



Экономический и социальный совет

Распространение: Общее
13 марта 2012 г.

Оригинал: английский

Европейская экономическая комиссия

Конференция статистиков Европы

Группа экспертов по переписи населения и жилого фонда

Четырнадцатое заседание

Женева, 24-25 мая 2012 г.

Пункт 4 предварительной повестки дня

Составление списка адресов/жилых домов

Составление адресного регистра для переписи населения 2011 года – Англия и Уэльс

Записка Офиса национальной статистики, Великобритания

Резюме

Качественный полный список адресов был основополагающим условием для планирования переписи населения 2011 г. в Англии и Уэльсе. Адресный список обеспечил точную доставку, сбор и дальнейшую работу по переписным листам, а также сыграл важную роль в сборе данных на местах и оценках. В этом документе описан примененный подход и приведены основные сложности и полученные уроки.

Примечание: Данный документ описывает ситуацию в Англии и Уэльсе. Главное регистрационное бюро Шотландии (GROS) и Агентство по статистике и исследованиям Северной Ирландии (NISRA) соответственно несли ответственность за перепись населения в Шотландии и Северной Ирландии. Разные обстоятельства и требования обуславливают различия в подходах, но регулярное взаимодействие и единый подход к основным элементам помогли сблизить подходы и основные результаты.

I. Роль адресного списка в проведении переписи населения 2011 г.

1. Данный документ был подготовлен Алистером Калдером, Офис национальной статистики (ОНС).
2. Адресный регистр был основополагающим компонентом в процессе переписи населения 2011 г. в Англии и Уэльсе. В отличие от предыдущих переписей населения, в 2011 г. переписные листы были разосланы по почте по всем адресам проживания. Только в случае с институциональными учреждениями («места, обеспечивающие проживание в соответствии с некоторыми правилами») переписные листы доставлялись и собирались обратно курьером. В основном, переписные листы возвращались людьми по почте или впервые в истории Соединенного Королевства путем заполнения переписного листа через Интернет.
3. Адреса, безусловно, использовались для отправки переписных листов по почте, но они также были важны при отслеживании их возврата. Каждому переписному листу был присвоен штрих-код и при получении обратно он сканировался. Счетчики посещали каждое домохозяйство, не выславшее переписной лист обратно, и содействовали заполнению формы. Адресный регистр был нацелен на придание основного значения этому последующему процессу, а также способствовал оценке для учета невозвращенных листов.

II. Основные требования

4. Основные требования к адресному регистру заключались в следующем:
 - Полнота: Офису национальной статистики был необходим полный список жилых домов (главным образом, тех мест, где кто-нибудь мог проживать) в Англии и Уэльсе;
 - Минимальный уровень недостаточного и избыточного охвата;
 - Охват и различие между адресами жилых домов и институциональных учреждений
5. Согласно переписи населения в Англии и Уэльсе, домохозяйство - это:
 - один человек живущий отдельно; или
 - группа людей, живущих по одному и тому же адресу, которые совместно используют кухню, жилую комнату, гостиную и столовую.
6. Совместное проживание распространено, а также возрастает число «скрытых» домохозяйств, в которых несколько семей живут в отдельных помещениях по одному и тому же адресу без открытого указания на совместное проживание. Адресная проверка, обсуждаемая позднее, была направлена на выявление таких адресов.

III. Цели по обеспечению качества

7. В целях по обеспечению качества была очевидна ключевая роль адресного регистра. ОНС был необходим регистр, где недостаточный охват составлял менее 1%, т.е. он должен был включать 99% всех существующих адресов в Англии и Уэльсе на День переписи (27 марта 2011 г.).

8. Вместе с тем, регистр должен был иметь низкий уровень избыточного охвата (ошибки, неверно кодированные торговые помещения, снесенные здания); необходимо было добиться менее 2,5% избыточного охвата.

9. В частности, ОНС придерживался цели достичь менее 1% дублирования в адресах (в рамках избыточного охвата) - где имело место несколько версий одного и того же адреса, возможно, имеющие разные описания. Эта последняя задача была важна, так как дублирование в списке привело бы к напрасным почтовым расходам, потенциально бесполезной трате времени на ненужное отслеживание и риск причинения беспокойства домохозяйствам, которые уже представили необходимую форму.

