



Европейская экономическая комиссия

Конференция европейских статистиков

Пятьдесят восьмая пленарная сессия
Париж, 8–10 июня 2010 года
Пункт 6 предварительной повестки дня
Пространственная статистика

Подготовка и представление в Финляндии сеточных данных со времени проведения переписи 1970 года

Записка Статистического управления Финляндии

Резюме

Статистическое управление Финляндии готовит и представляет сеточные данные с разрешением в 1 км² со времени проведения переписи 1970 года. Со временем процесс подготовки и представления сеточных статистических данных претерпел определенные изменения. Такие данные находят все более широкое применение, и помимо национальных статистических учреждений их преимущества признаются многими другими организациями.

В настоящем документе описывается история создания системы статистических сеток в Финляндии, ее использование в нынешних условиях, а также ее достоинства и недостатки по сравнению с традиционной статистикой.



I. Введение

1. Сеточные статистические данные представляют собой данные, имеющие прямую географическую привязку (геоданные). Такие данные привязываются к картам с прямоугольными системами координат. Каждая квадратная ячейка сетки обозначает географический район, территориальную единицу и инструмент территориальной классификации данных.

2. Важнейшим принципом территориальной классификации статистических данных традиционно является административно-территориальная классификация. Системы кодов административно-территориальных единиц лежат в основе региональной статистики. В то же время использование административно-территориальных единиц сопряжено с определенными трудностями. Административно-территориальные единицы постоянно меняются, их размер различается как в пределах стран, так и между ними, и они, как правило, являются довольно крупными образованиями, состоящими из нескольких разных частей, и сопоставимость единиц, относящихся к разным регионам и разным странам, а также во времени является довольно низкой.

3. Принятие на вооружение национальными статистическими учреждениями технологий географических информационных систем (ГИС) расширило возможности сбора и обработки пространственной статистики. В дополнение к кодам районов или административных образований стали применяться адреса и географические координаты. Результаты опроса, проведенного Евростатом в ноябре 2009 года, показали, что 17 из 27 национальных статистических управлений, от которых были получены ответы, разработали или намеревались разработать систему подробной географической привязки данных переписей 2010/2011 года.

4. В проекте ГЕОСТАТ, запущенном Евростатом в начале 2010 года, принимают участие девять государств – членов Европейского союза (ЕС). Этот проект призван содействовать получению и гармонизации сеточных статистических данных благодаря разработке руководящих принципов и сбору примеров существующей передовой практики (ГЕОСТАТ 2010-2011). Финляндия, имеющая длительный опыт подготовки сеточной статистики, спрос на которую постоянно растет, является участником этого проекта.

II. Предпосылки появления сеточных статистических данных в Финляндии

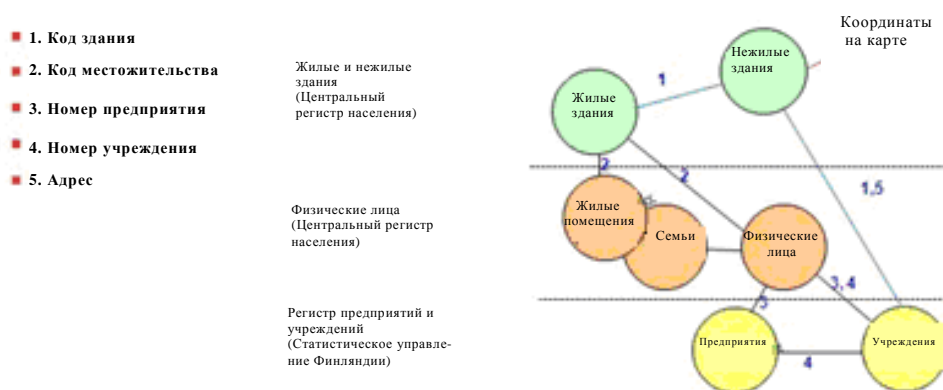
5. Отправной точкой в создании системы сеточной статистики стало решение определить в ходе переписи 1970 года географические координаты для центров всех зданий. После 1970 года появилась возможность распределять все имеющиеся отношения к зданиям данные по квадратам сетки.

