

# Использование гетерогенной пространственной информации при решении задач устойчивого развития территорий

*Кудашев Е.Б., Марков С.Ю., Попов М.А.*

## Аннотация

Статья посвящена проблемам интегрирования гетерогенных геопространственных данных при решении задач устойчивого развития территорий. Предложены концептуальные подходы к построению электронной инфраструктуры доступа к гетерогенной пространственной информации, проанализированы основные проблемы интегрирования разнородных данных, разработана модель интегрирования геопространственных данных на основе архитектуры SOA.

**Ключевые слова:** геопространственные данные, гетерогенность, метаданные, инфраструктура, интегрирование данных.

## Введение

Эффективное решение комплексных задач устойчивого развития территорий требует наличия максимально полной и разнообразной информации о территориальных объектах, причем большая часть этой информации имеет геопространственную природу. При этом глубокое рассмотрение и понимание происходящих процессов возможно только на основе интегрирования всех доступных разнородных данных. Однако при реализации интеграционных процедур возникает ряд проблем, основными из которых являются:

- отсутствие, низкое качество или плохая доступность метаописаний имеющихся наборов геопространственных данных (метаданных), что затрудняет поиск нужной информации под конкретные задачи;
- отсутствие или несовершенство инфраструктур для поиска и распространения геопространственных данных;
- несовместимость некоторых форматов геопространственных данных;
- гетерогенность геопространственных данных.

Для решения этих проблем, как показывает практика, целесообразно создать отраслевую или корпоративную инфраструктуру геопространственных данных, которая позволит наиболее эффективно использовать имеющиеся информационные ресурсы, описывающие территориальные объекты. Сегодня активно ведутся научные исследования в области создания локальных, региональных, национальных и глобальных инфраструктур геопространственных данных.

## 1. Концептуальные основы создания электронной инфраструктуры геопространственных данных

Основой создания электронной инфраструктуры геопространственных данных (е-Инфраструктуры) является комплексная оценка потребностей ее пользователей. При этом необходимо руководствоваться следующими принципами [1]:

- ориентированность на бизнес - процессы пользователей, а также на необходимые для их реализации геопространственные данные;
- гарантирование качества геопространственных данных / сервисов, предоставляемых пользователям е-Инфраструктурой;
- применение итеративного жизненного цикла разработки е-Инфраструктуры, что позволит своевременно корректировать ее архитектуру соответственно степени удовлетворенности пользователей;
- гибкое управление е-Инфраструктурой путем оперативного внесения изменений, в которых нуждаются пользователи;
- документирование потребностей пользователей на протяжении всего жизненного цикла разработки е-Инфраструктуры;
- приоритет архитектуры е-Инфраструктуры, от которой должны зависеть структуры данных, аппаратно-программное обеспечение и технологические решения;
- широкое использование национальных и международных стандартов.

Оценка потребностей пользователей является основной отправной точкой разработки е-Инфраструктуры. Такие оценки должны выполняться на всех этапах ее создания, чтобы вовремя корректировать перечень предоставляемых сервисов в зависимости от изменений требований пользователей. Кроме того, при построении е-Инфраструктуры очень важными являются общие принципы создания инфраструктур пространственных данных [2]:

- открытость, т. е. условие, что е-Инфраструктура должна базироваться на открытых стандартах и спецификациях. К таким спецификациям относятся, в частности, спецификации консорциума OpenGeoSpatial [3], которые в настоящее время являются стандартами "де-факто" открытых систем для предметной области геоматики;
- прозрачность, т. е. обеспечение правила, что пользователи данных / сервисов, которые предоставляются е-Инфраструктурой, не должны понимать механизмы и технологии информации, формирующие эти данные / сервисы; партнерство, что означает возможность совмещения на базе единых стандартов и технологий информации, циркулирующей как в государственных структурах, так и в частном секторе.
- эволюция, то есть усовершенствование ресурсов системы путем привлечения к этому процессу всех участников е-Инфраструктуры;
- своевременность, т. е. использование таких технологических решений, которые обеспечат доступ к информационным ресурсам в режиме реального времени;
- совместная поддержка, которая строится на общих усилиях участников е-Инфраструктуры в процессе ее поддержания в рабочем состоянии;
- самоорганизация, означающая возможность для участников, желающих сделать информационный вклад в е-Инфраструктуру, не тратить лишних усилий на внесение данных, а также то, что для функционирования инфраструктуры нужны лишь минимальные затраты на централизованное администрирование.

