



---

**Европейская экономическая комиссия****Конференция европейских статистиков****Пятьдесят восьмая пленарная сессия**

Париж, 8–10 июня 2010 года

Пункт 6 предварительной повестки дня

**Пространственная статистика****Создание национальной иерархической сетки в  
Словении – извлеченные уроки и задачи на будущее****Записка Статистического управления Республики Словения***Резюме*

Для осуществления и планирования различной человеческой деятельности в окружающей человека среде или для мониторинга трендовых изменений различных явлений в пространстве и времени требуются широкий круг пространственных статистических данных и адекватные ответные действия их поставщиков. Ряды данных за длительные периоды и базы регистровых данных, ведущиеся Статистическим управлением Республики Словения или иными органами, были признаны в качестве ценного вклада в решение этих задач, позволяющего использовать современные инструменты пространственного (статистического) анализа, т.е. географические информационные системы.

Распространение пространственных статистических данных в формате географических информационных систем является всего лишь еще одной формой распространения информации, дополняющей уже существующие (например, цифровые или печатные таблицы), а поскольку данные можно интегрировать в географическую информационную систему, сетки предоставляют широкие возможности как пользователям, так и статистическим органам. В этой связи Статистическое управление Республики Словения решило изучить возможность использования своих собственных сеточных данных, что привело в итоге к созданию национальной иерархической системы сеток и участию в различных международных геостатистических инициативах.

## I. Введение

1. Усиление антропогенной нагрузки на окружающую среду требует стратегического планирования человеческой деятельности и соблюдения принципов устойчивого развития. Географические информационные системы (ГИС) открывают новые горизонты, а заодно и расширяют возможности для понимания и распространения пространственных статистических данных. Для удовлетворения потребностей растущего круга пользователей пространственных данных статистическим органам необходимо принимать на вооружение новые инструменты и форматы распространения пространственной информации. Широкое применение в этой области находят среди прочего и статистические сетки. Статистическое управление Республики Словения (СУРС) уже изучило ряд источников данных для статистических сеток, и о сделанных им основных выводах и поставленных задачах речь пойдет ниже.

2. Для создания статистических сеток могут использоваться как пространственные (замкнутые геометрические фигуры – обследуемые районы), так точечные данные. Регистровая статистика в Словении представляет собой надежный фундамент для создания статистических сеток с высоким разрешением. Регистр пространственных единиц, который был создан по инициативе СУРС и который в настоящее время ведется Управлением картографии Республики Словения, стал первым шагом в разработке рационального территориального деления, позволившего впервые провести географическую привязку (точечную) статистических данных в Словении (перепись населения и жилищного фонда 1971 года). На основе данных переписи 1971 года был создан Центральный регистр населения (ЦРН), а жителям Словении были впервые присвоены идентификационные номера<sup>1</sup>, что в дальнейшем упростило использование данных отдельных регистров. Хотя в то время существовала возможность лишь сводить эти данные в таблицы, а не представлять их графически, как это делается сегодня благодаря ГИС, было принято решение на постоянной основе хранить информацию о территориальной привязке максимально возможного (или приемлемого) качества.

3. Прозорливость этого решения стала очевидной, когда в 1995 году было завершено составление графической части Регистра пространственных единиц. Хотя данные в таблицах и имели географическую привязку, до этого момента было очень сложно и даже невозможно анализировать их при помощи ГИС по всей национальной территории. На практике это означало, что с 1995 года данные о населении, полученные, например, в ходе переписи 1971 года, могли графически отражаться на карте по каждому конкретному лицу с точностью до координат его дома, постоянного места проживания или района регистрации. Когда СУРС стало работать со статистическими сетками пространственных данных, наиболее пригодными считались точечные данные из различных регистров, однако в последнее время опробуются новые методы повышения позиционной точности многоугольников, которые предполагают работу с их точечными координатами и их агрегирование в рамках сеток, о чем говорится ниже. Таким образом, статистические данные переписей 1971, 1981, 1991 и 2002 годов, а также информация ЦРН позволяют в историческом ракурсе взглянуть на то, как различные пространственные показатели изменились за последние 40 лет.