6. Попытки собирать сеточные статистические данные предпринимались и до переписи 1970 года, однако они в большей или меньшей степени сводились к определению координат по картам вручную и к оценкам распределения населения. Примером такой статистики может служить статистика городских населенных пунктов за период с 1960 по 1980 годы (всеобщая перепись населения 1960 года).

7. В 1970 году в Финляндии началось поэтапное создание регистровой системы проведения переписей, благодаря чему в 1987 году появилась возможность увязывать между собой все учетные единицы с использованием идентификационных кодов физических лиц, предприятий, жилых и нежилых зданий; кроме того, стало возможным увязывать физические лица с семьями и домохозяйствами по их жилищным единицам и работодателям. Это означает также, что после 1987 года все данные переписей можно найти на карте по географическим координатам через центры зданий.

Рисунок 1

Основные учетные единицы и их связь в регистрационной статистической системе Финляндии



8. Благодаря имеющим географическую привязку кодированным регистровым данным в Финляндии появились очень хорошие возможности для использования географических данных социально-экономического характера. С другой стороны, качество сеточной статистики во многом зависит от точности и качества географических координат зданий (в регистре жилых и нежилых зданий). За последние десять лет их качество и охват заметно улучшились. Это стало возможным благодаря тесному сотрудничеству между регистрационными органами и Национальным кадастром Финляндии, а также быстрому появлению качественных цифровых карт Финляндии.

9. Традиционно сеточные данные относились к разряду специальной статистики, не предназначенной для широкой общественности. Первоначально такие данные готовились лишь по специальным заказам. В стандартном виде они стали доступны лишь после 1995 года.

10. Сегодня Статистическое управление Финляндии стремится расширить возможности визуализации в виде карт всей региональной статистики в Интернете. Общий интерес к пространственной статистике также, безусловно, растет. Одна из причин этого, возможно, кроется в принятии директивы INSPIRE¹, а также в создании национальной инфраструктуры пространственной информации. Благодаря этой директиве и национальным постановлениям появилось 34 новых вида пространственных данных для использования в природоохран-

¹ Директива INSPIRE 2007/2/ЕС, принятая Европейским парламентом и Советом 14 марта 2007 года, предусматривает создание инфраструктуры пространственной информации в Европейском сообществе. См. <http://inspire.jrc.ec.europa.eu/>

ной области. К ним относятся, в частности, система географических сеток и демография. В настоящее время ведется обсуждение роли сеточных статистических данных и их применения в национальной инфраструктуре пространственной информации. Это означает, что в будущем Статистическое управление Финляндии вместе с другими национальными статистическими органами европейских стран, возможно, будет, среди прочего, представлять стандартные метаданные, спецификации данных и оказывать сетевые услуги применительно к демографическим сеточным данным.

III. Подготовка и представление сеточной статистики

A. Модель подготовки данных

11. Так же как и другие виды данных переписей, сеточные данные в большинстве случаев имеют пространственный аспект. В рамках нынешней модели, созданной в 2005–2008 годах, все виды переписных данных хранятся в единой базе данных в формате SQL. В этой реляционной базе данных в основе территориальных классификаций лежат различные виды географической информации, относящейся к разным зданиям. Это означает, что каждому зданию присвоены все географические коды, используемые в статистических целях: от провинциальных до муниципальных кодов и от почтовых кодов до сеточных и географических координат. Вся качественная информация в этой базе данных может быть привязана к "таблице местоположения зданий", из которой можно получить географические коды, необходимые для компилирования статистических данных по разным областям и разным территориальным единицам.

12. Границы статистических районов определены в центральной базе географических данных, содержащей сведения о сетках с двумя разными размерами ячеек. Статистические данные можно привязать к данным о границах при помощи географических кодов. С другой стороны, зданиям, расположенным в пределах разных территориальных единиц, географические коды присваиваются по пространственному принципу; это означает, что все здания в пределах определенного квадрата сетки или района с единым почтовым кодом получают соответствующий географический код. Этот принцип не распространяется на административные географические коды, которые обычно присваиваются первичным регистрационным данным.

