



Европейская экономическая комиссия

Конференция европейских статистиков

Группа экспертов по переписям населения и жилищного фонда

Восемнадцатое совещание

Женева, 28 - 30 сентября 2016 года

Пункт 6 предварительной повестки дня

Инновации в методологии и технологиях переписи и результаты тестирования

Стандартизация геокодирования как основа связи между данными переписи и пространственными данными и гибкого распространения окончательных статистических данных переписи

Доклад, подготовленный Центральным статистическим управлением Польши

Резюме

До переписи раунда 2010 г. оцифровка всех границ статистического деления проводилась в рамках Статистических управлений. Согласно принятым правилам границы единиц статистического учета при необходимости согласовываются с границами кадастровых единиц. В связи с необходимостью сохранения границ жилья и населения в единицах статистического учета и учитывая разнообразные показатели по территории и плотности населения, статистические районы и переписные участки в сельской и городской местности имеют разную площадь. В городе переписным участком может быть один жилой массив или квартал города. С другой стороны, в сельской местности, где плотность населения гораздо ниже, переписной участок может охватывать целое село. Это значит, что в пределах границ кадастровых кварталов может быть несколько единиц статистического учета и наоборот, при этом статистическое деление должно поддерживаться на регулярной основе статистической службой. Для сохранения коллинеарности в случае внесения изменений кадастровой службой, соответствующие изменения должны быть сделаны и со стороны статистической службы. В связи с этим необходимо рассмотреть возможность использования границ кадастровых единиц в качестве единиц территориального деления в целях статистики. В то же время, нынешняя тенденция

состоит в отходе от традиционной переписи к переписи на основе регистров. В этом случае необходимо пересмотреть определенные ограничения населения и зданий и законность их содержания. Необходимо использовать возможность гармонизации статистического деления (статистических районов и переписных участков) с кадастровым делением (единиц кадастрового учета), учитывая потребности официальной статистики.

I. Введение

1. В настоящее время процесс получения и предоставления статистической информации в значительной мере основывается на пространственной привязке. В Польше в ходе переписи раунда 2010 г. впервые было применено сочетание данных из административных источников и регистров, содержащих пространственные данные. Применение цифровых карт и технологий GPS обеспечило революционное изменение в части возможностей планирования и управления переписью, как до ее начала, так и непосредственно во время ее проведения.
2. Для этого использовались данные, полученные из государственных геодезических и картографических ресурсов, а также ортофотоснимки (обработанные аэроснимки). Используя материалы, полученные из геодезических и статистических ресурсов, стало возможно разрабатывать основы для построения выборки для переписей, включающие статистические адреса и их пространственную привязку.
3. Цифровые карты были неотъемлемым инструментом, помогающим в работе счетчиков переписи (когда речь идет о перемещении по району, проверке основы для выборки и т.д.), глав гмин (наименьшая административная единица) и воеводств (провинций) и центральных диспетчеров, которые могли по карте проверить достигнутый прогресс в ходе переписи и, например, маршрут или расположение счетчика, используя диспетчерское или ГИС приложение, что облегчает работу главы гмины.
4. В соответствии с принятыми организационными принципами до начала переписи был проведен раунд предпереписи. В ходе него счетчик переписи проверял существование зданий и дополнял регистр недостающими адресами. Дополнительный аспект предпереписи включал изучение района проведения переписи и возможное решение каких-либо неясностей, возникших в ходе пересмотра данных на уровне гмины.
5. Мобильные терминалы были оснащены ГИС-приложением, что позволило вносить изменения и показывать, в частности, текущее местоположение счетчика (GPS) и присвоенный ему адрес. Используя мобильное приложение, счетчик мог менять местоположение адресной точки, удалять адресную точку или добавлять адресную точку, не включенную в регистр, при помощи GPS устройства. На этапе предпереписи счетчик также отвечал за контроль всей территории переписных участков, порученных ему. Это было особенно важно, когда пересмотр района гмины осуществлялся только на основе ведущихся регистров и счетчик переписи был первым и во многих случаях единственным человеком, выполняющим полевые работы в рамках переписи. ГИС-приложение также активно использовалось во время переписи для контроля за ее ходом. Оно позволило осуществлять мониторинг и контролировать работу счетчика, а также отслеживать его перемещение по территории (в том числе в целях обеспечения его безопасности).