---

<sup>1</sup> Oblak Flander, A.: Opportunities and Challenges of a Register-Based Census of Population and Housing – the Case in Slovenia. Seminar on Registers in Statistics – methodology and quality, Helsinki, 2007.

4. Хотя подготовка баз данных, пригодных для использования со статистическими сетками, и стала важным этапом, необходимо было также обменяться опытом работы с сеточными статистическими данными или имеющими географическую привязку данными в целом. Реализация проекта совместно со Статистическим управлением Австрии и участие в Европейском форуме по геостатистике помогли заметно углубить понимание процесса создания, анализа и распространения сеточной статистики. Гармонизация этих процедур невозможна без международного сотрудничества, и ожидается, что итогом проекта "ГЕОСТАТ" (и последующих аналогичных проектов), осуществляемого в рамках Сети информационных центров Европейской статистической системы (ESSnet), станут широкоприемлемые и применимые решения.

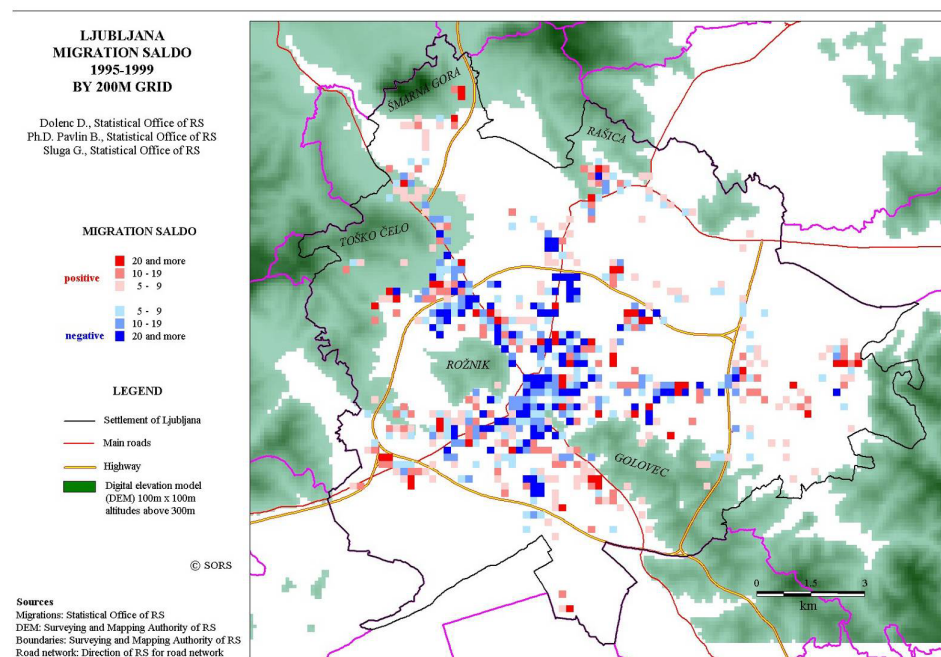
5. Изучая новые горизонты и методы анализа и использования сеточных данных, СУРС приступила в 2008 году к созданию национальной иерархической системы сеток. Наличие обновленной информации и образцовое сотрудничество между различными официальными учреждениями страны позволяют надеяться на то, что эта сетка получит широкое применение и сможет легко интегрироваться в различные инфраструктуры пространственной информации. После преодоления первых трудностей началась работа над определением системы данных для статистической сетки и механизмом контроля за раскрытием информации, которая ведется и по сей день. В то же время, эти усилия уже принесли некоторые важные результаты, которые также упоминаются в настоящем документе.

## II. Приобретение опыта

6. СУРС работает с пространственными данными на статистических сетках с начала 1990-х годов, в конце которых оно представило первые результаты анализа пространственной статистики (рис. 1).