Рисунок 2

Подготовка сеточных данных, аналогичных другим видам территориальной статистики с пространственными характеристиками



13. Статистическое управление Финляндии имеет две общенациональные сетки: одну с площадью ячейки в 1 км^2 и вторую с площадью ячейки в 250 м^2 . Эти сетки представляют собой векторные системы с фиксированной исходной точкой, покрывающие всю территорию Финляндии. Географической привязкой каждого квадрата сетки являются координаты его левого нижнего угла. Отдельные квадраты сетки имеют и иной идентификационный код – порядковый номер каждой ячейки. Квадраты сетки содержат также информацию о муниципалитетах, к которым они относятся. Всем квадратам, занимающим пограничное положение между двумя муниципалитетами, присваивается код того муниципалитета, которой занимает большую площадь квадрата.

14. В настоящее время финская сеточная статистика опирается на сети, построенные на основе единой национальной системы координат (GRID27E). Финляндия готовится заменить свою национальную систему координат Европейской земной системой координат (ETRS89)², которая должна стать одной из основных европейских систем наземной привязки и которая в будущем станет основной системой координат для сеточной статистики в Финляндии. Переход на новую систему координат должен осуществляться на уровне исходных данных из компиляции по квадратам сетки, что делает эту задачу довольно трудоемкой.

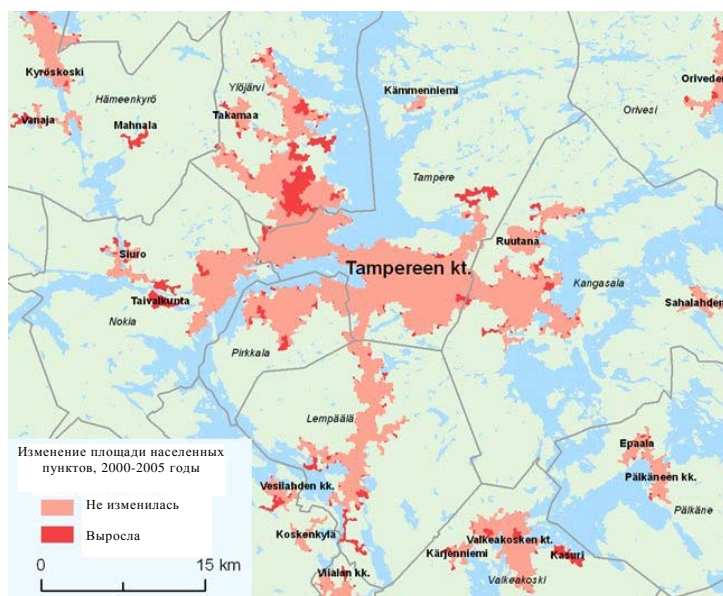
В. Продукты и услуги

15. Статистическое управление Финляндии раз в пять лет на основе сеточных данных проводит делимитацию границ населенных пунктов и городских поселений, которые затем принимаются за границы статистических районов как при подготовке общецелевой статистики, так и при классификации данных и муниципалитетов на сельские и городские³.

² <http://etrs89.ensg.ign.fr/>.

³ http://www.stat.fi/meta/luokitukset/kuntaryhmitys/001-2009/index_en.html.

Рисунок 3
 Делимитация границ городского поселения Тампере в два разных года.
 Темно-красные зоны обозначают растущие районы. Серые линии
 обозначают границы муниципалитетов



16. Большая часть сеточных данных имеет два вида применения: такие данные готовятся и хранятся для использования либо в стандартных продуктах, либо в специальных продуктах, заказываемых отдельными клиентами.

17. Одним из стандартных продуктов является база сеточных данных, в которой представлены все сетки, содержащие агрегированные данные восьми групп⁴. Ее можно получить либо в целом виде, либо по отдельным модулям. После подписания контракта клиент может скачать базу данных с вебсайта, используя свой идентификационный код. Цена такой услуги зависит от объема данных и от числа пользователей.

