



Европейская экономическая комиссия

Конференция европейских статистиков

**Группа экспертов по переписям населения
и жилищного фонда**

Семнадцатое совещание

Женева, 30 сентября – 2 октября 2015 года

Пункт 4 предварительной повестки дня

Статистические картограммы: возможности и проблемы

Инновационные продукты для распространения географической информации (часть II)

Записка Национального института статистики Испании

Резюме

Возможности, открывающиеся благодаря наличию геопространственной информации в рамках переписи 2011 года, огромны. В течение последних месяцев мы продолжали разрабатывать новые продукты, использующие такую информацию и содействующие удовлетворению выявленных потребностей пользователей. Настоящий документ, который является второй частью документа, представленного в прошлом году, содержит подробную информацию о самых последних разработках.

Кроме того, в настоящем документе поясняются некоторые аспекты, не упомянутые в первой части, включая используемые правила конфиденциальности и алгоритм, который делит территорию на ячейки со стороной до 50 метров в зависимости от размера выборки населения.



I. Введение

1. Успехи, достигнутые в последние годы во всех секторах благодаря геопространственной информации, не вызывают сомнений. Становится все более и более популярным использование систем, в которых сохраняются пользовательские координаты, а информация представляется в виде карт для более интуитивного восприятия.

2. В рамках переписи – статистической операции, дающей очень подробную информацию в этой сфере, – нельзя упускать возможность использования географической информации. Ниже приводятся примеры, свидетельствующие о важности наличия геопространственной информации в контексте переписи:

- сеть супермаркетов выбирает место для размещения своего нового магазина;
- местный орган власти определяет наиболее подходящее место для строительства школы или аэропорта;
- в наличии имеется подробная информация о населении, проживающем в пределах определенного радиуса.

II. Новые опубликованные географические продукты

3. В представленном в прошлом году документе¹ упоминался ряд продуктов, содержащих связанную с переписью географическую информацию, включая карты, таблицы заданного формата, файлы микроданных и т.д. Все они размещены на микросайте переписи 2011 года².

4. В этом году мы продолжили разработку новых продуктов и учли высказанные пользователями мнения относительно изменения функциональных свойств существующих продуктов.

5. В частности, были смягчены требования к конфиденциальности информации, с тем чтобы предоставить пользователям больший объем информации. Более подробно эти новые правила конфиденциальности рассматриваются в разделе 4 настоящего документа.

A. Географические тематические карты

6. В прошлом году мы включили в нашу систему распространения данных переписи функцию визуализации 7 географических карт разных уровней.

7. На карте пользователи могут напрямую выбрать географический район, который они хотели бы проанализировать. В зависимости от уровня масштабирования можно получить 3 различные карты:

- ячейки со стороной в 100 км (10 000 кв. км);
- ячейки со стороной в 10 км (100 кв. км);
- ячейки со стороной 1 км (1 кв. км).

¹ http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2014/mtg1/WP22_s5_Spain.pdf.

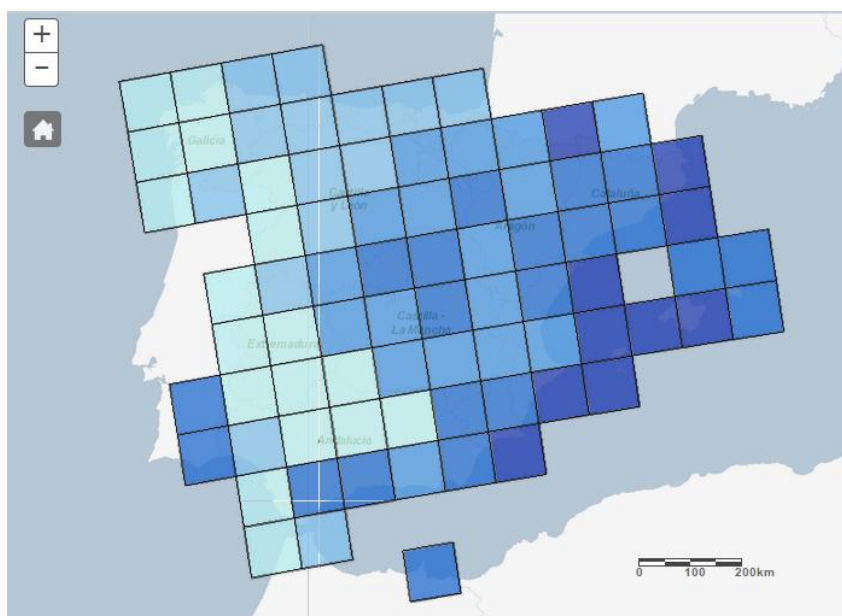
² http://www.ine.es/en/censos2011_datos/cen11_datos_inicio_en.htm.