Таким образом, концептуальными основами создания е-Инфраструктуры являются базовые принципы построения инфраструктуры

пространственных данных, однако при этом необходимо учитывать особенности бизнес-процессов основных пользователей, а также специфику проблем устойчивого развития, на которые будет ориентирована эта е-Инфраструктура.

## 2. Проблема гетерогенности геопространственных данных

Главным проблемным фактором при решении проблемы интегрирования информации в е-Инфраструктуре является, безусловно, гетерогенность данных. Выделяют [1] следующие основные виды гетерогенности, присущие данным:

- синтаксическая гетерогенность, которая обусловлена различными типами данных, форматами их сохранения, допустимыми значениями, единицами измерения и т. п.;
- схематическая, или структурная гетерогенность, которая обусловлена различными моделями данных, использованными при моделировании объектов реального мира (векторная, растровая, реляционная, объектная и т. п.);
- семантическая гетерогенность, которая проявляется в виде несоответствий в способах отображения одних и тех же объектов в компьютерной среде. Этот вид гетерогенности включает в себя:
  - концептуальную (когнитивную) гетерогенность, обусловленную тем, что разные люди при моделировании определенного явления создают собственное представление (абстракцию);
  - формализационную гетерогенность, которая является результатом использования различных языков (языковая гетерогенность) и терминов (репрезентационная гетерогенность) при формировании концепций;
  - контекстную гетерогенность, которая относится к изменению смысла концептуальных построений под влиянием различных условий и обстоятельств.

Необходимо отметить, что гетерогенность присуща не только данным, но и системам. Могут быть определены следующие разновидности гетерогенности информационных систем, которые содержат и обрабатывают эти данные:

- аппаратная гетерогенность, или гетерогенность аппаратного обеспечения;
- гетерогенность программного обеспечения, которую можно, в свою очередь, разделить на:
  - интерфейсную гетерогенность;
  - гетерогенность прикладных программ;
  - гетерогенность операционных систем.

Безусловно, такое разнообразие видов гетерогенности создает существенные проблемы интегрирования данных, технологий и информационных систем, что, в свою очередь, негативно влияет на решение комплексных задач устойчивого территориального развития. Особенно остро проблемы несовместимости данных / технологий / систем проявляют себя при необходимости интегрирования пространственной и непространственной информации.

Если речь идет о геопространственных данных, то их гетерогенность проявляется в нескольких измерениях [4]:

- пространство (прежде всего, система координат);
- время;
- масштаб;
- связи (например, топологические);
- непространственные атрибуты.

При реализации процедур интегрирования гетерогенной информации необходимо компенсировать гетерогенность по указанным измерениям, используя различные сервисы преобразования данных. Очевидно, е-Инфраструктура должна реализовывать такие сервисы и обеспечивать пользователям простой доступ к ним.

## 3. Базисы интегрирования геопространственных данных

В подавляющем большинстве задач интегрирования данных первой проблемой, которую необходимо решить, является проблема определения базисов интегрирования, т.е. определенных информационных характеристик или свойств, которые являются общими для всех данных, которые интегрируются.

Согласно Дж. Клиру [5], правильно определенные базисы должны отвечать следующим требованиям:

- базисы должны быть применимы для всех основных свойств системы, для которой они определены;
- базисы системы должны соответствовать ее назначению;
- наблюдения всех свойств системы должны однозначно определяться базисами системы.

