

УДК 004.6 (004.62)

С. К. Дулин^{1, 3}, д-р техн. наук, проф., гл. науч. сотр., e-mail: s.dulin@ccas.ru,

Н. Г. Дулина², канд. техн. наук, ст. науч. сотр., e-mail: ngdulina@mail.ru,

Д. А. Никишин³, канд. техн. наук, зав. сектором, e-mail: dmnikishin@mail.ru

¹ Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте (ОАО НИИАС) РАН

² Вычислительный центр им. А. А. Дородницына РАН

³ Институт проблем информатики РАН

Особенности моделей геоданных и методов их обработки в аспекте обеспечения семантической геоинтероперабельности*

Ключевой задачей обеспечения семантической геоинтероперабельности является создание единой концептуальной модели представления геоданных на основе интеграции пространственно-распределенной информации. Такая модель должна строиться на основе существующих метасхем баз геоданных с учетом многофакторности взаимодействия пользователей и семантики, заложенной в пространственные онтологии и/или геотезаурусы и классификаторы. Работа посвящена исследованию и анализу существующих моделей и методов представления геоданных и имеет целью оценку релевантности существующих структур геоданных методам и подходам обеспечения их семантической геоинтероперабельности при решении задач концептуального поиска геоданных. Для анализа были выбраны те структуры геоданных, которые совмещают в себе средства отображения всех трех групп пространственно-распределенной информации: семантической, метрической и топологической. Рассмотрены различные уровни представления геоинформационного контента, которые требуют принципиально различных методов обработки.

Ключевые слова: семантическая геоинтероперабельность, моделирование геоданных, онтологии, концептуальный поиск

Введение

За последнее десятилетие в области геоинформатики произошла значительная эволюция способов организации информационных процессов. Можно выделить следующие модели организации информационных процессов:

- автономная модель, объединяющая в себе как прикладные интерфейсы пользователя, так и функции хранения данных в едином аппаратно-программном комплексе, т. е. "все в одном". Эта модель является наиболее очевидным подходом, существовавшим достаточно долго в виде так называемых настольных геоинформационных систем;
- модель клиент-сервер (Client/Server, C/S) — данные хранятся централизованно в виде выделенного ресурса, к которому посредством стандартных запросов обращаются многочисленные клиенты-обработчики данных, т. е. представляет

собой "звездообразную" структуру, реализующую отношения "один ко многим";

- облачная, или сетевая, модель (GRID) — имеется множество серверов, предназначенных как для хранения данных, так и для выполнения некоторых базовых процедур, а также имеется множество клиентов, которые решают в этой среде свои прикладные задачи. То есть такая модель представляет собой сетевую структуру, реализующую отношения "многие ко многим". Важным компонентом такой системы должна являться посредническая структура, которая сообщает клиентам, где искать нужные им данные.

В настоящее время произошел практически полный переход от автономных, "настольных" комплексов в сторону распределенных геоинформационных систем [1]. В связи с этим немаловажным фактором работы с геоданными, ввиду их значительного объема, становится скорость их передачи по каналам связи в сети Интернет или в корпоративных сетях.

Вместе с тем, ежедневно в отрасли геоинформационных технологий проявляется и используется

* Работа выполнена при поддержке РФФИ (проекты 14-07-00040 и 14-07-00785).

все большее число географических данных. Высокая стоимость как создания и эксплуатации массивов геоданных, так и сбора собственно геоданных и их своевременного обновления, сопровождаемого необходимостью их согласования с уже существующими данными, обуславливает необходимость обеспечения возможности совместного и многократного использования геоинформационных ресурсов различными заинтересованными пользователями.

Обобщая сказанное, можно отметить, что геоинформационный контент в современных реалиях зачастую представляет собой сложную распределенную систему, включающую в себя совокупность множества источников геоданных и прикладных геоинформационных сервисов (в том числе облачных), создаваемых и эксплуатируемых различными организациями-операторами, каждая из которых, в свою очередь, может иметь собственные представления о способах организации геоданных и правилах их предоставления потребителям, использовать различные устройства для сбора геоданных.

