



Европейская экономическая комиссия

Конференция европейских статистиков

Шестьдесят восьмая пленарная сессия

Женева, 22–24 июня 2020 года

Пункт 5 b) предварительной повестки дня

**Заседание 2: Опыт и результаты конкретных шагов,
уже предпринятых сообществом НСУ и геопространственным
сообществом с целью модернизации их роли**

Реализация потенциала интеграции геопространственных и статистических данных в период расширения возможностей¹

**Записка, подготовленная Центральным статистическим
управлением Польши**

Резюме

Настоящий документ представлен для обсуждения на семинаре Конференции европейских статистиков на тему «Новая роль национальных статистических управлений и геопространственных агентств в формирующихся национальных экосистемах данных» для заседания 2: «Опыт и результаты конкретных шагов, уже предпринятых сообществом НСУ и геопространственным сообществом с целью модернизации их роли».

¹ Настоящий документ был запланирован к изданию после установленного срока в силу обстоятельств, не зависящих от представившей его стороны.



I. Введение

1. На протяжении ряда лет органы государственной статистики Польши занимаются разработкой и использованием пространственных данных. С учетом современных технологий, информационных стандартов и технологических потребностей при проведении переписей и обследований Центральное статистическое управление провело работу по обеспечению пространственной идентификации объектов на основе технологий пространственной привязки и ГИС. Геопро пространственная информация приобретает все большее значение в процессе статистического производства и повышает роль статистических управлений в интеграции геопро пространственных и статистических данных.

II. Модель выпуска статистической продукции

2. В настоящее время Центральное статистическое управление Польши разрабатывает *модель выпуска статистической продукции* (Model Procesu Produkcji Statystycznej – МВСП) на основе *Типовой модели производства статистической информации* (ТМПСИ)², которая была дополнена геопро пространственными компонентами. Для точного анализа внедренных в Центральном статистическом управлении Польши процедур в разрезе соответствия пространственных данных процедурам, рекомендованным в типовой модели, соответствующие бизнес-процессы сопоставлялись с моделью ТМПСИ. Данная модель представляет собой типовой эталонный стандарт процесса производства статистической информации.

3. В основу анализа было положено сопоставление бизнес-функций на каждом этапе ТМПСИ с фактическими и запланированными процессами и бизнес-функциями, реализованными в рамках статистического производства, с особым акцентом на пространственные данные.

4. Анализ соответствия проводился путем увязки различных мероприятий Центрального статистического управления Польши с этапами, процессам и подпроцессам, рекомендованными в ТМПСИ. Практическое осуществление соответствующих процессов, связанных с пространственными данными, и их сопоставление с моделью ТМПСИ показали, что модель не охватывает некоторые важные области. Благодаря выявлению этих областей были установлены потенциальные недостатки модели ТМПСИ. Недостатки в основном касались аспектов «пространственной привязки» статистических данных от этапа планирования методики сбора данных до геокодирования и анализа и распространения статистической продукции с пространственными характеристиками. Эти выявленные геопро пространственные подпроцессы, которые следует включить в архитектуру процесса статистического производства, приведены ниже:

- этап 2: подпроцесс «2.5а Проектирование генеральной совокупности геокодирования, выборки и сбора данных»;
- этап 4: подпроцессы «4.1а Генеральная совокупность и выборка геокодов» и «4.3а Сбор геокодов»;
- этап 6: подпроцесс «6.2а Подготовка пространственного анализа и карт»;
- этап 7: подпроцесс «7.2а Управление пространственным анализом и картами с использованием ГИС».

5. Благодаря включению перечисленных выше подпроцессов в модель выпуска статистической продукции добавляется пространственный компонент. Это позволит лучше понимать пространственные данные и их роль в процессе статистического производства, а также стандартизировать методики, объединяющие статистические данные с пространственными. Эти новые геопро пространственные подпроцессы (желтые блоки на диаграмме ниже) предусмотрены в *модели выпуска статистической*

² <https://statswiki.unece.org/display/GSBPM/IV.+Levels+1+and+2+of+the+GSBPM>.

продукции, на основе которой в Польше осуществляется статистическое производство и интеграция статистических и геопространственных компонентов в интересах более качественного комплексного геопространственного анализа.