18. База сеточных данных предлагается в формате MapInfo или в формате Института исследований экологических систем (ИИЭС). Эти данные могут использоваться с программным обеспечением ГИС напрямую даже без отдельной сетки, которую также можно получить в случае необходимости. База сеточных данных и соответствующие сетки с 2003 года ежегодно обновляются.

19. Содержание и формат подготовленных по заказу сеточных данных различаются в зависимости от доступности информации и степени ее дезагрегирования. При этом всегда должны учитываться также качество данных и необходимость принятия мер по защите конфиденциальности. В соответствии с Законом о статистике конфиденциальные сеточные данные, так же как и все другие статистические данные, могут предоставляться в статистических и исследовательских целях при соблюдении определенных требований и при условии подписания письменного соглашения об ограничении сферы их использования.

20. Кроме этого, "сырая" сеточная статистика время от времени уточняется самим статистическим ведомством, которое предлагает клиентам лишь результаты пространственного анализа. Примерами результатов такого анализа могут

⁴ http://tilastokeskus.fi/tup/ruututietokanta/index_en.html.

служить статистические данные о населении малонаселенных районов или о расстоянии до некоторых объектов, таких как электростанции, автомагистрали или береговая линия.

С. Защита данных

21. В таких малонаселенных странах, как Финляндия, сеточной статистике присущи проблемы конфиденциальности, особенно применительно к сельским районам. Увеличение размера квадрата сетки не всегда является решением проблемы. С другой стороны, изъятие данных по соображениям конфиденциальности может сильно отразиться на результатах пространственного анализа. Таким образом, для того чтобы понимать возможные последствия для результатов анализа, пользователи должны знать, какой обработке подвергались конфиденциальные данные.

22. Для учета особенностей географической информации был разработан метод местной ограниченной условной оценки (МОУО) (Markkula, 2003). МОУО обеспечивает защиту данных на местном уровне, что гарантирует их точность на иерархически более высоком уровне. Этот довольно многообещающий метод лучше всего подходит для исследований, проводимых на базе географической информации.

23. На сегодняшний день МОУО не применяется к сеточной статистике как таковой. Защита данных осуществляется в соответствии с общими руководящими принципами контроля за раскрытием демографической статистики путем простого изъятия данных. Эффективность таких мер, как правило, зависит от того, к какой группе относятся соответствующие статистические переменные. Поэтому считается, что риск разглашения персональных данных зависит в основном от чувствительности переменных и числа соответствующих случаев в каждом квадрате сетки (Tammilehto-Luode 2001).

Д. Качество данных

24. Пространственная точность исходных данных (данных, связанных со зданиями и их координатами на карте) играет важную роль при компиляции статистики по небольшим районам, например сеточной статистики. Сегодня, когда данные о населении и местах его проживания привязаны к зданиям, координатный охват составляет 99%. Когда к зданиям привязываются места работы, координатный охват составляет 92%. Для восполнения пробелов в данных очень важно сотрудничество между регистрационным органом (Регистром жилых и нежилых зданий), Статистическим управлением Финляндии и пользователями данных. Процедура обратной связи с центральным регистром по-прежнему находится в стадии разработки.

25. Для повышения качества пространственных данных Статистическое управление Финляндии ежегодно проводит выверку данных. Изучаются и корректируются все случаи, когда здания не имеют координат или когда их местоположение не соответствует муниципальному коду. Проводится перекрестная сверка данных Регистра коммерческих предприятий и зданий и Регистра жилых и нежилых зданий. Несовпадающие адреса из этих двух источников изучаются более подробно. Дополнительным методом обеспечения качества данных является присвоение зданиям географических кодов на основе адресных данных до-