10. Обследование охвата переписи (CCS), которое было направлено на проверку результатов и подтверждения оценок, состоялось двумя месяцами позже переписи населения, но адресный регистр не использовался. Обследование охвата переписи было проведено как исчерпывающее обследование домохозяйств в определенных местах и без адресного списка, и поэтому выступало в роли независимого подтверждения качества регистра.

11. Принимая во внимание важность регистра в качестве поддержки переписи населения, разработка адресного регистра была основной частью исследовательской работы, проведенной до начала переписи населения 2011 г. Несмотря на наличие ряда качественных и, в некоторой степени, конкурирующих между собой национальных адресных списков, исследования на раннем этапе показали, что ни один из них не отвечает требованиям переписи. ОНС был вынужден создать новый составной адресный регистр, объединив лучшие компоненты национальных источников. В данном разделе приводится краткий обзор нашей стратегии для выполнения этой работы.

IV. Составление списка жилых домов

12. Подход к разработке адресного регистра жилых домов состоял из пяти этапов.

A. Сопоставление

13. Основа для адресного регистра была сформирована путем сопоставления ключевых национальных наборов данных – Архива почтовых адресов с индексами Королевской почты (PAF) и Национального географического справочника земель и собственности (NLPG), которые ведутся местными органами. Процесс сопоставления также включал в себя координаты из Адресного слоя 2 (AL2), которые ведет Ordnance Survey (Национальное картографическое агентство).

14. Автоматизированное сопоставление адресов проводилось от имени ОНС компанией «Manchester Geomatics Limited» (MGL), судподрядчиком компании «Lockheed Martin». Процесс был следующим:

- Адресные списки были стандартизированы в согласованную адресную структуру. Классификации адресов были также стандартизированы, для того чтобы облегчить различие между классификациями.
- AL2 и PAF были объединены с помощью уникальных определителей адресов. На всех последующих этапах эти два продукта рассматривались как единый адресный список. В данном случае намерение заключалось в увеличении пользы от точно геокодированных и классифицированных адресов из AL2 при

помощи более обновленной информации, имеющейся в PAF Королевской почты.

- Адреса из NLPG и PAF/AL2 были затем сопоставлены по тексту с помощью серий алгоритмов, включая точные и нечеткие сопоставления.
- Оставшиеся адреса, которые не могли быть сопоставлены по тексту, были потом сопоставлены пространственно.

15. В этом процессе было сопоставлено около 94,5% адресов в масштабе страны, хотя коэффициент сопоставления различался у разных местных органов власти, у некоторых из них коэффициент сопоставления составил от 70% до 80%. В целом, уровни сопоставления были ниже в Лондоне, чем в других районах (в среднем 90%), вероятно, из-за большего числа квартир, которые не полностью совпадали в исходных продуктах.

16. Эти основные сопоставленные адреса, которые совпадали в двух национальных продуктах (и классифицировались как жилые), были изначально предположены как действительные (решение, основанное на доказательствах пробной переписи населения).

17. В ходе разработки адресного регистра данный процесс сопоставления был проведен 16 раз, каждый раз усовершенствуя сопоставление и подтверждая целесообразность подхода. И наши процессы, и сами исходные продукты постоянно улучшались вплоть до начала переписи.

18. Вместе эти сопоставления обеспечили базу адресов, около 95% регистра, которые совпадали в двух национальных продуктах. Было сделано предположение о том, что данные адреса являются действительными. Следующие этапы направлены на те адреса, которые не совпадают, и на аномальные адреса.

В. Устранение аномалий организациями, предоставляющими данные

19. Оставшиеся несовпадающие адреса были изучены в целях устранения аномалий и более 500 тыс. адресов были направлены местным органам или в Королевскую почту для проверки и решения проблемы. В данном процессе приняло участие более 90% местных органов, а также персонал Королевской почты в Англии и Уэльсе. В конечном итоге в результате этой работы, около 40% этих аномальных адресов были определены как недействительные и исключены из нашего списка.

С. Проверка на местах

20. Летом 2010 г. ОНС проверил на местах 3,6 млн. адресов. Адресная проверка проводилась с мая по август 2011 г., охватывая 15% адресов в Англии и Уэльсе. ОНС намеревался охватить те почтовые индексы (в среднем около 15 адресов), которые содержали высокое число несовпадающих адресов или адресов совместного проживания с двумя или более количеством домохозяйств (что представляет особую проблему для переписи). В проверке было задействовано 240 человек и, в основном, проверялись адреса проживания.