Таблица
Модель выпуска статистической продукции

1 Specify Needs	2 Design	3 Build	4 Collect	5 Process	6 Analyse	7 Disseminate	8 Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update outputs systems	8.1 Gather evaluation inputs
1.1a Prepare Statistical Programme (SSPPS)	2.2 Design variable descriptions	3.2 Reuse or build processing & analysis components	4.1a Geocode frame & sample	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.2 Consult & confirm needs	2.3 Design collection	3.3 Reuse or build dissemination components	4.2 Set up collection	5.3 Review & validate	6.2a Prepare spatial analyses & maps	7.2a Manage spatial analyses & maps using GIS	8.3 Agree an action plan
1.3 Establish output objectives	2.4 Design frame & sample	3.4 Configure workflows	4.3 Run collection	5.4 Edit & input	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	
1.4 Identify concepts	2.5 Design processing & analysis	3.5 Test production system	4.3a Geocode collection	5.5 Derive new variables & units	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5a Design geocoding frame, sample & data collection	3.6 Test statistical business process	4.4 Finalise collection	5.6 Calculate weights	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production system & workflow	3.7 Finalise production system	4.5 Verify plans	5.7 Calculate aggregates		7.6 Satisfaction research	
1.7 Prepare a general plan for the implementation of the production process	2.7 Design quality indicators	3.8 Verify projects and construction plans		5.8 Finalise data files		7.7 Prepare an improvement plan for the next research cycle	
1.8 Prepare a detailed plans for subsequent phases	2.8 Verify the general plan and detailed plans						
1.9 Verify the concept to improve satisfaction							

III. Географическая информационная система, поддерживающая процесс статистического производства

6. Геопространственные компоненты играют все более значительную роль в процессе статистического производства в Польше. Для проведения анализа, показывающего территориальную составляющую динамики различных явлений, необходимы статистические данные с очень точной информацией о местоположении. Только официальная статистика собирает все необходимые данные статистических единиц с привязкой к определенным точкам (в прямоугольной системе координат), что необходимо для дальнейшей обработки и получения агрегированных данных, позволяющих проводить достоверный и высококачественный геопространственный анализ. Для целей переписи населения также нужны данные с привязкой к местности на уровне прямоугольной системы координат.

7. В целях представления результатов предыдущего цикла переписей – сельскохозяйственной переписи 2010 года и переписи населения и жилищного фонда 2011 года – Центральное статистическое управление Польши использовало технологию ГИС. В целях содействия распространению геокодированной статистической информации была создана и запущена платформа визуализации данных для внутренних и внешних пользователей – Геостатистический портал³.

³ <https://geo.stat.gov.pl/imap/?locale=en>.

8. Геостатистический портал является инструментом интерактивного картографического представления и публикации агрегированной статистической информации, получаемой на основе различных видов пространственных данных и анализа в виде карт. Он предназначен для сбора, представления и распространения информации среди широкой аудитории, включая органы государственного управления, предпринимателей, индивидуальных пользователей и научно-исследовательские учреждения. Портал является весьма полезным инструментом принятия стратегических решений на каждом уровне управления. Он представляет собой комплексное решение, отвечающее европейским стандартам, и все данные благодаря проводимой обработке представляются в обезличенном виде, обеспечивающем конфиденциальность статистической информации.

9. На нынешнем этапе появления новых возможностей Центральное статистическое управление Польши работает над расширением Геостатистического портала, концентрируя усилия на еще одном проекте по разработке инновационного передового инструмента нового поколения, который будет служить Статистическому управлению подспорьем при выполнении функций учреждения, занимающегося геопространственными статистическими данными и более сложным детальным анализом данных.

10. Этот дополнительный проект носит название «Пространственные статистические данные в информационной системе государства» (ПДС). Его целью является расширение охвата и повышение доступности статистической информации и методов геостатистического анализа с использованием ресурсов официальной статистики. Проект ПДС разработан с учетом потребностей, выявленных после запуска существующего Геостатистического портала, а также новых источников данных, IT-инструментов и технологий с открытым исходным кодом. Эта цель будет достигнута путем разработки современной ИТ-платформы, позволяющей использовать в статистических исследованиях географическую информационную систему и проводить более современный и сложный геопространственный анализ.