Для обеспечения как самой возможности совместной работы, так и ее эффективности географическая информация должна быть рационально организована таким образом, чтобы любой пользователь, независимо от специфики решаемых задач или используемого им аппаратно-программного обеспечения, мог обратиться к тем или иным геоданным через эффективный и удобный метод запроса, чтобы быстро получить точную и исчерпывающую информацию по интересующему его вопросу [2].

При этом должна обеспечиваться возможность предоставления геоданных с различной степенью обобщения (генерализации) как в пространственном, так и в семантическом аспектах.

Пространственный аспект предполагает возможность представления геоданных (в частности, пространственного положения геообъектов) с различной степенью точности в пространствах различной размерности (2D, 3D), различных системах координат и проекциях.

Семантический аспект может предусматривать отбор и обобщение данных в различных смысловых категориях.

Все это обуславливает проблемы совместимости и сопоставимости геоданных различного происхождения.

Под *совместимостью* будем понимать принципиальную возможность совместного использования данных различного происхождения в том или ином информационном процессе.

Под *сопоставимостью* будем понимать соответствие данных из различных источников определенным требованиям, в частности, по точности, полноте, достоверности, актуальности и т. п. [3].

Проблема совместимости геоданных, или их интероперабельности, в последние годы является одной из актуальных тем исследований в области геоин-

формационных технологий [4, 5]. В некоторых работах [2] используется термин "функциональная совместимость", близкий по смыслу приведенному выше определению интероперабельности. В дальнейшем изложении будем пользоваться термином "интероперабельность" как более удобным.

Основополагающими предпосылками исследований в области обеспечения геоинтероперабельности являются [1]:

- переносимость структур геоданных в различные предметные области;
- использование открытого программного обеспечения;
- обеспечение возможности обогащения структур геоданных дополнительными семантическими характеристиками;
- обеспечение возможности автоматизированного формулирования логических выводов между семантическими описаниями;
- использование общепринятых и распространенных стандартов для обеспечения функциональной совместимости.

Актуальность проблемы обеспечения геоинтероперабельности проявляется наиболее наглядно в тех ситуациях, когда необходимо обеспечить взаимодействие геопорталов, использующих сервисы и приложения, в которых реализованы разные модели представления геоданных или синтаксис их описания и которые различаются методическими и функциональными подходами к обработке данных.

Сложность решения этой проблемы существенно возрастает, если модель представления данных или синтаксис ее описания имеет динамический аспект, т. е. изменяется во времени. В этом случае говорят об обеспечении свойства динамической геоинтероперабельности [6].

При ближайшем рассмотрении проблема интероперабельности распадается на ряд частных проблем как в плане представления геоданных в информационном процессе (протоколов передачи геоданных, моделей геоданных или форматов файлов и т. п.), так и в плане составляющих модели геоданных (пространственные, семантические, а также их топологические расширения).

Целью данной работы является исследование, в первую очередь, проблемы семантической интероперабельности геоданных относительно способов их организации. С этой целью в работе рассмотрен ряд традиционных моделей геоданных, на примере которых анализируются проблемы интероперабельности геоданных вообще и их семантической интероперабельности в частности.

Модели, методы и технологические схемы представления геоданных

Пространственные данные устроены гораздо сложнее и разнообразнее традиционных видов деловой информации, поэтому проектирование храни-

лишь пространственных данных требует специальных компетенций из области географии, геодезии, картографии, наук о Земле и других областей естествознания.

Модели пространственной информации обычно объединяются в две большие группы: полевые и объектные [7]. Применительно к геоданным эти группы образуют соответственно классы "растровых" геоизображений и "векторных" моделей местности.

Полевая модель преобладает в сфере дистанционного сбора первичных данных о Земле. Получаемые в результате аэро- или космической съемки материалы представляют собой в основном растровые геоизображения; исключением можно считать разве что данные лазерного сканирования, представляемые в виде диффузного облака точек (MTD-модель).

Представление геопространства как непрерывного поля тех или иных свойств является наиболее естественным и простым, но в то же время такой подход затрудняет некоторые виды анализа и преобразования этих данных.