Дж. Клиром показано, что решение многих сложных задач с интеграцией данных, систем и технологий требует базиса, который включает в себя такие категории, как пространство, время и группа. Необходимо отметить, что перечисленные базисы не являются единственно возможными, поскольку они не позволяют полностью отобразить перечисленные выше измерения гетерогенности. Однако существующая на сегодня теория и практика использования геопространственных данных / сервисов показывает, что наиболее распространенными базисами интегрирования гетерогенной геопространственной информации являются пространство и время. Это, прежде всего, связано с тем, что для компенсации пространственной и временной гетерогенностей в геоматике имеется хорошо разработанные инструменты, реализованные в распространенных видах программного обеспечения. Инструментарий компенсации других видов гетерогенности (прежде всего, семантической) пока что находится в стадии научных исследований (например, онтологические подходы в геоматике), и их практическая реализация - дело будущего. По этой причине в качестве основных платформ интегрирования разнородных данных в е-Инфраструктуре взяты пространство и время.

### 3.1. Пространство как базис интегрирования

Как известно, главным признаком любого набора геопространственной информации является обязательным наличие в нем атрибутов, с помощью которых эта информация привязывается к определенному пространству (термин "пространство" присутствует даже в самом названии информации такого типа). Поэтому пространство является основным универсальным базисом интегрирования любой

геопространственной информации.

Ключевым вопросом реализации пространственного базиса интегрирования является обеспечение совместимости систем координат, которые используются для пространственной привязки геоданных. При этом основные проблемы заключаются в следующем:

- наличие большого количества систем координат и картографических проекций, которые используются с различными данными;
- необходимость в выполнении координатных преобразований при интегрировании данных, которые сопровождаются возникновением ошибок;
- возникновение неизбежных ошибок во время перехода с эллипсоидальной модели поверхности Земли к плоскости;
- ограниченная точность собственно координатной основы (прежде всего, геодезической сети), которая используется для формирования пространственного базиса.

Таким образом, возникает потребность в формировании универсального трехмерного пространственного базиса, который давал бы возможность однозначного математического описания любой системы координат, в которой привязаны определенные пространственные данные.

Базис, отвечающий данному положению, описан, в частности, в [6-8]. При построении этого базиса использовались следующие принципы:

- в основу положена трехмерная декартова система координат XYZ;
- в качестве единого начала отсчета в декартовой системе координат выбран центр масс Земли;
- плоскость XY совпадает с плоскостью земного экватора;
- количественная оценка погрешностей производится через среднеквадратическое отклонение, а при преобразованиях координат в данные, с которыми производятся манипуляции, не должны привноситься дополнительные погрешности;
- в локальных приложениях пользователи работают с разностями координат, а не с самими координатами.

Основываясь на перечисленных принципах, базовой координатной системой избрана универсальная трехмерная декартова система координат (УТСК) с началом в центре масс Земли, оси X и Y которой расположены в плоскости экватора, ось X направлена в направлении Гринвича, ось Z совпадает с осью вращения Земли, ось Y дополняет систему координат до правой тройки векторов.

На практике используется довольно много различных систем координат. Поэтому весьма важным является установление однозначной связи между этими системами и УТСК, для чего используются известные геодезические зависимости.

Проблема реализации координатных преобразований упрощается при использовании для описания различных систем стандарта ISO 19111 "Пространственная привязка с использованием координат" [9]. Согласно требованиям этого стандарта, координаты пространственного объекта описываются путем определения следующих пакетов (согласно терминологии языка UML):

- системы координат;
- датум;
- привязанная система координат;
- координатные операции.

Такой подход обеспечивает однозначное описание существующих систем координат и единые правила описания новых систем координат, создаваемых пользователем.

Во многих источниках пространственных данных, особенно представленных в символическом виде, пространственная информация выражена неявно, т.е. отсутствуют прямые сведения о координатах определенного объекта, хотя косвенно эта информация присутствует (например, в виде почтового адреса, номера телефона или топонима). Пространственная привязка данных такого рода реализуется через так называемые пространственные идентификаторы, или газетеры.

Под газетиром понимается база данных, используемая для реализации пространственной привязки определенных географических атрибутов (например, названий городов) к их географическим координатам [10].