На сеточном уровне были опубликованы следующие 7 карт:

- общая численность населения;
- средний возраст населения;
- процентная доля иностранных граждан;
- общее количество жилых единиц;
- процентная доля сезонного и пустующего жилья;
- общее число домохозяйств;
- средний размер домохозяйств.

Рис. 1

Процентная доля иностранных граждан в разбивке по квадратам сетки



В. 25 карт заданного формата с размером ячейки 1 кв. км

8. В начале 2015 года была опубликована единственная печатная публикация материалов переписи – небольшая брошюра под названием «Испания: какая она? 25 карт, рассказывающих о каждом кв. км страны».

Рис. 2
Обложка данной публикации



9. Эта публикация содержит 25 карт, составленных с помощью информации по представляющим наибольший интерес переменным переписи:

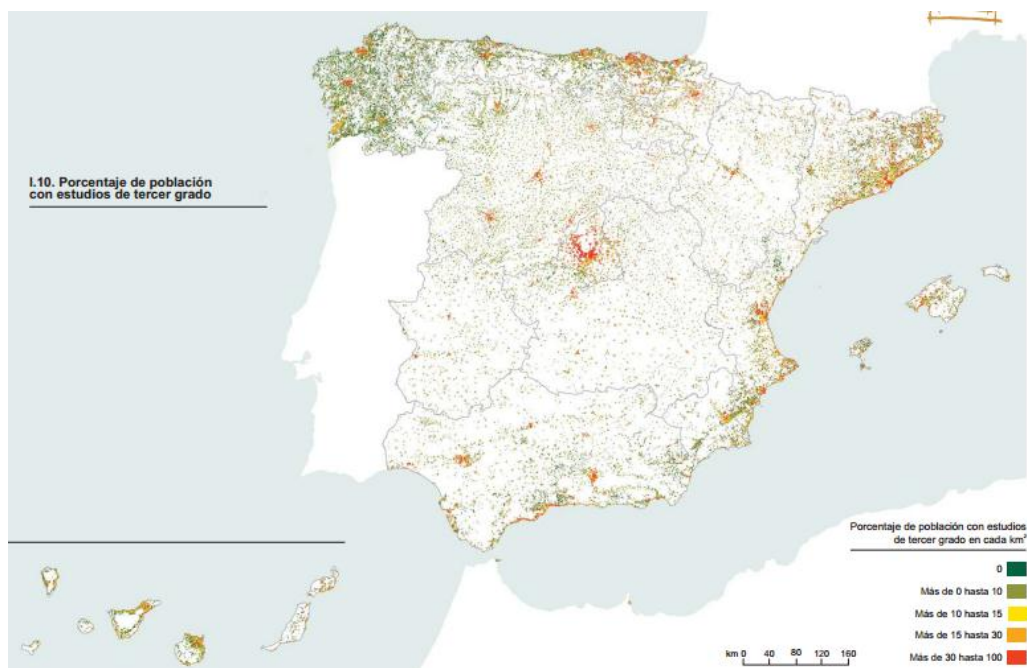
а) население (14 карт): общая численность населения, процентная доля иностранных граждан, наиболее распространенное иностранное гражданство, средний возраст, процентная доля населения в возрасте 65 лет и старше, процентная доля лиц в возрасте от 25 до 34 лет, проживающих со своими родителями, процентная доля лиц, проживающих в месте, отличном от места своего рождения, год прибытия в Испанию (в среднем), средний уровень образования, процентная доля населения с высоким уровнем образования, среднее число детей, среднее время, затрачиваемое на дорогу до места работы или места учебы, и процентная доля лиц, проживающих в течение 14 суток или более на территории другого муниципалитета или другой страны;

б) домохозяйства (3 карты): средний размер, процентная доля домохозяйств, состоящих из одного человека, процентная доля домохозяйств, состоящих из представителей двух или более поколений;