Для расширения возможностей анализа полевой модели и повышения эффективности ее обработки могут применяться квадротомическое, трихотомическое, гексатомическое, октотомическое представления, основанные на разделении пространства поля на отдельные участки/блоки, организованные в виде иерархической структуры [8].

Форматы растрового представления данных более или менее сопоставимы друг с другом, что обусловлено простотой растровой модели данных. Их многообразие объясняется, в значительной мере, различиями в способах описания метаданных, сопровождающих такое изображение¹.

Объектная модель предполагает выделение (декомпозицию) в геопространстве отдельных объектов — структур, обладающих схожими комплексами качеств (классами); формальное описание качеств, их совокупность и образуемые ими структуры определяют модель данных того или иного класса объекта.

Такой подход предполагает наличие некоторой систематизированной совокупности моделей объектов различного рода, эта совокупность обычно называется концептуальной схемой, или онтологией. Классы объектов могут существовать независимо друг от друга или входить в структуру более сложных объектов (например, иерархическую). Существует множество различных онтологий, относящихся к различным областям знания, например, топографии, геологии, геоморфологии, инженерным задачам. При этом в каждой из областей могут использоваться различные альтернативные онтологии,

¹ Например, из почти 100 растровых форматов, поддерживаемых в ArcGIS, многие являются специфическими форматами для представления данных, получаемых различными спутниковыми съемочными системами.

имеющие свою область применения, специфику, степень конвенциональности и т. п.

Геообъект как объектная модель географических данных предполагает наличие двух категорий атрибутов: пространственных (координатных) и непространственных (смысловое описание объекта, представленное либо в текстовом виде, либо в виде его цифрового кода). Например, еловый лес в пространственном отношении определяется полигональным контуром своих границ, а непространственным атрибутом является характеристика древостоя со значением "ель".

Способы представления числовых или текстовых полей данных, а также состав и структура геообъекта в целом определяются задачами, которые будут решаться на его основе. В связи с этим среди векторных форматов геоданных наблюдается значительно большее разнообразие по сравнению с растровыми форматами.

Объектная модель существенно расширяет возможности анализа. Характерными прикладными применениями объектной модели являются задачи кадастрового учета земельных участков или построения графов транспортных сетей (например, дорог).

В контексте объектных моделей можно выделить еще один аспект — топологический, определяющий взаимосвязи между несколькими геообъектами. В данном аспекте также можно выделить пространственную и непространственную составляющие. Первая определяет взаимное пространственное расположение объектов, например, предыдущий/следующий участок дороги в маршруте как сегменты дорожного графа. Вторая отвечает за смысловую, семантическую взаимосвязь объектов, например, объединение всех лесных массивов в общий слой "леса". Топологические зависимости по своей природе могут быть пространственными, временными, функциональными, структурными и т. п.

Задача обеспечения преемственности геообъектов в пространственно-временном континиуме обусловлена тем, что геообъект в общем случае имеет ограниченный по времени период своего существования.

Один из примеров: некоторый участок коренного леса был вырублен — превратился в район вырубленного леса, затем распахан — превратился в сельхозугодие, которое, в свою очередь, было заброшено, на его месте появился кустарник или молодая поросль, которые впоследствии превратились в новый лес (но уже не коренной). В этом случае геообъект как определенная территория (скажем, землевладение) не менял своей пространственной локализации во времени, однако происходило изменение его сущности, т. е. семантической составляющей.

Другой пример: в определенный момент шахта была построена и введена в эксплуатацию, впоследствии была закрыта и, наконец, разрушилась и прекратила свое существование как геообъект.

Однако в результате проседания грунта на месте шахты образовалось озеро, т. е. возник другой геообъект. Таким образом, некоторая область местности (не определенная по своим границам) в различные периоды времени принадлежит различным геообъектам: различным как в семантическом плане, так и по своей пространственной конфигурации. При этом между этими объектами существует причинно-следственная связь, которая может быть отражена в базе геоданных.

Еще одной задачей топологии может быть обеспечение полимасштабного представления геообъектов, заключающееся в определении взаимосвязей между серией контуров, описывающих пространственную локализацию объекта с различной степенью детализации и предназначенных для использования в том или ином масштабе.