6. Введение координат x,y и адресных точек в статистические данные позволило изменить предыдущую систему пространственной идентификации и перейти от присвоения территорий (переписных участков) к присвоению точек. Это имело принципиальную важность для применения геоинформатики в официальной статистике. Изменение режима присвоения позволило более гибко группировать данные, собранные в государственной статистике, даже по самым малым географическим территориям. Это также способствовало созданию пространственно-ориентированной базы микроданных, что позволило проводить геостатистический анализ.

7. В настоящее время статистика в Польше представляет плотность населения на основе сетки в нескольких видах на платформе визуализации статистических данных – Геостатистическом портале (geo.stat.gov.pl). Опыт, полученный при подготовке этих визуализаций, показывает, что представление данных в ячейках сетки, как правило, является очень точным, позволяет легко проводить сравнения, поскольку ячейки имеют одинаковый размер и являются стабильными в течение долгого времени. Кроме того, сетки легко интегрируются с другими научными данными (например, метеорологической информацией) и системы сеток могут строиться иерархически относительно размера ячейки и, таким образом, совпадая с исследуемой территорией. Ячейки сетки также могут быть собраны для формирования территорий с учетом конкретных целей и охватывая исследуемую область (горный район, водосбор).

8. Портал используется для пространственного представления крупнейшего информационного ресурса Польши, он позволяет публиковать агрегированные статистические данные в форме различных видов пространственного анализа, представляемых на картах со статистической достоверностью. Геостатистический портал является инструментом для интерактивного картографического представления и размещения данных, полученных по результатам переписи. Он выполняет следующие функции:

- хранение,
- представление,
- обмен информацией с широким кругом получателей.







9. Портал функционирует на двух уровнях: для внутренних (официальной статистики) и внешних пользователей, и объем представляемых данных определяется посредством соответствующих ролей и авторизаций. Внутренние пользователи имеют доступ, как к данным единиц, так и к агрегированным данным, в то время как внешние пользователи имеют доступ только к агрегированным данным, публикуемым с учетом статистической достоверности.

10. Интерфейс Геостатистического портала позволяет пользователям легко и быстро получить доступ к необходимой статистической информации. Данные представляются с применением таких методов представления картографических данных как картограммы (фоновые карты, хороплет) и различные картодиаграммы. Также можно установить собственные параметры для визуализации тематической области для заданной картограммы. К ним относятся размер, уровень агрегации (единица административно-территориального деления), количество интервалов и т.д. Помимо возможности использования готового пространственного анализа в Геостатистическом портале внутренние пользователи могут составлять специализированные тематические карты, исходя из выбранной характеристики модели данных, используя пространственный анализ, т.е. линейный или дистанционный анализ, или буферизацию объектов.

11. Классификация видов анализа, проводимого по точкам с координатами x,y также дает возможность стать независимым от изменений границ (в территориальном делении страны), которые обычно приводят к изменениям переписных участков и трудоемким перерасчетам. Это способствует проведению сравнительного анализа временных рядов вне зависимости от изменений, происходящих в административно-территориальном делении. Дополнительным плюсом является возможность агрегирования данных как в структуре NUTS (Номенклатура территориальных единиц для целей статистики) административного деления, так и в ячейках СЕТКИ, разработанной Центральным статистическим управлением в рамках проектов ГЕОСТАТ.