Рисунок 1

Сальдо миграции в Любляне в 1995–1999 годах (на 200-метровой сетке)



7. С тех пор спрос пользователей на статистические данные в формате ГИС постоянно рос, что убедило СУРС в необходимости расширять работу с сеточными статистическими данными и их распространение. Выводы, сделанные по результатам отдельных практических исследований и распространения их результатов, послужили стимулом к созданию национальной иерархической системы сеток. С самого начала такая система сеток рассматривалась в качестве составной части Регистра пространственных единиц, ведущегося Управлением картографии Республики Словения.

8. Под сетками традиционно понимаются пространственные единицы, которые:

- a) являются независимыми от политических или административных единиц;
- b) не меняются во времени;
- c) распределены равномерно (сопоставимы);
- d) полезны для анализа на микро- и макроуровнях;
- e) могут обрабатываться с помощью стандартных инструментов ГИС;
- f) могут легко составляться на основе точечных данных;
- g) позволяют осуществлять более эффективный контроль за раскрытием статистической информации;
- h) позволяют получать более точные оценки по небольшим районам;
- i) позволяют проводить более качественный выборочный анализ.

9. Вышеуказанные характеристики частично применялись и проверялись на практике в ходе совместного проекта, реализованного Статистическим управлением Австрии и СУРС в 2005 году при поддержке Программы Phare Европейского союза. К тому времени Статистическое управление Австрии уже наладило систематическое распространение пространственной статистики и разработало протокол защиты сеточных данных. Поскольку речь шла о соседней стране, было очень удобно представлять различные социально-экономические данные по обеим странам на одной карте. Трансграничный анализ был проведен на расширенной австрийской сетке в системе координат в универсальной поперечной проекции Меркатора, на которую были нанесены агрегированные данные, имеющие географическую привязку. Поскольку использовалась лишь третья зона этой проекции, западная часть Австрии оказалась исключенной из анализа. Обмен знаниями и опытом продолжался при активном участии Европейского ГридКлуба и Северного форума по геостатистике, который впоследствии был преобразован в Европейский форум по геостатистике. Являясь членом Европейского форума по геостатистике, СУРС участвует также в проекте "ГЕОСТАТ", осуществляемом в рамках Сети информационных центров Европейской статистической системы. Группа экспертов в составе европейских геостатистиков стала отличной платформой для обсуждения актуальных вопросов сеточной статистики и геостатистики в целом.

10. Учитывая состояние инфраструктуры пространственной статистики в Словении и опираясь на опыт работы с сеточными данными статистических управлений ряда европейских стран, СУРС воспользовалось давними традициями регистровой статистики для нанесения пространственных статистических данных на сетки с высокой степенью позиционной точности.

### III. Национальная иерархическая система сеток

11. В 2008 году СУРС выступило с инициативой создания национальной иерархической системы сеток в Словении. Принять участие в ее осуществлении согласились три учреждения: СУРС должно было обеспечивать методологическую поддержку, а Геодезический институт Словении и Управление картографии Республики Словения – техническую поддержку.

12. Этот совместный проект преследовал следующие цели:

- a) создать векторные квадратные сетки с семью основными размерами ячеек;
- b) определить систему обозначений квадратов сеток в соответствии с их иерархической структурой;
- c) определить точку отсчета иерархической системы сеток;
- d) определить квадраты сеток как в прежней (D48/GK), так и в нынешней (D96/TM) национальной системе координат.

13. К семи основным сеткам относятся 100, 200, 500, 1 000, 2 500, 5 000 и 10 000-метровые сетки. Самая мелкая сетка с размером ячеек 100 x 100 метров была выбрана с учетом потребностей пользователей в пространственных данных с высокой степенью разрешения и необходимости обеспечения согласованности с другими существующими в Словении базами пространственных данных.