21. Проверяющие адресов были обеспечены распечатанными списками по всему почтовому индексу (не только аномальными адресами). Работа была разделена на 28 тыс. отдельных объемов работ, среднее количество почтовых индексов в каждом объеме составляла около 10 индексов. Они требовались для проверки этих списков, а также для выявления новых, ранее не включенных, мест проживания. Специальные

проверяющие адресов (более 85 работников было рассредоточено по всей стране в соответствии с распространением институциональных учреждений) посетили тысячи более сложных коллективных домохозяйств, включая все университеты, тюрьмы, больницы и воинские части.

22. В ходе данного процесса были выявлены некоторые недостающие адреса и около полумиллиона адресов, которые подлежали исключению из списка. Для оценки качества работы проверяющих адресов применялись меры по контролю качества. Адреса, предоставленные проверяющим, удалялись из списка адресов (но не из окончательного списка), затем подсчитывалось количество адресов, определенных как «новые адреса». Удаленные адреса являлись случайной выборкой совпадающих адресов, равной 1%. Проверяющие адресов нашли 77% этих удаленных адресов, что показало, что, несмотря на то, проверки на местах выявляли недостающие адреса, они не обеспечивали абсолютно точные списки.

23. Адресная проверка была завершена вовремя и в пределах бюджета и была признана очень успешной. Вместе с подтверждением действительности адресного списка на участках с большим количеством аномальных или сложных адресов, проверка обеспечила реальное понимание качества сопоставляемых данных и других имеющихся источников доказательств.

D. Построение доказательства

24. Информация, полученная в результате работы по устранению аномальных адресов и собственной проверки ОНС, была объединена в доказательную базу путем сбора всей имеющейся информации по каждому адресу. Использование более одного источника для создания адресного регистра привело к необходимости принятия трудных решений касательно того, что следует включать, а что необходимо исключить. Применялись бизнес-правила, но ОНС придерживался осторожного подхода, когда в случае возникновения сомнений, адреса включались в список, а не исключались, так как сведение недостаточного охвата до минимума являлось доминирующей целью. (См. обсуждения ниже.)

E. Составление окончательного списка

25. Вся работа, указанная выше, была интегрирована для формирования окончательного списка. Большинство адресов (25,1 млн.) было предоставлено для печати на переписных листах в октябре 2010 г. Второй список из порядка 300 тыс. адресов был составлен в январе 2011 г., который включал в себя последующие изменения, а также те адреса, по которым, вероятно, в период между составлением списка и днем проведения переписи будут построены и заселены жилые дома. Одним из основных недостатков переписи населения 2001 г. было использование адресного списка 3-летней давности; на этот раз ОНС обеспечил максимально возможную актуализацию списка.

V. Корректировка доказательств

26. Принцип составления адресного регистра концептуально был простым. Те адреса, которые совпали в двух базах данных, были признаны действительными. Те адреса, которые не совпали, были признаны действительными, если только не было доказательств обратного (хотя это могло привести к ненужному дублированию). Количество аномальных адресов и степень дублирования сокращались путем

постепенного совершенствования и улучшения сопоставлений адресов. ОНС улучшил качество и снизил избыточный охват путем построения доказательств и устранения аномальных адресов.

27. Так как избежание недостаточного охвата было важным для ОНС (т.е. упущение действительных адресов), те адреса, которые существовали только в одной из баз данных, были включены в список для отправки переписных листов, кроме тех случаев, когда у ОНС было явное доказательство ошибочности. Соответственно форма переписи была отправлена по всем адресам, включенным в NLPG, и по всем адресам, включенным в PAF, кроме случаев, когда имелось явное доказательство для того, чтобы предположить, что адрес не существует или является дублирующим.

28. Несмотря на то, что принцип был весьма простым, оставалась одна важная проблема: решения по аномальным адресам и доказательства, имеющиеся у ОНС, были часто противоречивыми. Разумеется, Королевская почта верит в правильность собственной версии правильности адреса, в то время как местные власти предпочитают версию NLPG, при этом обе версии адреса могут быть одинаково верными.