12. До раунда переписи 2010/2011 гг. оцифровка всех границ статистического деления производилась в рамках статистических управлений. Согласно принятым правилам границы единиц статистического учета при необходимости согласовываются с границами кадастровых единиц. В связи с необходимостью сохранения границ жилья и населения в единицах статистического учета и учитывая разнообразные показатели по территории и плотности населения, статистические районы и переписные участки в сельской и городской местности имеют разную площадь. В городе переписным участком может быть один жилой массив или квартал города. С другой стороны, в сельской местности, где плотность населения гораздо ниже, переписной участок может охватывать целое село. Это значит, что в пределах границ кадастровых кварталов может быть несколько единиц статистического учета и наоборот, и что статистическое деление должно поддерживаться на регулярной основе статистической службой. Для сохранения коллинеарности в случае внесения изменений кадастровой службой, соответствующие изменения должны быть сделаны и со стороны статистической службы. В связи с этим необходимо рассмотреть возможность использования границ кадастровых единиц в качестве единиц территориального деления в целях статистики. В то же время, нынешняя тенденция состоит в отходе от традиционной переписи к переписи на основе регистров. В этом случае необходимо пересмотреть определенные ограничения населения и зданий и законность их содержания. Необходимо использовать возможность гармонизации статистического деления (статистических районов и переписных участков) с кадастровым делением (единиц кадастрового учета), учитывая потребности официальной статистики.

13. Со ссылкой на описанный выше опыт в Польше ниже представлено предложение по инвентаризации уровня интеграции пространственных объектов, используемых в статистике и геодезии – «Модель десяти уровней» - в целях гармонизации рамочных структур в статистике и геодезии.

Geodetic System	Layers (suitable for geocoding)	Statistical System
+	NUTS1 - Administrative level 1	+
+	NUTS2 - Administrative level 2	+
+	NUTS3 - Administrative level 3	+
+	LAU1 - Administrative level 4	+
+	LAU2 - Administrative level 5	+
 Cadastral units Cadastral parcels	INDIVIDUAL UNITS	 Statistical regions Enumeration areas
+	 POLYGON	?
?	 GRID	+
+	 LINE	?
+	 POINT	+

II. Административные уровни (уровень 1-5)

14. В целом, в обеих системах (в геодезической и статистической) функционирует пять слоев, относящихся к административно-территориальному делению страны (в Европе NUTS 1, 2, 3 и LAU 1, 2), и с точки зрения синхронизации данных эти слои рассматриваются одинаково в обеих системах. Это означает, что данные, собираемые в геодезии и силами статистической службы, привязываются к той же топологии, которая уже создана Картографическим управлением. Это обеспечивает возможность применения этой топологии к процессу геокодирования статистических данных. Однако этот процесс происходит на высоком уровне агрегирования, который чаще всего не является удовлетворительным для пользователей статистики.

III. Индивидуальные единицы для внутренних целей (уровень б)

15. В геодезии - кадастровые единицы и кадастровые участки, а в статистике – статистические районы и переписные участки. Топологии в этих делениях имеют свое предназначение в этих двух системах - они независимы. Суть этого предложения состоит в том, чтобы провести некоторую работу для того, чтобы попытаться гармонизировать их и создать механизм для объединения пространственных данных со статистическими данными. Гармонизация вызывает некоторые проблемы, поскольку в статистике используются статистические единицы и, к сожалению, в геодезии не принято такое деление пространства и предпочтение отдается кадастровой системе. Эта область требует дальнейшего обсуждения. Одним из решений может быть перекодировка первично кодируемых явлений (например, статистических единиц или кадастра) с применением согласованного и гармонизированного уровня предложенной модели, но это непростая задача. Основная проблема возникает в случае явлений, которые относятся к диапазонам, отличным от приведенных выше.

IV. Многоугольник (уровень 7)

16. В геодезии широко применяется многоугольный уровень. В случае с экологическими явлениями достаточно сложно определить их многоугольные диапазоны в связи со сложностью определения местонахождения этих явлений в пространстве. Для получения данных, собираемых с необходимым уровнем точности, необходимо разработать новый многоугольный слой, специально для целей статистики. Но такой слабо стандартизированный слой был бы сильно подвержен изменчивости и характеризовался бы разнообразием обследованных многоугольников. Как следствие, для целей статистики он станет запутанным с течением времени и соответственно бесполезным для статистического анализа и сравнения.