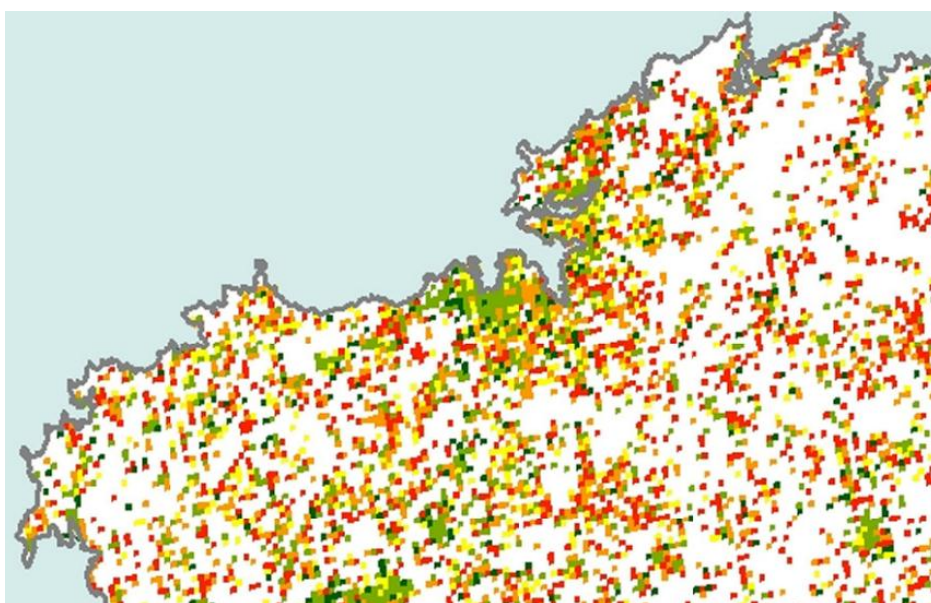
в) жилища (8 карт): процентная доля вторичных или пустующих жилищ, средняя полезная площадь, процентная доля жилищ, подключенных к сети Интернет, процентная доля жилищ с системой отопления, процентная доля арендуемых жилищ, процентная доля жилищ, находящихся в ипотеке, и средний период строительства.

10. Эти карты освещают информацию с нестандартной и инновационной точки зрения (все ячейки имеют одинаковый размер, равный 1 кв. км) и независимы от привычной пользователям карты административно деления. В большинстве случаев, и особенно в случае больших городов, карты с ячейками размером 1 кв. км содержат более подробную информацию и лучше отражают реальную ситуацию, чем муниципальные карты.

Рис. 3
Пример карты с заданным форматом (процентная доля населения с высшим образованием)



11. Одним из наиболее важных преимуществ этих карт является возможность мгновенной визуализации подробных данных. Кроме того, данная публикация рекомендована для всех категорий пользователей, а не только для тех из них, кто хорошо знаком с данными переписи. Сопоставление данных между различными регионами Испании максимально упрощено.



Ш. Ячейки размером менее 1 кв. км: алгоритм разбивки

12. Используемая для распространения географической информации НИС Испании картограмма с ячейками размером 1 кв. км представляет собой сетку, утвержденную Евростатом³. Согласно нашим данным, насчитывается 79 857 ячеек, в которых расположено по крайней мере одно жилище, и в 63 527 из этих ячеек проживают люди.

13. Сайт переписи НИС позволяет пользователям получать информацию в разбивке по ячейкам размером 100, 10 и 1 кв. км.

14. Такая картограмма может быть недостаточно подробной, особенно в случае городских районов. Следует напомнить, что данная геопространственная информация получена с помощью обследования, которое являлось одним из элементов переписи Испании. Было обследовано около 9% населения. Ячейки, содержащие больше единиц выборки (физических лиц), мы подвергли дополнительной разбивке (на ячейки размером до 50 метров с каждой стороны) в зависимости от размера выборки населения в ячейках.

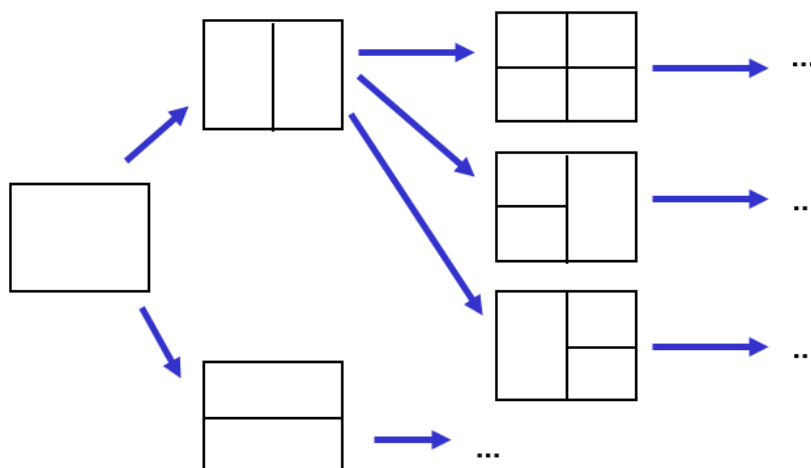
15. Пользователи могут ознакомиться с проведенной разбивкой по следующему адресу (выбрав опцию «Unlimited navigation»): <http://www.ine.es/censos2011/visor/>.