Любые данные, независимо от своей природы, должны в том или ином виде сопровождаться информацией, которая описывает сущность и свойства данных, хранящихся в базе данных или файлах, — метаданными. При отсутствии такой информации данные оказываются бессмысленными и, соответственно, бесполезными. Вместе с тем, метаданные — описание свойств конкретного информационного объекта, которое может отчуждаться от этого объекта [5].

Среди метаданных можно выделить две составляющие:

- метаданные, определяющие структуру модели данных и способы представления значений данных;
- метаданные, отражающие свойства предметной области, моделируемой данными.

Согласно стандарту ГОСТ Р 51353—99 [9] метаданные позволяют описывать содержание, объем, положение в пространстве, качество (точность, полноту, достоверность и современность) и другие характеристики.

Стандарт ISO 19115:2003 — Geographic information — Metadata определяет методологию формирования метаданных географической информации и содержит концептуальную модель метаданных [10]. Стандарт ISO 19139:2007, в свою очередь, определяет правила кодирования этой модели и ее реализацию в виде XML-схем [11]. Стандарт ISO 19115-2:2009 содержит расширения ISO 19115:2003 для описания сеточных данных, а также изображений, полученных от камер, инфракрасных и мультиспектральных сканеров, радаров и подобных устройств [12].

Согласно стандарту ISO 19115:2003 имеют место следующие определения:

- *метаданные* — это данные о данных;
- *элемент метаданных* — это дискретная единица метаданных, элемент метаданных должен быть уникален в пределах сущности метаданных, элемент эквивалентен понятию "атрибут" в терминологии UML;

- *сущность метаданных* — это набор элементов метаданных, описывающих один и тот же аспект данных; может включать в себя одну или несколько других сущностей метаданных, сущность эквивалентна понятию "класс" в терминологии UML;
- *раздел метаданных* — это подмножество метаданных, которое состоит из связанных (родственных) сущностей и элементов метаданных (это понятие эквивалентно понятию "пакет" в терминологии UML).

Другими словами, разделы метаданных представлены в виде UML-пакетов. Каждый пакет содержит одну или более сущностей (UML-классов). Сущности содержат элементы (UML-атрибуты и UML-ассоциации). Сущности могут являться дочерними или родительскими для других сущностей. На рис. 1 в качестве иллюстрации представлена диаграмма пакетов метаданных с отношениями между ними [13].

Стандарт ISO 19115:2003 также определяет ядро метаданных для географических данных — минимальное подмножество элементов метаданных, которое следует использовать при описании набора данных.

Элементы ядра метаданных требуются для идентификации набора данных и позволяют документировать следующие характеристики набора данных: тематику, формат, язык описания, местоположение, временной период, к которому относятся данные, контактную информацию лица, ответственного за данные.

Лишь часть элементов ядра является обязательной для заполнения, однако стандарт ISO 19115:2003 рекомендует использовать все элементы для достижения большей совместимости метаданных и для более полного понимания пользователями природы и содержания набора данных.

Стандарт ISO 19115:2003 определяет и другие структурные элементы модели данных:

- *модель* — это абстракция некоторых аспектов предметной области;
- *набор данных* — идентифицируемая совокупность данных, которая в частном случае может состоять из одного элемента;
- *тип данных* — совокупность характеристик, наличие которых у группы данных позволяет выделить эту группу из множества других данных. Примеры типов данных: Integer, Real, Boolean, String, Date, GMPoint;
- *структурированные данные* — данные, разделенные на элементы с однозначно трактуемой семантикой в соответствии с заранее известными синтаксическими правилами (соглашениями) [5];
- *схема данных* — стандартизованное машиночитаемое описание типового способа представления структурированных данных об определенной разновидности объектов (сущностей) или сово-

