определение «рынков труда» и т.д.

24. Компании мобильной связи имеют в своем распоряжении большое количество такой информации, которая может оказаться очень интересной для следующей переписи, поскольку каждый раз, когда мы звоним, отправляем сооб-

шение или подключаемся к сети, мы оставляем след на ближайшей к нашему местоположению антенне.

25. Из трех рассмотренных в этом документе примеров источников данных этот является наиболее инновационным. Сегодня мы пытаемся получить информацию с мобильных телефонов, однако на настоящий момент здесь нет никакой гарантии. Сейчас вся имеющаяся у нас информация основана на экспериментальном проекте, проводимом только одной компанией.

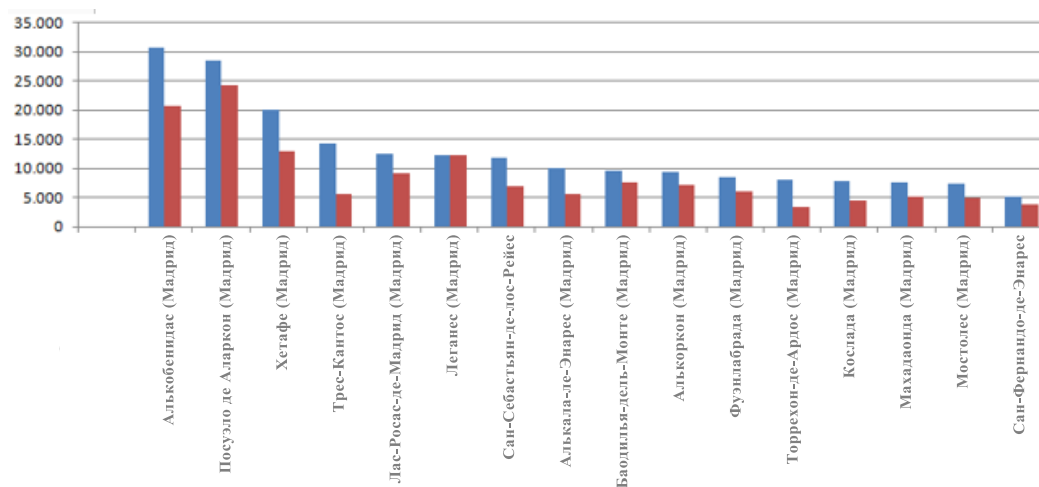
26. В первую очередь, используя предоставленную операторами информацию, Национальный статистический институт будет определять переписной участок, в котором проживает человек, а также целевой переписной участок (для работы или учебы) для каждого мобильного телефона.

27. Вот некоторые конкретные результаты на уровне секторов переписных участков, которые могут описываться с помощью этой информации:

- Результат 1: Матрица ежедневных перемещений между переписными участками за определенный день года (например, один день в ноябре). Ячейка «строка А – столбец В» матрицы покажет число людей, которые в среднем ездят из области А в область В, оставаясь там в течение нескольких часов выше определенного порога.

Рис. 3

**Сравнение места поездок населения Мадрида. Данные переписи населения в 2011 году в сопоставлении с данными мобильных телефонов (экспериментальное исследование)**



- Результат 2: Матрица населения переписных участков в разное время в определенный день (например, один день в ноябре). Этот результат может показать то, что уже было найдено другими организациями и что пользуется высоким спросом у пользователей, т.е. данные о численности населения сектора в дневное и ночное время.
- Результат 3: Матричная таблица связанного населения для определенного количества рассматриваемых дней. В ячейке «строка А – столбец В» будет указано число людей, которые живут в секторе А, но в течение определенного количества часов могут находиться в секторе В. Идея заключается в том, чтобы рассчитать эту матрицу за примерно 10 конкретных дней года: 5 дней в июле/августе и еще 5 дней в другое время года (Пасхальная неделя, Рождество и т.д.).

28. Есть один более тонкий момент, который следует иметь в виду в связи с этими результатами: в том виде, в котором она описана выше, эта процедура измеряет перемещение мобильных телефонов, но может быть применен некоторый поправочный коэффициент, который позволит использовать ее для измерения населения.

### **III. Выводы**

29. То, что регистр населения и другие источники, в которых содержатся основные данные переписи, улучшились по качеству и увеличились по охвату за последние годы, позволит Испании в 2021 году провести перепись на основе регистров.

30. Использование административных регистров позволяет добиться значительно большей глубины информации, чем традиционные переписные листы, где список показателей был уменьшен из-за ограниченности места на переписном листе. Еще одно преимущество этого метода – возможность периодически публиковать информацию по мере ее необходимости, а также в любое время адаптировать ее с учетом требований.

31. Благодаря новому методу существует возможность того, что даже те источники данных, которые не показывают связи на уровне микроданных или не могут допускать опубликования данных о физических лицах из-за соображений конфиденциальности, но которые имеют информативное содержание, могут также стать частью содержания переписи и таким образом дополнить количество доступной пользователям информации.

---