16. Алгоритм разбивки ячеек размером 1 кв. км применяется в трех различных ситуациях:

- Ситуация 1: количество единиц выборки в ячейке размером 1 кв. км не превышает 40. В этом случае какой-либо дополнительной разбивки не проводится.
- Ситуация 2: количество единиц выборки в ячейке размером 1 кв. км равно или больше 40 и меньше 250. В этом случае применяются последовательные (вертикальные и горизонтальные) половинчатые разбивки, которые оставляют по крайней мере 20 единиц выборки по каждую сторону разбивки.

Рис. 5

Механизм разбивки ячеек



³ Более подробная информация о сетке, утвержденной Евростатом, размещена по адресу http://www.efgs.info/data/eurogrid/Grid_ETRS89_LAEA_ES_1K.zip/at_download/file.

- Ситуация 3: количество единиц выборки в ячейках размером 1 кв. км равно или больше 250. В этом случае на первом этапе ячейка размером 1 кв. км разбивается на 400 подъячеек со стороной, равной 50 метрам. Затем рассчитывается переменная «р»:

$$p = 20 * \left[\left(\sqrt{\frac{\text{sampling_units}}{20}} \right) + 1 \right]$$

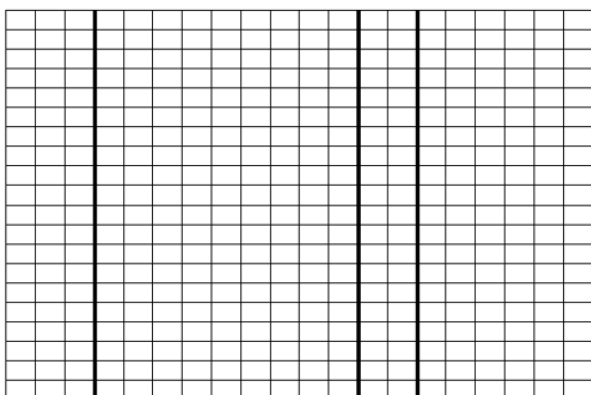
17. Наиболее густонаселенная ячейка размером 1 кв. км в Испании содержит 3 209 единиц выборки (примерно 52 845 жителей) и разбита на 273 подъячейки⁴.

18. На последнем этапе выполняются следующие три шага:

- начертить максимально возможное число вертикальных линий, оставив не менее «р» единиц выборки по каждую сторону.

Рис. 6

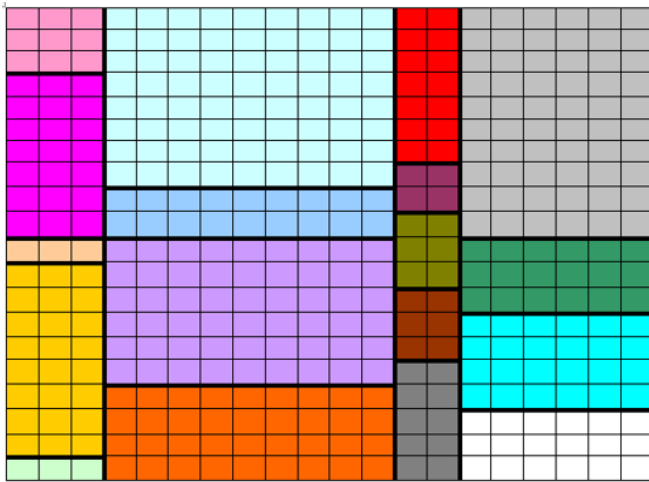
Первый шаг (n=500, p=120). 4 деления



- начертить максимально возможное число горизонтальных линий, оставив не менее 20 единиц выборки по каждую сторону.

⁴ Расположена в L'Hospitalet del Llobregat (Барселона).

Рис. 7
Второй шаг. 18 делений



с) начертить максимально возможное число вертикальных линий, оставив не менее 20 единиц выборки по каждую сторону.