29. Соответственно ОНС не мог обосновать свое решение по включению или исключению какого-либо адреса на основании одного имеющегося доказательства. Вместо этого ОНС предпочел сохранить полный список адресов и на каждом этапе фиксировал представленные доказательства по каждому адресу.

30. На Рисунке 1 представлен (упрощенный) пример концепции хранения доказательства по каждому возможному адресу. Каждый код представляет позицию отдельной стороны по данному адресу.

31. Через некоторое время ОНС был создан поток доказательств – индивидуальный «DNA», относящийся к каждому адресу. Затем появилась возможность в любое время, но в особенности при подготовке ОНС окончательного списка, применять ряд бизнес-правил для принятия решения о том, какой аномальный адрес стоит проверить или куда следует выслать форму.

Рисунок 1
Концепция доказательной базы

Доказательство	IA/LA ILPG	RM PAF	ONS FIELD	VOA CT ?	3rd ?	? ?
9 High Street	1L	1F	1R	1L	1L	1V
10 High Street	0C	1F	1R	0C	1L	1V
11 High Street	0S	1F	0X	1L	1L	0X
11a High Street	0X	1F	0X	0X	0X	0X
11b High Street	0X	1F	0X	0X	1D	0X
Flat 1/11 High Street	1L	0X	1R	1L	0X	1V
Flat 2/11 High Street	1L	0X	1R	1L	0X	1V

32. Несомненно, реализация этой идеи была более сложной, чем предполагает этот принцип, и в некоторых случаях были использованы особо сложные правила. Однако концепция построения доказательной базы для последующего применения к сопоставленному набору данных обеспечила реальную гибкость, тем самым, позволяя ОНС сопоставлять различные источники доказательств относительно друг друга и проверять степень уверенности, которая требовалась для определенных целей. Этот подход также обеспечил возможность оценить относительное качество отдельных источников данных и принимать решения о том, какие формы доказательств являются наиболее надежными.

33. Были разработаны и усовершенствованы со временем бизнес-правила, сначала исходя из логики, а затем опираясь на доказательства, полученные на местах. Данный подход был признан успешным.

VI. Актуализация регистра

34. Адресный регистр сыграл гораздо менее значимую роль в 2001 г., чем в 2011 г., поэтому его качество было менее важным. Однако явным недостатком списка, использованного в 2001 г., было то, что он сильно устарел. Срез адресного регистра был сделан за три года до его использования в ходе переписи (!), поэтому он не отражал изменения и, в особенности, новые адреса, которые появились в промежуточный период (хотя доставка на руки и проверка адресного списка на местах в 2001 г. сделали проблему менее острой).

35. Одной из основных проблем является высокая степень изменений, которые происходят во всех исходных продуктах. Новые адреса постоянно добавляются в отдельные списки – в количестве 30-40 тыс. новых адресов в месяц. Если к этому добавить такое же количество удалений и изменений адресов, то важность обновления и синхронизации списков станет совершенно очевидной.

36. Ключевой принцип новой стратегии заключался в обновлении основного регистра вплоть до дня переписи. В течение всего этого времени новые данные получались, как для NLPG, так и PAF, и эти данные применялись для постоянной актуализации основного регистра. Это обеспечило ОНС механизм нахождения доказательств в отношении новых адресов, прежде всего, тех, которые включались в обновления Местного географического справочника земель и собственности (LLPG), ежемесячно предоставляемого местными властями.

37. Для того, чтобы начать печатание конвертов, основной срез адресного регистра должен был быть сделан более чем за шесть месяцев до начала переписи. Однако в течение этого периода ОНС продолжал отслеживать изменения и в феврале 2011 г. подготовил дополнительный список для печати. Вероятные будущие изменения были включены путем добавления в процессе доработки (потенциальные адреса от строительных компаний, еще не построенные и т.д.). Было принято решение включить большое количество адресов, где доказательство об их существовании было незначительным, т.е. лучше рискнуть получить в результате избыточный охват, чем упустить какие-либо новые объекты.

Окончательный регистр устарел лишь на несколько *дней*, а не *лет*.

VII. Составление адресного регистра институциональных учреждений

38. Требовался отдельный список институциональных учреждений (мест, обеспечивающих проживание в соответствии с некоторыми правилами, такие как дома престарелых, воинские части и тюрьмы), поскольку таковые учитывались при помощи переписных листов и методов, отличных от применяемых к домохозяйствам.