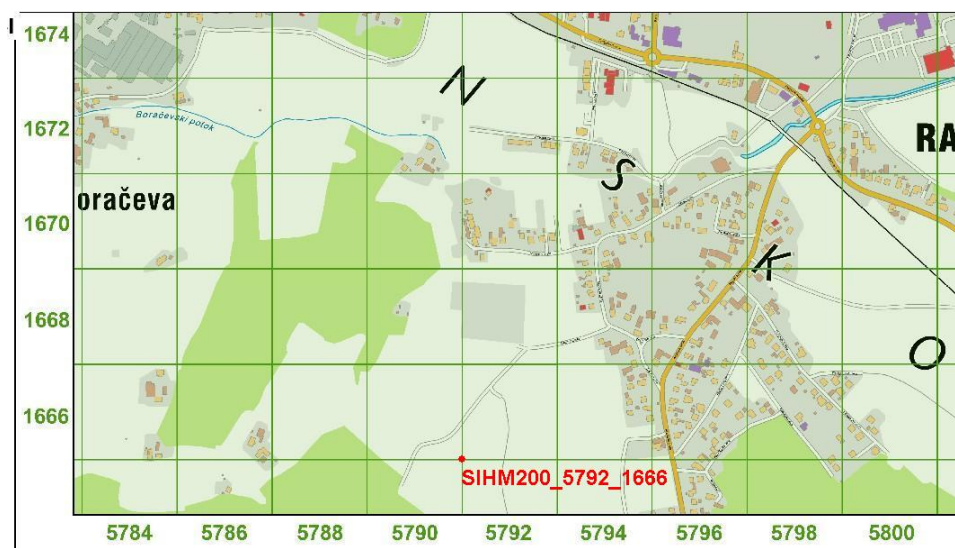
14. Для решения проблемы преобразования данных из одной системы координат в другую было принято решение создать векторные квадратные сетки в системе D96/TM, а затем преобразовать их в формат D48/GK, в котором квадраты сеток в обеих системах координат имеют один и тот же идентификатор. Преобразованные в прежнюю систему координат D48/GK ячейки сеток слегка теряют свою квадратную структуру, покрывая, тем не менее, прежнюю территорию. Таким образом, все официальные пространственные статистические данные или пространственные данные самих пользователей, в основном существующие в прежней системе координат D48/GK, попросту наносятся в агрегированном виде на сетку в формате D48/GK, а затем преобразуются в систему D96/TM благодаря идентификатору ячейки. Точка отсчета иерархической системы сеток находится на пересечении меридиана, идущего вдоль 15° восточной долготы и параллели, расположенной в 5 000 000 метрах к северу от экватора. 500 000 метров были добавлены к 15-градусному меридиану для того, чтобы избежать появления отрицательных значений к западу от этого меридиана. Координата параллели составляет 0 метров. Ось Y идет параллельно экватору в восточном направлении, а ось X – параллельно центральному меридиану (15°) в северном. Таким образом, точка с координатами (500 000, 0) является точкой отсчета системы сеток, совпадающей с точкой отсчета нынешней системы координат D96/TM, что упрощает увязку между собой разных пространственных данных на территории Словении.

15. Обозначение квадрата сетки (например, SIHM200\_5792\_1666) состоит из буквенной части SIHM (Slovene hierarchical grid), размера квадрата в метрах и координат по оси Y (\_Y) и оси X (\_X), укороченных на количество лишних нулей. Таким образом, идентификатор квадрата задает как его размер, так и координаты.

Таблица 1  
Примеры сокращенных обозначений квадратов сеток

<i>Полное обозначение квадратов</i>	<i>Укороченное обозначение квадратов</i>
SIHM100_463000_103700	SIHM100_4630_1037
SIHM200_463000_103600	SIHM200_4630_1036
SIHM500_463000_103500	SIHM500_4630_1035
SIHM1000_463000_103000	SIHM1000_463_103
SIHM2500_462500_102500	SIHM2500_4625_1025
SIHM5000_460000_100000	SIHM5000_460_100
SIHM10000_460000_100000	SIHM10000_46_10

Рисунок 2  
Примеры обозначений квадратов 200-метровой сетки



LEGEND:

- Cell SIHM200

#### IV. Пространственные статистические данные в сеточном формате (метод агрегированного отображения)

16. Пространственные статистические данные, имеющие географическую привязку (точечные координаты) или географический код (территориальную привязку, например к административным районам), могут в агрегированном виде наноситься на сетки. Источниками сеточных данных являются различные регистры и результаты четырех переписей населения (1971, 1981, 1991 и 2002 годов). Данные регистров и переписи 2002 года имеют географическую привязку, а данные остальных переписей являются геокодированными.