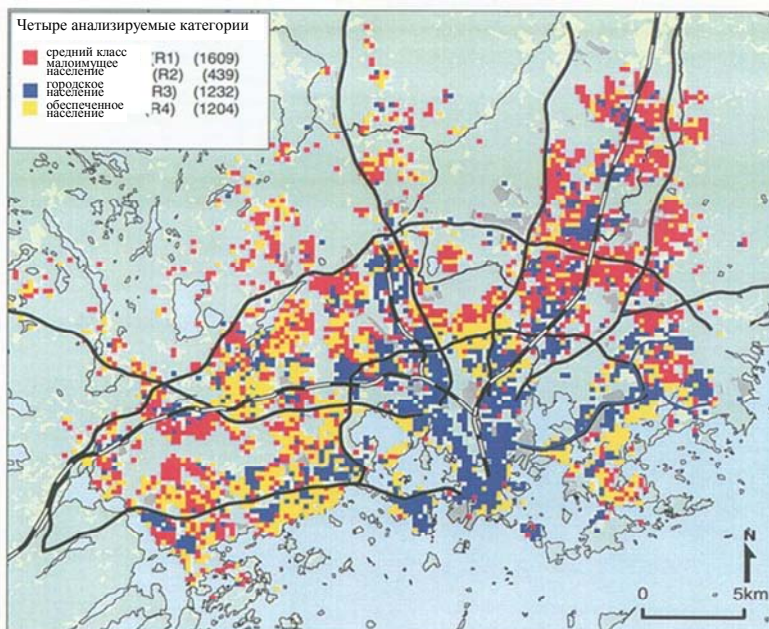
рожной сети. Используется также инструментарий ГИС, позволяющий актуализировать привязку адресов к координатам на карте.

IV. Примеры конечного использования и конечных потребителей

26. На протяжении длительного времени пользователями сеточных данных являлись главным образом исследователи. Сеточная статистика рассматривалась как практический вычислительный инструмент моделирования и изучения пространственных изменений, не зависящих от административных границ. Исследователи с кафедры географии Университета Улу стали первопроходцами в деле использования социально-экономических данных сеточной статистики. Насчитывающая почти 20 лет программа "Динамика региональной структуры" представляет собой их рамочную исследовательскую программу, содействующую анализу различных аспектов финских региональных структур и различных явлений человеческой деятельности на региональном уровне (см., например, Rusanen 2009). На основе сеточных статистических данных в Университете Улу было написано несколько докторских диссертаций (например, Vaattovaara 1998). С начала 1990-х годов другие университеты и исследовательские институты стали время от времени пользоваться сеточными данными в целях дифференциации жилых районов, изучения экологических последствий, управления рисками, изучения проблем потребительского рынка, городских и сельских районов, а также географии здравоохранения. Многие факультеты географии финских университетов используют сеточные статистические данные в качестве учебного материала для студентов.

Рисунок 4

Дифференциация жилых районов Хельсинки на четыре основных класса по результатам факторного анализа сеточных данных по 20 переменным (Vaattovaara 1998)

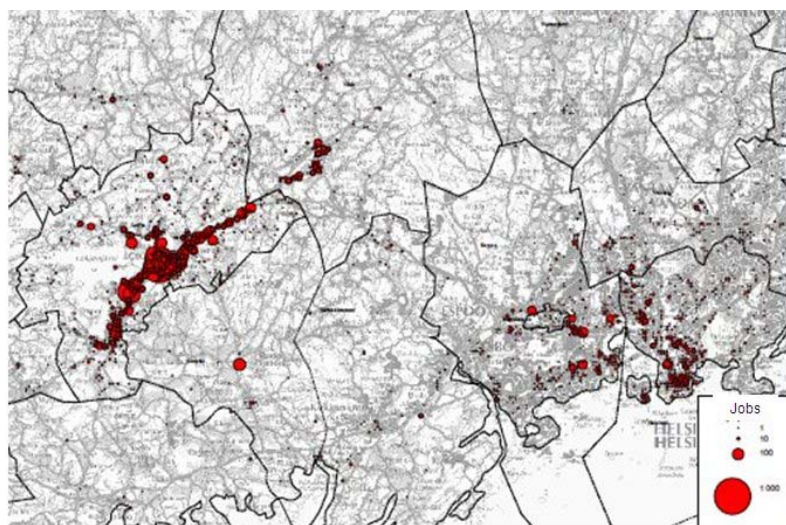


Источник: Статистическое управление Финляндии.