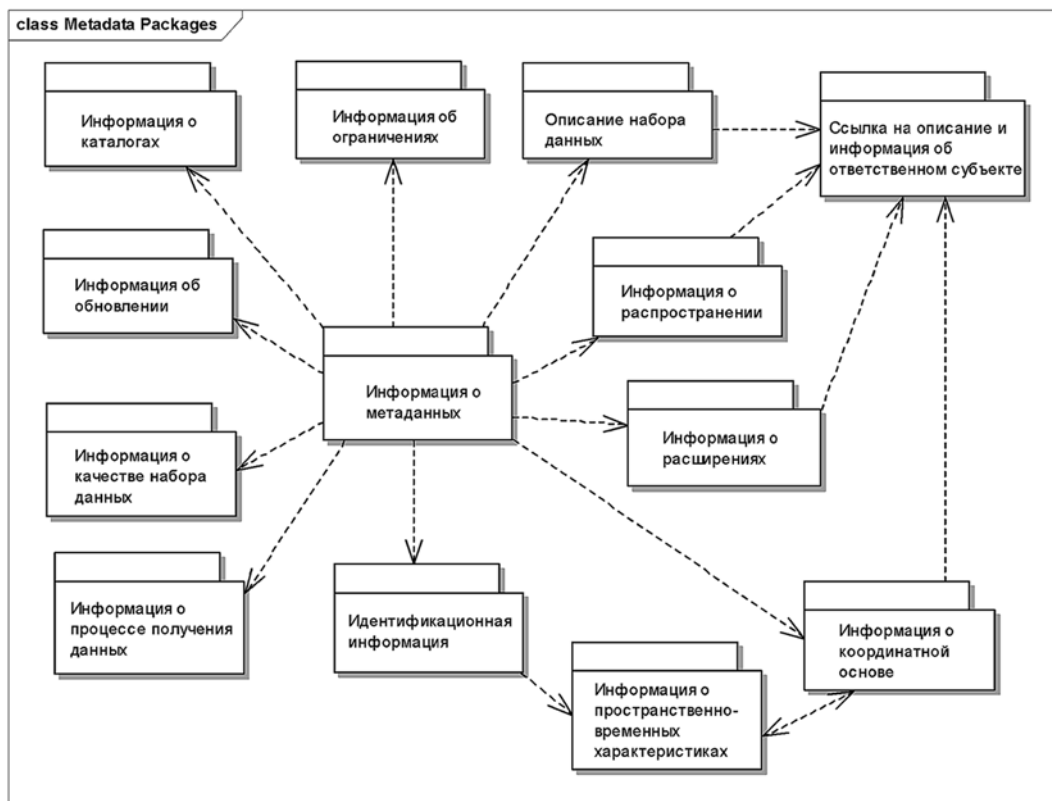


Рис. 1. Диаграмма пакетов метаданных

купности объектов и отношений между ними в определенной предметной области. Схема данных устанавливает формализованные требования к взаимному расположению и иерархии элементов данных, их кодировке, обозначениям, допустимым типам и значениям;

- *формат массивов данных* (файлов или БД) — набор синтаксических, структурных и логических правил кодирования и упорядочивания (организации) данных, обеспечивающих их корректное и эффективное восстановление в виде структуры, соответствующей определенной модели данных.

Формат данных может быть оптимизирован как под способы решения определенных задач (например, хранение "пирамиды" для визуализации изображения, использование особых моделей цвета типа СМΥΚ, разбиение изображения на блоки/полосы), так и под особенности обработки данных на различных аппаратно-программных платформах (например, LE/BE-последовательность байтов, выравнивание на границу слова и т. п.).

В форматах, предназначенных для архивного хранения или для передачи данных, могут быть заложены возможности выявления ошибок и восстановления данных в случае их повреждения, обычно реализуемые за счет избыточности.

Еще одной задачей, решаемой посредством так называемых обменных форматов, является обеспе-

чение возможности более или менее корректного и полного преобразования одной модели данных в другую, например, между различными приложениями — т. е. обеспечение интероперабельности.

Особенности пространственного описания геоданных

Пространственное моделирование включает в себя два функциональных аспекта:

- аналитический — модель данных, описывающая пространственное положение и форму геообъекта в функциях пространственного анализа данных;
- графический — модель данных, обеспечивающая частную, но совершенно необходимую задачу отображения пространственной информации (на экране/принтере).

Как аналитические, так и графические построения в "векторной" графике основаны на аппроксимации контуров геообъектов между пространственно определенными узловыми точками (вершинами) с помощью тех или иных функциональных зависимостей, для этого обычно используются линейные функции или полиномы (полиномиальные функции, сплайны).