V. СЕТКА (уровень 8)

17. Неким компромиссом, который обеспечивает хорошее решение в этой области, является идея использовать сетку как особый вид многоугольника. Картографирование полученных результатов на сетке квадратов, как в исследованиях, проводимых в статистике, так и наблюдениях, проводимых в геодезии, увеличивает шансы достижения большей согласованности между этими двумя системами, в частности, в отношении экологических явлений, требующих пространственной локализации только в форме многоугольника. Такая стандартизация многоугольника обеспечивает сетку с надлежащим образом выбранным размером ячеек. Проблема заключается в том, что объекты СЕТКИ, как правило, должны вноситься в существующую геодезическую систему, которая не столь популярна в картографическом ведомстве, и может оказаться непростой задачей. Но этот шаг гарантирует надлежащее геокодирование экологических явлений, представляемых в статистике. В настоящее время в статистике используется сетка с размером квадрата 1 км² в основном для представления и публикации данных по численности населения. Учитывая, что в геодезии проводятся исследование явлений с пространственными пределами, которые также сложно определить, использование хорошо подобранного размера квадратов в сетке, по сути, также решает эту проблему.

Немаловажным является то, что сетка с размером квадратов 1 км² является достаточно гибкой, что позволяет виртуальное соединение всех данных на любом уровне пространственного деления.

VI. ЛИНИЯ (уровень 9)

18. Геодезические данные также представляются с использованием линейных объектов. В настоящее время в статистике нет обследований, которые могли бы быть представлены объектами такого рода. Это тенденция, которая может получить признание в будущем при распространении статистических данных, относящихся, например, к транспорту, водным путям или линейным инвестициям. Возможность создания линейной статистики появится в ближайшем будущем и это позволит легко связывать линейные статистические данные с топологией, предлагаемой в геодезии (линейное геокодирование).

VII. ТОЧКА (уровень 10)

19. На самом нижнем уровне геокодирования (самый высокий уровень точности) в обеих системах выступают точки, отражающие пространственное положение, обычно в виде координат x, y . В настоящее время в этой области наблюдается активный прогресс в развитии сотрудничества между статистической и геодезической службами. Причина заключается в том, что в ходе последней переписи в большинстве стран успешно применялась топология адресных точек и с того времени она стала важным связующим звеном между статистическими и пространственными данными (точное точечное геокодирование).

VIII. Ожидаемые выгоды

20. Далее приведены ожидаемые плюсы от Модели десяти уровней, представленной выше:

- Инвентаризация уровней интеграции пространственных объектов, используемых в статистике и геодезии, в целях гармонизации;
- Поддержка статистической геопространственной структуры;
- Повышение добавленной стоимости статистических и пространственных данных;
- Выгоды от гармонизации обеих систем в результате методологической работы по данному вопросу;
- Геокодированные данные единичных записей в условиях управления данными;
- Использование основополагающей геопространственной инфраструктуры и геокодирования;
- Повышение качества геопространственных данных, интеграция статистических и пространственных данных;
- Повышение качества пространственного анализа;
- Повышение совместимости наборов данных, возможность анализа;
- Установление прочного сотрудничества между картографическим ведомством и статистической службой с поддержанием актуальности обеих систем;
- Возможность подчеркнуть необходимость и начать сотрудничество между картографическим ведомством и органами статистики.

21. В заключение, вопросительные знаки в предложенной модели (отсутствие сетки в геодезии и отсутствие линейных объектов в статистике, разработка механизма перехода со статического деления на геодезическое деление) должны стать предметом интенсивной работы для того, чтобы сломить существующие барьеры и, в качестве отправной точки, - достичь практического прогресса в методологии объединения пространственных данных со статистическими данными.