27. Одним из основных пользователей сеточных статистических данных из государственного сектора является Финский институт окружающей среды (ФИОС). Этот институт создал систему географической информации под названием "Система мониторинга изменений в городской структуре", которая широко используется органами государственного управления в области городского и регионального планирования. Эта система содержит общенациональные сеточные данные о населении, рынке жилья, рабочих местах и их удаленности от мест проживания рабочей силы за период с 1980 по 2008 год. Эта система, которая работает на базе Интернет, позволяет пользователям обобщать сеточные данные в пределах определяемых ими районов, отслеживать изменения во времени и применять стандартные методы тематического анализа (Oinonen 2007).

Рисунок 5

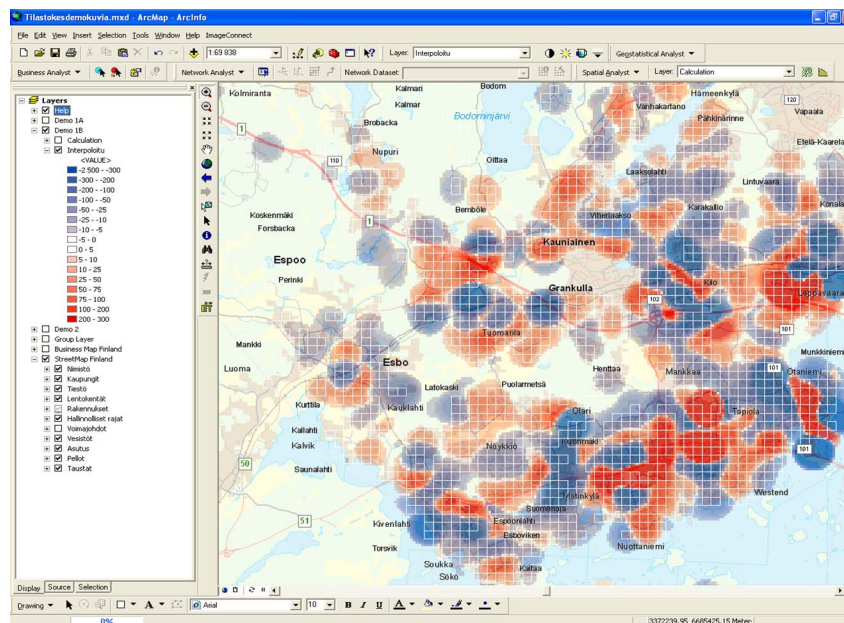
Распределение рабочих мест работников наемного труда, проживающих в городе Лохья



28. Число частных предприятий, пользующихся сеточными статистическими данными в коммерческих целях, растет из года в год. В последние три года частный сектор являлся источником почти 70% поступлений за использование базы сеточных данных. Типичными потребителями такой информации являются предприятия связи, банки, предприятия розничной торговли, страховые и консалтинговые компании. Сфера применения ими сеточной статистики очень широка: от маркетинга, выбора мест для хозяйственной деятельности, управления рисками и планирования до создания различных видов услуг. Существуют также предприятия, перепродающие информацию из базы сеточных данных, главным образом продавцы программного обеспечения ГИС, однако на сегодняшний день их вклад нельзя назвать существенным.

Рисунок 6

Анализ (проведенный частной компанией) изменения числа рабочих мест в период с 2003 по 2005 год в южных районах Финляндии. Области, где число рабочих мест сократилось, окрашены голубым цветом, а те, где их число увеличилось, – красным. Этот анализ был проведен в целях планирования строительства нового жилья в районах, где наблюдается рост количества рабочих мест



29. Время от времени запросы на сеточную статистику поступают из-за рубежа, главным образом от международных предприятий, планирующих обосноваться на финском рынке, или от исследователей, занимающихся трансграничными исследованиями.