Число и взаимное расположение узловых точек определяются требованиями к точности модели, т. е. к допустимому расхождению между траекторией аппроксимирующей функции и положением ре-

альных контуров. Различные требования к точности контуров на реальном примере двух карт различного масштаба [14] иллюстрирует рис. 2 (см. третью сторону обложки).

Полилинейная аппроксимация является наиболее распространенным способом, что обусловлено ее единообразным и простым с точки зрения пространственного и топологического анализа представлением метрики геоинформационного контента в виде линейных сегментов. Недостатком этого метода является более значительный объем данных за счет большего числа узловых точек.

Полиномиальная (сплайновая) аппроксимация является методом, более емким в информационном отношении, но более сложным с точки зрения пространственного и топологического анализа. Например, здесь становится неоднозначным распространенный способ выявления смежности объектов по совпадению координат узловых точек соответствующих сегментов их общей границы.

В векторном моделировании пространственных объектов можно выделить два уровня: примитивы и собственно объекты, состоящие из примитивов. Для сложных или составных объектов роль примитивов могут играть более простые объекты.

Можно выделить следующие топологические классы пространственных моделей объектов:

- точечная, морфологическая локализация которой не имеет смысла ввиду малоразмерности объекта;
- ориентированная — по сути это точечная модель, для которой определена ориентация вдоль определенного направления;
- линейная (в том числе замкнутая), образуемая осевой линией объекта;
- площадная, образуемая граничной линией (контуром) объекта и описывающая разделяемые им внутреннюю и наружную области геопространства.

Моделирование поверхностей является более сложной задачей.

В векторном виде модель рельефа описывается как топологическая сеть на плоскости посредством различных способов:

- изолиниями: изогипсами (горизонтали) или изобатами. Изолинию можно представить как область пространства с высотами не ниже/выше высоты данной горизонтали. В частности, горизонталь образует область, снаружи которой высоты не могут превышать значение высоты горизонтали, а внутри, соответственно, быть ниже. Таким образом, возникает топологическая зависимость, накладывающая ограничения на соседство горизонталей с несоответствующими высотами. В практике картографии такая зависимость отражается так называемым бергштрихом, указывающим направление ската;
- структурными линиями (в этом числе ориграфические). Аналогично изолинии структурную

линию можно представить как область с определенным градиентом уклона местности;

- TIN (Triangle Irregular Model) — моделью рельефа, основанной на принципе аппроксимации поверхности посредством нерегулярной триангуляционной сети. Ее структурными элементами являются высотные точки (пикеты), между которыми назначаются треугольники, аппроксимируемые плоскостями. В триангуляционной модели качество аппроксимации зависит от плотности высотных точек (узлов триангуляционной сети) на определенном участке модели. На практике TIN в той или иной степени опирается на сеть структурных линий. Модель TIN допускает ее непосредственную визуализацию как поверхности (рендеринг). Существуют различные методы построения триангуляции, различающиеся способами представления и поиска необходимых узлов, ребер и треугольников [15].

Все три перечисленные выше способы теоретически инвариантны друг другу. Для передачи немасштабных особенностей рельефа используются линейные и точечные объекты.

Для моделирования поверхностей используются также полевые (регулярные) модели. Для моделирования рельефа как функции высоты от планового положения точки используется регулярная модель высот (Digital Elevation Model, DEM), которая в русскоязычной литературе соответствует понятию матрицы высот. Для моделирования непространственных свойств местности используются так называемые матрицы качеств. По своим основным свойствам они аналогичны другим растровым изображениям.

Основным недостатком регулярных моделей является громоздкость данных, здесь остро стоит проблема выбора между степенью детализации такой модели (размером ячейки) и объемом данных.

Подводя итоги, можно сказать, что основными препятствиями к обеспечению интероперабельности пространственных данных являются:

- преобразования в пространства более низкой размерности — обратное преобразование становится неполноценным;
- преобразования, связанные с понижением (загрублением) точности представления координат или с прореживанием узловых точек;
- нарушения, связанные с неполным соответствием исходной и конечной пространственных моделей. Например, если в одной модели порядок цифрования объекта не определен, то его перенос на модель, которая использует такой признак, чреват непредсказуемыми последствиями (например, озеро может превратиться в остров и т. п.) [16].