17. Данные, привязанные к территориальным единицам, как правило, считаются более пригодными для использования в рамках процедуры дезагрегированного отображения, в рамках которой информация, например, о населении муниципалитетов наносится на сетки в дезагрегированном виде при помощи некоторых вспомогательных данных, например данных о землепользовании и землеустройстве. Этот метод может применяться в тех случаях, когда отсутствуют более точные пространственные статистические данные. Тем не менее в силу вполне приемлемой позиционной точности имеющиеся геокодированные данные словенских переписей позволяют использовать их и в рамках "восходящего" метода.

18. Геокодированные данные переписей 1971, 1981 и 1991 годов привязаны к территориальным округам. Такие округа являются самыми мелкими административными единицами в Словении, и, поскольку они охватывают определенные районы, их агрегированное отображение на сетке отличается от нанесения на сетку данных, имеющих точную географическую привязку. Процедура преобразования геокодированных данных в данные с географической привязкой на практике требует определения средней точки (центроида) в форме координат конкретного объекта в конкретном территориальном округе. Центроиды территориальных округов в Словении уже удовлетворяют этому требованию, поскольку они в большинстве случаев совпадают с местом наибольшей концентрации населения на территории конкретного округа. Таким образом, центроиды совпадают с координатами крупных объектов, например школ. Для территориальных округов, не имеющих таких крупных объектов, центроиды определяются по важным естественным признакам, например:

- a) по центру тяжести плотно застроенного района на территории округа;
- b) по центру тяжести всех зданий на территории округа, где они расположены;
- c) по центру тяжести территориального округа, если на его территории нет зданий.

19. Соответственно, любое изменение территориального округа неизбежно повлечет за собой изменение его центроида. Несмотря на это, центроиды территориальных округов при необходимости дополнительно проверялись и корректировались, поскольку распределение населения по территории некоторых областей в последние десятилетия сильно изменилось. Эта корректировка производилась по текущему состоянию центроидов зданий, причем на основании года сооружения для этой цели отбирались лишь те здания, которые существовали и были заселены в период проведения переписи. В других случаях, например, когда имеются данные о числе рабочих мест, центроид территориального округа может быть скорректирован с учетом центра притяжения всех зданий, в которых осуществляется хозяйственная деятельность. Помимо этого, позиционная точность данных переписей населения 1981 и 1991 годов, отображенных на сетках в агрегированном виде, может быть проверена по имеющим географическую привязку данным Центрального регистра населения, уже имеющимся за соответствующие периоды.

20. В отличие от данных, имеющих географическую привязку, геокодированные данные задают размер квадрата сетки по средней площади. В густонаселенных районах территориальные округа имеют меньшую площадь, чем в малонаселенных. Итоги ряда пространственных исследований говорят о том, что данные переписи могут отображаться в агрегированном виде на сетках с разме-

рами квадрата 100 x 100 м или 200 x 200 м в случае густонаселенных районов, 500 x 500 м в районах со средней плотностью населения и 1 x 1 км в малонаселенных районах. В таблице 1 площадь территориальных округов соотносится с площадью, охватываемой на 100, 200, 500 и 1 000-метровыми сетками, а также с численностью населения. Около 47% населения можно напрямую отображать 100, 200 и 500-метровых сетках, тем самым добиваясь высокого разрешения пространственных данных в густонаселенных районах.

Таблица 2

**Сопоставление территориальных округов и сеток**

<i>Площадь территориальных округов в км<sup>2</sup></i>	<i>% всех территориальных округов</i>	<i>% населения</i>	<i>% национальной территории</i>	<i>Плотность населения на 1 км<sup>2</sup></i>
площадь ≤ 0,01	11,80	10,85	0,05	20 078
0,01 < площадь ≤ 0,04	14,28	15,70	0,27	5 770
0,04 < площадь ≤ 0,25	19,40	20,85	1,90	1 095
0,25 < площадь ≤ 1	22,15	20,14	11,18	179
площадь > 1	32,36	32,46	86,60	37

21. Поскольку квадрат сетки определенного размера может не покрывать большей части территориального округа, т.е. территориальный округ может захватывать несколько квадратов, проводилась еще одна проверка с тем, чтобы центроид территориального округа оказывался в том квадрате сетки, в котором находилась основная часть (не менее 75%) населения округа. Когда это оказывалось невозможным, использовался более крупный квадрат сетки.