11. В рамках проекта ПДС функциональные возможности прежних сервисов Геостатистического портала будут расширены за счет включения функций проведения статистического анализа в рамках любой пространственной единицы, определяемой, например, внешними пользователями, скопированной из внешних пространственных сервисов (WFS), на основе сетки динамических квадратов, а также возможности комбинирования статистических данных с собственными данными пользователя или геокодирования заданных пользователем характеристик для использования в геостатистическом анализе. Кроме того, будут разработаны новые сервисы, которые облегчат применение различных видов исследовательского анализа пространственных данных с использованием статистической информации и анализ в области геостатистического моделирования и будут допускать дополнение собственных данных пользователей геостатистической информацией и результатами геостатистического анализа.

12. В результате проекта ПДС появятся три совершенно новых внешних сервиса:

- исследовательский геостатистический анализ данных с использованием имеющейся на портале статистической информации;
- геостатистический анализ методом моделирования;
- дополнение пользовательского контента геостатистической информацией и результатами анализа, имеющимися на портале.

13. Исследовательский геостатистический анализ данных с использованием имеющейся на портале статистической информации позволит пользователям:

- импортировать и дополнять собственные данные статистическими данными, имеющимися в системе ПДС;
- дополнять данные дополнительными переменными (так называемыми производными переменными), рассчитанными на основе уже импортированных

и дополненных данных пользователя с использованием имеющегося набора математических и геостатистических функций;

- исследовать собранные данные (дополненные производными переменными) в геостатистическом разрезе, оценивать их качество, а также дополнять и изменять данные;
- визуализировать и публиковать полученные результаты анализа.

14. Пользователям будет доступен широкий круг методов статистического пространственного анализа, включая статистику центра распределения, статистический разброс, кластерный анализ и пространственную автокорреляцию. Кроме того, сервис геостатистического анализа методом моделирования даст возможность создавать и применять вероятностные модели, позволяющие проводить исследование (оценку) величины зависимой переменной на основе результатов обследования по вероятностной выборке (т. е. на основе собранных данных).

15. Дополнение пользовательского контента геостатистической информацией и результатами анализа, имеющимися на портале, позволит пользователям:

- импортировать собственный контент для анализа;
- проводить анализ контента с использованием механизмов «анализа текстовых данных» для выявления ключевых слов, относящихся к категориям объектов, имеющихся в системе ПДС, и надлежащей классификации пользовательского контента;
- выбирать объекты в системе ПДС, больше всего совпадающие с выявленными словами и ключевыми фразами, и устанавливать корреляцию между объектами, разыскиваемыми в системе ПДС, для отбора имеющихся в этой системе видов анализа;
- конфигурировать и проводить пространственный анализ по выбору пользователя в графической форме;
- обеспечивать доступ к результатам отобранных видов пространственного анализа в графической или табличной форме.

16. Кроме того, в рамках проекта ПДС будет создан новый внутренний сервис, облегчающий эффективное ведение баз пространственных данных об адресах.

IV. Выводы и рекомендации

17. Статистические данные с точной географической привязкой необходимы для облегчения сравнительного анализа и повышения качества подготовки статистических геопространственных данных. В период расширения возможностей статистическое ведомство может повысить свою роль, создавая инструменты статистического анализа нового поколения.

18. Геостатистический портал используется для пространственного представления крупнейшего ресурса статистической информации Польши и позволяет публиковать агрегированные статистические данные по итогам различного рода пространственного анализа в виде карт с соблюдением принципов статистической конфиденциальности. Созданный портал ПДС позволит проводить весьма сложные виды анализа статистической информации, например исследовательский геостатистический анализ данных или геостатистическое моделирование. Сильные стороны и уникальный характер существующего Геостатистического портала и планируемой новой платформы ПДС объясняются возможностью проведения надежного пространственного анализа на основе данных статистических единиц, имеющихся только у органов официальной статистики.

19. Система ПДС будет разработана в соответствии с *моделью выпуска статистической продукции* и станет элементом более широкой архитектуры статистического производства в официальной статистике Польши. Данный портал

будет обеспечивать доступ к официальным статистическим данным и результатам статистического анализа, крайне необходимым для функционирования государства, местных органов власти и местных сообществ. Информация будет доступна в удобном графическом формате, ускоряющем процесс принятия решений. На следующем этапе поставлена цель обработки и представления результатов цикла переписей 2020–2021 годов.