V. Потенциал и проблемы

30. В Финляндии источником сеточных данных в основном являются переписные данные, которые регулярно обновляются сведениями из административных регистров. Существуют и другие потенциальные источники сеточных данных, например регистр коммерческих предприятий, однако на сегодняшний день они не находят столь широкого применения, как данные переписей.

31. Финские данные из административных источников имеют точную географическую привязку (или они могут быть привязаны к зданиям с координатами на карте), что обеспечивает гибкий пространственный подход ко многим статистическим темам. В то же время бесплатные статистические данные имеют лишь по крупным административным районам. С другой стороны, для создания качественной пространственной информации и защиты конфиденциальности статистических данных по небольшим районам требуются ресурсы, наличие которых сегодня обеспечивают поступления от продажи соответствующих услуг.

32. В настоящее время растет спрос на сеточную статистику и все виды географической информации, а также потребности в них. Директива INSPIRE и правила ее осуществления устанавливают дополнительные требования к публи-

куемой пространственной статистике. В этой области, безусловно, кроется огромный потенциал, для реализации которого требуются дополнительные средства.

33. Сегодня ведется обсуждение того, какую роль национальные статистические органы могли бы играть в регулярной разработке новых видов показателей на основе сеточных данных. На сегодняшний день сеточные данные, охватывающие всю национальную территорию, имеются в 11 европейских странах (GEOSTAT 2010). Примеры новых интересных видов статистики, в которых используются сеточные данные, появляются в скандинавских странах. Речь идет о данных о доступности важнейших услуг резидентам, зеленых зонах и общественном транспорте, населении городских и сельских районов и малонаселенных районов, населении, проживающем за полярным кругом, концентрации некоторых видов рабочих мест и т.д.

34. В целом использование статистических сеток для расчета сопоставимых территориальных данных и статистических временных рядов открывает широкие возможности (Tammilehto-Luode et al. 2003). Эти сетки могут составляться на гибкой основе по крупным районам или по районам, ограниченными естественными границами, по расстояниям или по другим пространственным характеристикам. Они представляют собой простую в использовании и надежную базу для отслеживания пространственных изменений по различным параметрам. Помимо этого, они являются хорошим инструментом гармонизации данных по разным видам территориальных единиц в тех случаях, когда требуется систематизация данных по признаку местонахождения.

35. Стабильная сеточная статистика может стать хорошей альтернативой статистическим данным, компилируемым по административным районам, поскольку территориальное деление не меняется и данные, собираемые по территориальному признаку, сопоставимы между регионами. Так, например, последовательная компиляция статистических данных по квадратам сетки позволит получить сопоставимую статистику по городским и сельским районам разных стран, которая отсутствует на сегодняшний день (Backer et al. 2002).

36. Использование сеточных данных открывает колоссальные возможности в сфере трансграничных исследований и в разработке различных показателей, в случае которых зависимость данных от пространственных образований, к которым они относятся, является очень высокой. Поскольку по сравнению с традиционными статистическими районами размер сетки, как правило, невелик, она может лучше описывать реальное пространственное распределение изучаемого явления. Использование небольших, соразмерных территориальных единиц, подобных сеткам, позволяет частично избежать проблем, возникающих при использовании средних показателей для описания региональных различий (Martin 1998).

37. К преимуществам сеточной статистики относится возможность отображения "незаселенных" районов или других несуществующих переменных в пространстве. Было бы довольно непросто описать процесс депопуляции в странах северной Европы, если статистические данные имелись бы лишь в разбивке по административным единицам (Harala – Tammilehto – Luode, 1999).

38. Сегодня сеточные данные компилируются главным образом для национальных нужд в отсутствие каких-либо международных стандартов. Однако стандарты требуются как минимум для трансграничных и международных исследований. Введение стандартов планируется в ближайшее время в рамках

деятельности по созданию Инфраструктуры пространственной информации в Европейском сообществе (INSPIRE).