22. Применяемая методология настоятельно рекомендует архивировать данные переписи (или иные сведения) с максимально точной позиционной привязкой, если это возможно с юридической и технической точек зрения. Позиционную точность можно повысить путем определения точечных координат всех важных пространственных объектов (например, зданий), к которым относится основная часть данных, причем сделать это придется всего один раз. Большим преимуществом преобразованных подобным образом геокодированных данных является то, что их можно отобразить в агрегированном виде на произвольно выбранной сетке, независимо от ее картографической проекции или системы координат, конечно же, при условии соблюдения размерности квадратов. Такие данные приобретают и все остальные преимущества сеточных данных, о которых говорилось выше.

23. Местоположение имеющих географическую привязку данных в Словении определяется координатами центроидов зданий или частей зданий, которым присвоены номера домов. Координаты этих центроидов хранятся в Регистре пространственных единиц, и благодаря общим идентификаторам центроиды можно увязать с отдельными данными из разных регистров. К числу регистров, которые могут использоваться для целей агрегированного отображения данных на сетках, относятся:

- a) Центральный регистр населения, данные которого с 1981 года имеют достаточную степень позиционной точности;
- b) Коммерческий регистр Словении;
- c) Налоговый регистр;
- d) Статистический регистр занятости.



Регистры зданий и недвижимости планируется создать в ближайшем будущем.

24. Имеющие географическую привязку данные перечисленных регистров могут не только наноситься на сетки с произвольными картографическими проекциями или системами координат, но и на сетки с произвольным размером квадратов. В то же время, несмотря на многие преимущества регистровой статистики, проведение переписей на основе данных регистров неизбежно ведет к утрате некоторых важных сведений, например о ежедневной миграции, истинном местонахождении рабочих мест и т.д.

## V. Политика защиты сеточных данных

25. Географическое местоположение респондента является тем атрибутом, который позволяет раскрыть как его самого, так и информацию о нем. Поэтому распространение сеточных данных и пространственной статистики в целом требует выбора приемлемой пространственной точности данных с точки зрения процедур их защиты. В отсутствие международных руководящих принципов защиты информации на статистических сетках страны сами определяют свою политику защиты данных, что может стать препятствием для осуществления трансграничных проектов. Приступая к созданию национальной иерархической системы сеток, СУРС решило определить стандартный круг данных, которые могут отображаться на сетках, а также правила их защиты. Пространственные статистические данные наносились на сетки раньше, но лишь по заказу пользователей, а не в качестве стандартного предложения СУРС, причем каждый такой заказ рассматривался комитетом по защите данных.

26. Первыми кандидатами для отображения на сетках стали данные демографической статистики, поскольку за подобными данными пользователи обращаются чаще всего. Для этой цели был выбран набор демографических данных, уже распространяемый на уровне поселений (начиная с муниципалитетов, т.е. с уровня LAU 2), который содержит следующие сведения:

- a) численность населения;
- b) пол;
- c) пятилетние возрастные группы.

27. Поскольку доступность этих данных не ограничивается по соображениям конфиденциальности, было принято решение временно использовать эту практику и применительно к сеткам, а поскольку самое маленькое поселение в Словении укладывается в 100-метровый квадрат, это решение не встретило возражений. В 2011 году ситуация может измениться, поскольку комитет по защите данных готовит новый протокол защиты данных переписи 2011 года, который будет распространяться даже на ту пространственную информацию, которая не связана с переписями. Во-вторых, увеличение числа атрибутов сеточных данных, особенно в таких чувствительных областях, как образование или доходы, также требует исключения менее чувствительных данных, например за счет сокращения численности рассматриваемого населения. Отказ от исключения таких данных повлечет за собой повышение уровня защиты атрибутов, однако важно помнить о том, что потребителей пространственных данных обычно больше интересуют те области, где то или иное конкретное явление происходит чаще. Поэтому минимальный уровень защиты устанавливается и для численных значений.