Что касается растровых изображений, то любое их преобразование, связанное с перепроектированием на новую сетку разбиения (геометрическое

трансформирование), связано с безвозвратными потерями исходной информации и, соответственно, снижением его информационных свойств.

Основные аспекты интероперабельности

Геоинтероперабельность (англ. — Geointeroperability) — это способность к взаимодействию в геопространстве в широком смысле этого слова.

При более подробном рассмотрении проблемы интероперабельности можно выделить несколько ее аспектов, которые упоминаются в работах [4, 2]:

1. Свойство интероперабельности применительно к геоданным означает, что при решении задач обеспечивается возможность совместного использования геоданных из различных источников (баз геоданных) независимо от их происхождения, моделей для описания семантической и пространственной составляющих данных, синтаксических особенностей представления этих моделей и т. п. Решение проблемы интероперабельности данных реализуемо, например, посредством автоматизированного преобразования данных на основе обменных форматов геоданных или стандартизированных запросов к базам пространственных данных. Процесс преобразования должен быть прозрачным пользователям.

2. Аналогично можно определить функциональную интероперабельность программных модулей как возможность использования тех или иных стандартных методов и их программных реализаций (процедур/функций) для решения частных задач, возникающих в ходе решения основной задачи. Альтернативные процедуры, например, могут различаться методом получения одного и того же результата, способом оптимизации вычислительного процесса под конкретную аппаратно-программную платформу или другими особенностями. Такая возможность обеспечивается использованием парадигмы процедурной абстракции, основанной на стандартизации интерфейсов системных вызовов функций и программных модулей или стандартов удаленного вызова процедур. Необходимым условием для функциональной интероперабельности является обеспечение интероперабельности данных.

3. Наконец, можно определить интероперабельность геоинформационных систем (в частности, геопорталов) как возможность для решения той или иной прикладной геоинформационной задачи в сложной распределенной среде, с использованием как различных источников геоданных, так и различных инструментальных средств (программных реализаций методов обработки данных). Это дает пользователям свободу выбора платформы географической информационной системы в гетерогенной вычислительной среде.

Таким образом, решение проблемы интероперабельности в геоинформационных технологиях сосредотачивается на обеспечении обмена данными между геоинформационными системами с различ-

ной архитектурой и обеспечении возможности совместного использования как данных, так и функциональных методов их обработки.

В работе [2] также упоминаются два подхода к обеспечению интероперабельности, основанные на стандартизации спецификаций:

- способов представления и кодирования геоданных (стандартных форматов);
- интерфейсов методов доступа к геоданным и их функциональной обработки, главным образом через приложения, с помощью программных интерфейсов (API). В этом случае проще организовать сетевое взаимодействие, но, вместе с тем, у прямого доступа к глобальным сетевым данным могут быть некоторые проблемы, такие как безопасность, полные узкие места, юридические ограничения.

Принятие стандартов — важное средство для формирования и развития взаимодействующих геопространственных данных и служб. Деятельность на рынке геоинформационных услуг должна основываться на немногих согласованных и хорошо продуманных стандартах и действиях. В настоящее время многие международные организации работают над проблемой стандартизации, например, Консорциум открытой геопространственной информации (Open Geospatial Consortium, OGC) и Технический комитет международной организации по стандартизации 211 (ISO's Technical Committee 211. Geomatics, Geographic Information) [4]. Например, OGC имеет концепцию OGIS, которая включает в себя единую открытую модель геоданных, модель функционального взаимодействия и информационную модель группы [2]. В спецификациях OGC функциональная совместимость осуществлена применительно к C/S (клиент-серверной) модели.

Для успешного взаимодействия различных геоинформационных систем должны обеспечиваться [5]:

- аппаратная совместимость — соответствие размеров, уровней электрических сигналов, способов физической записи информации на машинный носитель и т. п.;
- коммуникационная совместимость — единство используемых протоколов, форматов служебных сообщений и иных средств, используемых для организации передачи данных и команд по каналам связи;
- совместимость моделей/форматов данных.