39. Квадрат сетки представляет собой абстрактную искусственную пространственную единицу, которую иногда непросто объяснить тем, кто не знаком с данной системой. Часто требуется представлять результаты наглядно, указывая местоположение сеток на карте. При составлении таблиц квадраты сетки приходится укрупнять до административного или какого-либо иного известного регионального уровня с тем, чтобы облегчить толкование конечных данных.

40. Сеточная статистика хорошо сочетается с картографическими алгебраическими приемами, применяемыми во многих пакетах программного обеспечения ГИС. В то же время необходимо уметь ей пользоваться и правильно толковать результаты. Та легкость, с которой сеточными данными можно манипулировать, требует ответственного отношения к их отображению на картах.

41. Ошибки в сеточных данных с трудом поддаются обнаружению и исправлению. Ошибки в основном возникают из-за низкого качества "сырых" данных, например из-за использования неточных и неверных методов географической привязки исходных данных. Данные должны иметь свою спецификацию качества, которую иногда довольно трудно составить. Решающее значение имеет размер используемого квадрата сетки, который следует выбирать с учетом соображений качества и конфиденциальности.

42. Если в квадратах сетки зарегистрировано, например, небольшое число жителей или работников, контроль за разглашением данных должен быть очень строгим. Часто исследователям необходимо знать точное местоположение каждого квадрата. В то же время для некоторых потребностей, например для планирования хозяйственных объектов и рыночного анализа, требуется информация обо всех сетках. В статистике небольших районов, таких как сеточная статистика, контроль за раскрытием информации представляет собой одну из важнейших задач, значение которой признается потребителями пространственной статистики во многих странах.

VI. Справочная литература

Backer, L., M. Tammilehto-Luode & P. Gublin (2002). Tandem GIS_I. A Feasibility study towards a common geographical base for statistics across European Union. Eurostat. Working papers.

Business Register. http://stat.fi/tup/yritysrekisteri/index_en.html

Eurostat (2010). Geo-referencing of Data from the 2011 Censuses of Population and Housing. Consultation November 2009. Document: E/GIS/103/EN. Meeting of Working Party "Geographic Information Systems for Statistics". Luxembourg, March 08-09, 2010.

General Census of Population 1960. Non-administrative urban settlements and their boundaries, Official Statistics of Finland VI C: 103. Helsinki 1965.

Grid Database. http://stat.fi/tup/ruututietokanta/index_en.html

GEOSTAT 2010-2011. Representing census data in a European population grid. ESSnet project. Grant agreement for an action with multiple beneficiaries. Agreement number 50502.2009.004-2009.860. Eurostat.

Harala, R. & M. Tammilehto-Luode (1999). GIS and Register-based Population Census. *Statistics, Registers and Science*. Edited by J. Alho. Statistics Finland. Helsinki.

INSPIRE 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). <http://inspire.jrc.ec.europa.eu/>

Markkula, J. (2003). Geographic Personal Data, their Privacy Protection and Prospects in a Location-based Service Environment. *Jyväskylä Studies in Computing 30*. University of Jyväskylä. Jyväskylä

Martin, D. (1998). Census output areas. From concept to prototype. *Population Trends 94*, 19-24.

Oinonen, Kari (2007). Monitoring system for changes in urban structure. Nordic Forum for Geo-Statistics Seminar 2007. Helsinki.

http://www.stat.fi/geostatistics2007/session2_oinonen_presentation.pdf

Rusanen Jarmo (2009). List of publications. <http://www.oulu.fi/geography/staff/professors/publications/rusanen.pdf>

Tammilehto-Luode, M. (2001). Disclosure control for demographic statistics. *Statistical Journal of the United Nations Economic Commission for Europe*. Volume 18:4. ISSN 0167-8000.

Tammilehto-Luode, M., M. Ralphs & L. Backer (2003). Tandem II: Towards a common geographical base for statistics across Europe. The Final Report. Unpublished. Eurostat. Luxembourg.

Vaattovaara, Mari (1998). Residential differentiation within the Metropolitan area of Helsinki. Pääkaupunkiseudun sosiaalinen erilaistuminen. City of Helsinki Urban Facts Research Series 1998:7.