Рис. 8
Третий шаг. 20 делений

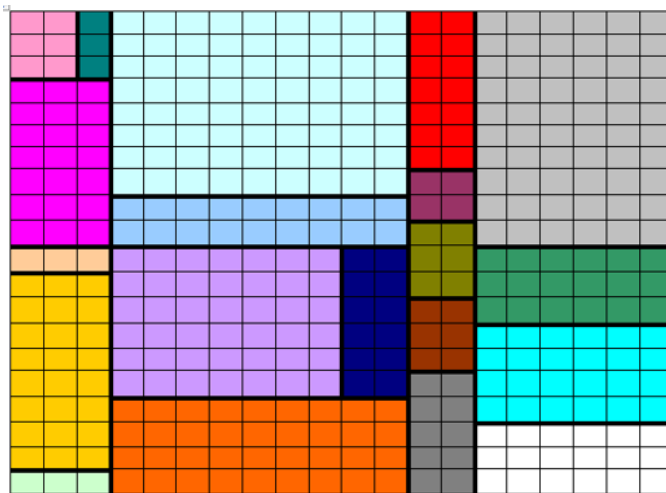


Рис. 9
Пример применения алгоритма в городе Барселона



IV. Вопросы конфиденциальности

19. Одним из наиболее важных аспектов при распространении геопрограмной информации является подход к конфиденциальности данных.

20. НИС Испании применил для обеспечения конфиденциальности географических продуктов те же подходы, что и к другим категориям продуктов. Информация распространялась в тех случаях, когда выполнялось одно из следующих условий:

а) если соотношение между числом имеющихся единиц выборки и числом категорий⁵ в запросе составляет не менее 5, то распространяется вся информация, специфицированная в запросе;

б) во всех других случаях, когда в любой единице разбивки запроса содержится не менее пяти единиц выборки, распространяется информация по тем единицам разбивки, которые удовлетворяют вышеупомянутому условию.

21. Принцип работы этого правила разъясняется на следующем примере. Предположим, что пользователь хочет получить информацию об официальном семейном положении по району, который содержит 20 единиц выборки (5 категорий):

а) в первую очередь, рассчитывается соотношение между 20 и 5. Полученный результат – 4 – не превышает 5, поэтому вся информация, специфицированная в запросе, не распространяется;

б) во-вторых, предположим, что эти 20 единиц выборки описываются следующим образом: 7 лиц, не состоящих в браке, 8 лиц, состоящих в браке, 2 вдовых и 3 разведенных лица. В этой ситуации будет распространена информация о лицах, не состоящих в браке, и о лицах, состоящих в браке, поскольку число единиц выборки по этим единицам разбивки превышает 5.

⁵ В случае некоторых чувствительных переменных применяется «множитель чувствительности», составляющий число больше единицы, и требуется выполнить условие, в соответствии с которым соотношение между числом имеющихся единиц выборки и произведением числа категорий и множителя чувствительности задействованных в запросе переменных должно быть не меньше 5.

22. Кроме того, в целях защиты конфиденциальности распространяемой информации и предотвращения идентификации на основе различий все данные округляются до ближайшего целого числа-множителя 5 (например, если результатом запроса является число 422, то число, которое будет распространено, равно 420).

23. И, наконец, что касается тематических карт с географической информацией, то они всегда разрабатываются таким образом, чтобы в таблице условных обозначений использовались числовые интервалы, с тем чтобы ни при каких условиях было невозможно разглашение персональных данных.

24. Весь комплекс мер, направленных на защиту информации, гарантирует что, несмотря на весьма подробный географический масштаб, как например 1 кв. км, разглашение персональных данных невозможно ни при каких условиях.

V. Выводы

25. Геопространственная информация, получаемая в результате переписи, является весьма разнообразной. В отличие от административного уровня, характеризующегося крайне разнородным географическим делением данных и сложностью проведения единообразных сопоставлений, использование геопространственной информации в рамках системы переписи НИС позволяет проводить сопоставление единиц, всегда имеющих одинаковую географическую площадь, или создавать практически бесконечное множество районов для анализа.

26. Кроме того, она служит весьма интересным инструментом для анализа ситуации в приграничных районах или межстранового сопоставления информации.

27. Некоторые учреждения приняли к сведению возможности такой информации⁶. Евростат стремится поощрять страны к сбору такой информации (хотя она не может быть включена в Регламент переписи 2021 года).

28. Вследствие этого весьма вероятно, что число стран, планирующих распространять географическую информацию в ходе следующей переписи, значительно возрастет.

⁶ Например, ее планируют использовать в качестве источника данных при распределении бюджета европейских фондов. Кроме того, другие учреждения, такие как EFGS <http://www.efgs.info/>, в течение многих лет занимаются содействием развитию статистических картограмм.