Практическое применение газетиров связано с рядом проблем:

- совпадения названий различных пространственных объектов;
- наличие неофициальных и временных топонимов, присущих определенным территориям и соответствующих определенному времени;
- некоторые топонимы могут меняться со временем (например, многократные переименование пространственных объектов во время смены политических режимов, изменение границ административно-территориальных единиц и т. д.);
- наличие многих авторских наименований пространственных объектов (особенно в науках о Земле);
- неоднозначная и разнообразная транслитерация топонимов пространственных объектов в различных странах.

Унификация процедур привязки пространственных объектов с помощью газетиров реализуется на базе стандарта ISO 19112:2003: "Geographic information - Spatial referencing by geographic identifiers" (Географическая информация - привязка через географические идентификаторы) [11].

Для реализации пространственного базиса интегрирования данных в е-Инфраструктуре необходимо, как минимум, реализовать сервис координатных преобразований и сервис использования газетиров.

## 3.2. Время как базис интегрирования

Отображение временных изменений в информации о территориальных объектах является очень важной задачей при решении проблем устойчивого территориального развития, поскольку без отражения динамики геопространственных данных нельзя представить себе возможность реализации мониторинговых задач, которые иногда являются определяющими в этой предметной области.

На практике чаще всего можно наблюдать следующие виды пространственно-временных изменений в объектах реального мира [12]:

1. Постоянное изменение пространственных атрибутов объектов во времени (распространение оползней, подтопление территорий и т.д.). Эти изменения всегда относятся к определенному периоду времени;

2. Пространственные атрибуты являются постоянными на протяжении определенного времени, а затем быстро меняются (изменение границ объекта недвижимого имущества). Этим изменениям соответствует определенный момент времени;
3. Постоянное изменение во времени тематических атрибутов пространственных объектов (изменение типа почвы земельного участка);
4. Дискретное изменение во времени тематического атрибута пространственного объекта (смена собственника объекта недвижимости);
5. Дискретные изменения пространственных объектов во времени, в результате чего появляются новые объекты (слияние нескольких объектов недвижимости в один).

Пространственно-временные модели, которые целесообразно реализовывать в е-Инфраструктуре, должны предоставлять возможность отражать особенности таких изменений геопространственных объектов. На практике наиболее популярной в настоящее время в геомашике является версионная пространственно-временная модель, реализованная, например, в базе геоданных ArcGIS [13]. В этой модели устанавливается, что существуют определенные «атомарные» или «инкрементные» изменения в геопространственной базе данных, которые не могут быть разделены на более мелкие части. Механизм управления версиями позволяет восстановить состояние базы геоданных на момент запрошенного инкрементного ее изменения.

Для реализации временного базиса интегрирования данных в е-Инфраструктуре необходимо поддержание сервиса версионности, имеющего функциональность, подобную описанной выше.

## 4. Модель использования гетерогенной пространственной информации

Модель использования гетерогенных геопространственных данных в е-Инфраструктуре строилась с учетом необходимости решения перечисленных выше проблем, а также на основе широко известной модели архитектуры, ориентированной на сервисы (SOA) [14]. Структура этой модели показана на рис. 1.

Для всех гетерогенных геопространственных данных создаются базы метаданных на основе профилей соответствующих стандартов. Кроме того, в е-Инфраструктуре предусмотрена реализация базовых сервисов, которые позволяют выполнять преобразование данных, чтобы компенсировать влияние различных видов гетерогенности, перечисленных выше.

Модель интегрирования работает следующим образом. Пользователи, использующие базы метаданных, библиотеки алгоритмов обработки геопространственных данных, стандарты (профили стандартов) и нормативные документы, юридической и институциональной базы е-Инфраструктуры, а также собственные знания и опыт, в результате интерактивного взаимодействия с модулем определения алгоритма интегрирования выбирают процедуру обработки гетерогенных геопространственных данных, которая лучше всего отвечает характеру конкретной задачи интегрирования и типам используемых данных. В результате модуль определения алгоритма интегрирования передает в модуль интегрирования геопространственных данных выбранный алгоритм, вызываются необходимые для его реализации сервисы и выполняется процедура интегрирования.