28. Существует несколько методов решения проблемы, создаваемой исключением квадратов сетки, и с учетом пространственного распределения различных явлений в Словении было решено отдать приоритет раскрытию большего объема данных с высоким разрешением, например по густонаселенным районам, вместо того, чтобы раскрывать больше общих данных или снижать их разрешение в результате объединения квадратов сетки. Менее крупные квадраты объединяются лишь в рамках существующей иерархической системы сеток и лишь в тех случаях, когда это ведет к неоправданному снижению степени разрешения пространственных данных. Например, три исключенных 100-метровых квадрата сетки могут быть объединены с четвертым таким квадратом в 200-метровый квадрат лишь в том случае, если суммарный численный показатель (например, численность населения) по четырем объединенным квадратам превышает пороговый уровень, позволяющий раскрывать тот или иной атрибут (например, образование), и, по крайней мере, равняется ему.

29. Ожидается, что исходное фиксирование набора данных окажется для пользователей очень полезным, поскольку они будут заранее знать, какую информацию можно получить на сетках и при каких условиях. Кроме того, снизится нагрузка и на СУРС, поскольку данные будут готовиться заранее с учетом уже известных правил защиты данных.

## VI. Выводы

30. Как и многие другие европейские национальные статистические управления, СУРС считает сетки важным дополнением к существующим формам распространения или представления статистических данных. Сетки рассматриваются всего лишь как еще один формат, альтернативный традиционным бумажным документам или цифровым таблицам, который имеет в то же время свои собственные характеристики и сферу применения. Сетки обеспечивают такую пространственную точность данных, которая позволяет использовать их для различных видов пространственного анализа или отображения, делает их удобными в использовании и гарантирует статистическую конфиденциальность публикуемой статистической информации. Накопленный СУРС богатый опыт работы с ГИС и адекватный уровень развития пространственной (статистической) инфраструктуры в стране позволяют с полным основанием рассчитывать на достижение качественных результатов. Определение наборов данных для отображения на сетках и выработка политики их защиты считаются на сегодняшний день двумя важнейшими задачами. Идя в ногу с глобальными тенденциями в области использования пространственной статистической информации, а также следуя рекомендациям Директивы, касающимся инфраструктуры пространственной информации в Европейском сообществе (INSPIRE), СУРС намерено и впредь участвовать в международных обсуждениях и проектах, связанных с использованием статистических сеток и геостатистики в целом (например, в рамках Европейского форума по геостатистике). Оно будет стремиться также обеспечивать более широкое использование таких сеток в Словении, поощряя поставщиков пространственных данных применять сетки и в процессе сбора данных (например, данных о почвенно-растительном покрове, метеорологических данных) и тем самым расширять базу национальной сеточной статистики. Развитие ГИС и статистических сеток, бесспорно, позволяет взглянуть на предназначение национальной статистики под иным углом зрения, повышая доступность и расширяя применимость статистических данных.

## VII. Справочная литература

Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)

Izdelava hierarhičnih mrež Slovenije – zaključno poročilo ob izvedbi projekta. Geodetski inštitut Slovenije, Ljubljana, 2008.

Oblak Flander, A.: Opportunities and Challenges of a Register-Based Census of Population and Housing – the Case in Slovenia. Seminar on Registers in Statistics – methodology and quality, Helsinki, 2007.

Tammilehto-Luode, M.: Tandem II – Towards a common geographical base for statistics across Europe. Meeting of the WP “Geographical Information System for Statistics”, Luxembourg, 2003.

Wonka, E., Kaminger, I., Kuzma, I.: Regionalstatistisches grenzübergreifendes Projekt zwischen Slowenien und Österreich zum Thema Raster. Statistische Nachrichten 1, 2007

---