39. Для разработки списка институциональных учреждений были использованы разные источники. Вместо национальных адресных баз данных было использовано около 50 специальных источников для разных видов учреждений. Однако для определения содержания окончательного списка применялся такой же подход: оценка доказательств для каждого адреса, включая представленные местными органами, в целях определения адресов, которые были включены. Большинство сопоставлений было проведено вручную ввиду сложности этих адресов. И вновь ОНС придерживался осторожного подхода и включил в список адреса, где имелись сомнения в их существовании и/или в том, что кто-либо там проживал.

40. Составление списка институциональных учреждений было одним из сложных этапов проекта по подготовке адресного регистра. В некоторых случаях проблему представляло дублирование адресов в списках институциональных учреждений и жилых домов. Несмотря на приемлемость конечного результата, в ретроспективе ОНС следовало начать работу раньше и обеспечить более высокую степень согласованности в списках жилых и институциональных домохозяйств.

VIII. Оценка качества

41. Сопоставляя список жилых домохозяйств с вышеуказанными целями, можно сказать следующее:

- Цель по проценту недостаточного охвата не более 1% достигнута, по оценкам ОНС он составляет 0,8%. К тому же, отсутствовали доказательства наличия проблемы недостающих адресов в небольших районах.

- Цель по проценту избыточного охвата на уровне не более 2,5% не достигнута, по оценкам ОНС он составил 3,3%. Полученный результат был сочтен приемлемым и, в некоторой степени, является следствием сознательных решений. Цель по уровню недостаточного охвата всегда была доминирующей;
- Цель по точности позиционирования на уровне 99,0% была превышена, по оценкам ОНС, точность позиционирования составила 99,7%.
- В отношении институциональных учреждений в ходе своей работы ОНС выявил незначительное количество новых учреждений (1,9%), которые отсутствовали в первоначальном списке. Избыточный охват составил 27% вследствие двух факторов:
- осторожности при составлении списка: например, ОНС включил объекты, где была вероятность нахождения обычных резидентов (отели, кемпинги и т.д.), при этом зная, что на многих объектах никого не будет; и
- дублирования со списком домохозяйств.

42. Выше чем ожидаемый уровень дублирования, как в рамках списка домохозяйств, так и между списками домохозяйств и институциональных учреждений, привели к некоторым проблемам на других этапах процесса переписи (работа на местах, обработка данных и контактный центр) и неудобствам для некоторых членов общества. Мы сожалеем об этом, но в итоге это было ценой, которую стоило заплатить за сведение недостаточного охвата до минимального уровня.

IX. Основные выводы и извлеченные уроки

- В принципе, подход по сопоставлению адресных баз данных, проверке аномальных адресов и применению доказательной базы был верным; несмотря на возможности для улучшений, план был реализован.
- Разработка доказательной базы и применение правил обеспечили высокий уровень охвата, но осторожность в правилах привела к более высокому, чем ожидалось, уровню избыточного охвата и дублирования записей. ОНС следовало привлечь больше ресурсов для усовершенствования бизнес-правил.
- Несмотря на то, что цель по уровню избыточного охвата не была достигнута, на тот момент это было сознательное решение (включать адреса при каких-либо сомнениях). Преимущества достижения цели по минимизации недостаточного охвата далеко превосходили незначительные дополнительные издержки, обусловленные избыточным охватом, при работе на местах.
- ОНС не смог бы достичь такого высокого качества при составлении адресного регистра без поддержки и участия организаций, предоставивших адресные базы, и местных органов власти. Партнерство с провайдерами данных было, безусловно, важным и являлось одним из ключевых успехов проекта.
- Списки домохозяйств и институциональных учреждений должны разрабатываться одновременно в целях сведения к минимуму дублирования записей.

43. Хотя ОНС может отметить на будущее некоторые возможные улучшения, оглядываясь назад, можно сказать, что адресный регистр превзошел ожидания по самой важной задаче по достижению высокого уровня охвата, по оценкам, в данный регистр вошло 99,2% существующих домохозяйств. Учитывая все это, данный

сложный проект, который обеспечил фундамент для проведения переписи, имел значительный успех.