В качестве основных (но не единственно возможных) сервисов используются следующие:

- геокодирование;
- преобразование систем координат;
- преобразование форматов геопространственных данных;
- согласование семантики атрибутивных данных;
- согласование топологических отношений между геопространственными данными;
- согласование структуры метаданных;
- генерализации;
- управление версиями для темпоральных данных;
- интегрирование моделей данных.

Указанный перечень сервисов перекрывает большинство процедур интегрирования, которые востребованы на практике. Безусловно, в процессе дальнейшего развития е-Инфраструктуры перечень сервисов должен расширяться, а функциональность самих сервисов совершенствоваться.

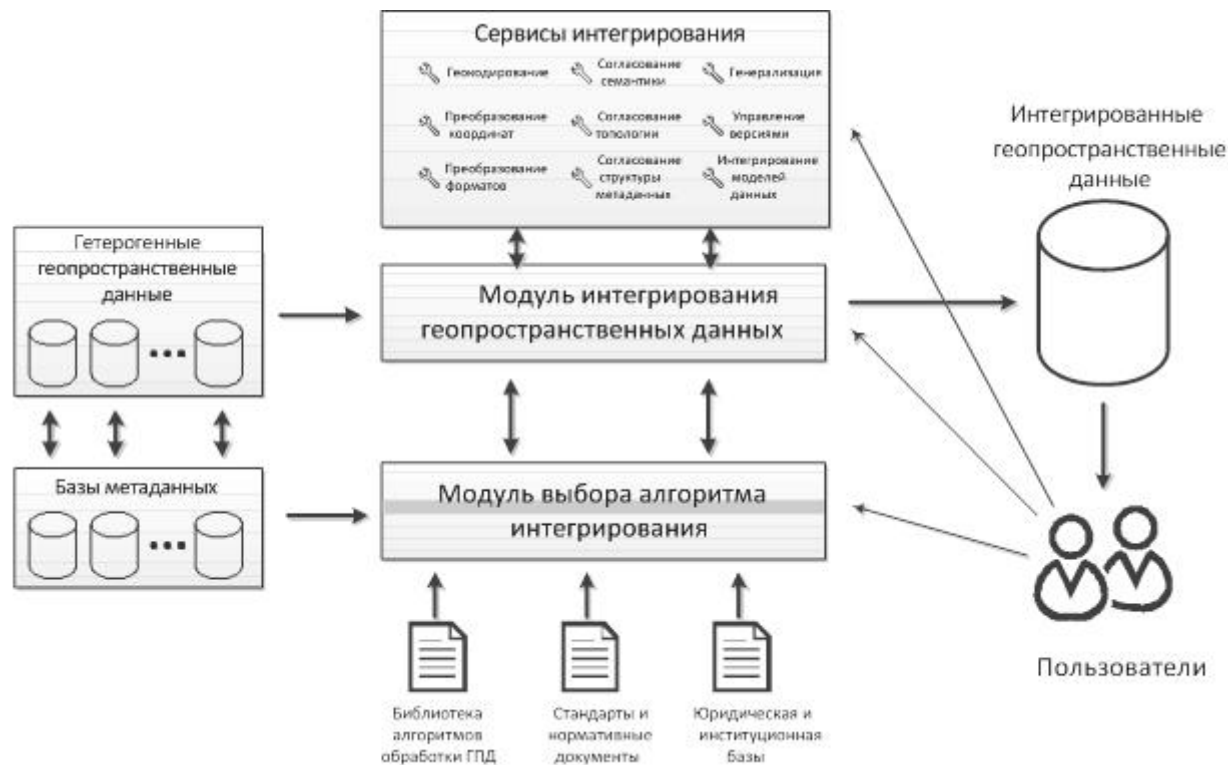


Рис. 1. Модель использования гетерогенной пространственной информации

После завершения выполнения выбранного алгоритма в модуле интегрирования геопространственных данных, интегрированные данные становятся доступными пользователям.

## 5. Выводы

Предложенная модель интегрирования гетерогенной геопространственной информации является реализацией широко известной концепции инфраструктур геопространственных данных, примененной к предметной области дистанционного зондирования Земли и решению проблем устойчивого развития территорий. При этом основной проблемой является реализация в распределенной информационной среде базисов интегрирования. Если пространственный и временной базисы на сегодня можно считать достаточно хорошо разработанными и практически реализованными в программных продуктах, то обеспечение семантической интероперабельности между информационными системами, обрабатывающими геопространственные данные, пока является проблемой. Решение задач интегрирования гетерогенной геопространственной информации в различных предметных областях потребует адаптации предложенной модели как с точки зрения используемых данных, так и с точки зрения функциональности сервисов интегрирования. Однако, на наш взгляд, предложенная в данной работе общая структура модели интегрирования может быть взята за основу при создании подобных моделей в других областях применения пространственных данных.

## 6. Благодарности

Исследование, описанное в данной статье, проведено в рамках выполнения совместного Российско-Украинского проекта «Российско-Украинский сегмент глобальной е-Инфраструктуры постоянного доступа к научным ресурсам ДЗЗ для решения задач устойчивого развития территорий» при финансовой поддержке Российского и Украинского фондов фундаментальных исследований.

## Список литературы

1. Yeung A.K.W., Hall G.B. Spatial database systems: Design, implementation and project management. Dordrecht: Springer, 2007.
2. The Canadian Geospatial Data Infrastructure. Achieving the vision CGDI. - URL: [http://geoconnections.org/publications/tvip/Roadmap\\_E/CGDI\\_Roadmap\\_final\\_E.html](http://geoconnections.org/publications/tvip/Roadmap_E/CGDI_Roadmap_final_E.html).
3. OGC® Standards and Specifications.- URL: <http://www.opengeospatial.org/standards>.
4. Shi W.Z. Spatial data quality and uncertainty: ASPRS, 2009.
5. Клар Дж. Системология. Автоматизация решения системных задач: Пер. с англ. - М.: Радио и связь, 1990. - 544с.
6. Mikhail E. Observations and least squares. New York: Harper & Row, 1976.
7. Leick A. GPS satellite surveying, 3rd ed. New York: John Wiley, 2004.
8. Wolf P.R., Ghilani C.D. Adjustment computations: Statistics and least squares in surveying and GIS. New York: John Wiley, 1997.
9. ISO 19111:2007: Geographic information - Spatial referencing by coordinates.- URL: <http://www.iso.org>.
10. Gazetteer Service - Application Profile of the Web Feature Service Implementation Specification.- URL: <http://www.opengeospatial.org>.
11. ISO 19112:2003: Geographic information -- Spatial referencing by geographic identifiers.- URL: <http://www.iso.org>.

12. Peiquan J., Yue L., Gong Y. Design and implementation of a unified spatio-temporal data model // Advances in Spatio-Temporal Analysis. London: Taylor & Francis, 2008.

13. Understanding versioning. ArcGIS Desktop Help.- URL: <http://webhelp.esri.com>.

14. Geospatial Service-Oriented Architecture (SOA).- URL: <http://www.esri.com/library/whitepapers/pdfs/geospatial-soa.pdf>.

## **Об авторах**

*1. Кудашев Ефим Борисович - член-корреспондент Российской Инженерной Академии, д.т.н., ведущий научный сотрудник Института космических исследований Российской академии наук, профессор Механико-Математического Факультета МГУ им. М.В. Ломоносова, Москва, Российская Федерация.*

*2. Марков Сергей Юрьевич - к.т.н., доцент, старший научный сотрудник Научного Центра аэрокосмических исследований Земли Института геологических наук Национальной академии наук Украины, Киев, Украина.*

*3. Попов Михаил Алексеевич - д.т.н., профессор, заместитель директора по научной работе Научного Центра аэрокосмических исследований Земли Института геологических наук Национальной академии наук Украины, Киев, Украина.*

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (гранты № 11-07-90404 и 11-07-00006) и проекта APARSEN Европейской Комиссии "Alliance Permanent Access to the Records of Science in Europe Network".