Оставляя за рамками вопросы аппаратной и коммуникационной совместимости, сосредоточимся на вопросах совместимости моделей данных.

При оценке функциональной совместимости тех или иных данных следует учитывать как совместимость моделей данных вообще, так и соответствие свойств того или иного источника² геоданных условиям решаемой задачи.

² А возможно, и свойств конкретного информационного объекта, например, набора данных/файла.

В плане совместимости моделей данных необходимо учитывать следующие аспекты:

- характер используемой модели данных (полевая/объектная), размерность модельного пространства и т. п.;
- способ сегментации геопространства на геообъекты, определяемый используемой объектной моделью предметной области (например, в разных моделях могут использоваться объекты типа "кварталы" и "улицы", их инвариантность будет определяться условиями конкретной задачи);
- структура модели геоданных, особенности топологического связывания элементов геопространства (в частности, моделей геообъектов), различия в способах описания модели (например, полилинейная или сплайновая модель пространственных данных);
- способы представления числовых и текстовых полей данных, определяемые требуемой конечной степенью детализации пространственной и семантической составляющих модели геоданных: точностью координат (или габаритными размерами моделируемого пространства — extent), а также используемыми системами семантических классификаций и способами их кодирования;
- гарантированная номенклатура и способы представления метаданных;
- синтаксические особенности массивов данных (определяемые форматом, например, кодировка данных, об этом речь будет идти ниже).

В плане оценки возможности использования для решения задачи того или иного источника геоданных (даже при условии совместимости моделей данных) необходимо учитывать:

- происхождение данных, ограничения на их использование и т. п.;
- конкретные особенности описания метрики или семантики, например, система координат, способ представления координат (в частности, в градусной/ радианной мере и т. п.);
- точность конкретного источника данных — даже использующие одинаковый тип данных (например, double) данные, поступающие из различных источников и с различных устройств измерения, могут существенно различаться действительной точностью координат. Это обстоятельство определяет пригодность данных для решения задачи. Вместе с тем, при решении задач в глобальном масштабе необходимо обеспечить закругление высокоточных координат до требуемого уровня (например, координат пунктов государственной геодезической сети при нанесении их на карту). Аналогичный подход применим и в практике засекречивания истинных координат важных объектов.

Различные приложения также могут предъявлять различные требования к степени детализации (разрешению) геоданных.

Семантические и синтаксические аспекты интероперабельности

Одной из ключевых задач, связанных с обработкой данных вообще и геоданных в частности, является одинаковая и однозначная трактовка их сути, т. е. семантики.

В том случае, когда данные интерпретируются людьми, подобная проблема стоит не так остро, при ручном вводе оператор самостоятельно "разбирает" данные документа и без строгой формализации может определить, что является датой, что инициалами, а что, например, исходящим номером. Оператор может даже выявить и исправить ошибки в структуре и содержании заполненных бланков и форм [5].

В случае, когда данными обмениваются электронные автоматизированные системы (например, клиентское приложение и сервер геоданных или два геопортала между собой), однозначная интерпретация передаваемых данных становится уже необходимостью.

Синтаксис данных, как правило, не зависит от их смыслового содержания и диктуется только формальными правилами. И наоборот, семантически идентичные данные могут быть выражены с помощью различных синтаксисов.

Например, в текстовом виде может быть представлено как текстовое описание местности, так и "векторная" карта местности (в виде текстовой структуры, содержащей координаты узловых точек геообъектов, например, форматы TXF, SHP, KML и т. п.). Несмотря на принципиальное различие этих данных с точки зрения семантики их синтаксис будет одинаков — они используют одинаковые кодировку и способ разбиения на строки (различия, относящиеся к способам структурирования данных, в данном контексте не рассматриваем). Но на более высоком синтаксическом уровне возникают различия между упомянутыми выше форматами: в одних используется обычное представление текста, разделенного табуляторами и строками, в других используется более сложный синтаксис, например KML, основанный на синтаксисе XML.

Вместе с тем, "векторная" карта местности может быть представлена и в бинарном формате (например, SXF). Хотя синтаксис текстовых и бинарных файлов несопоставим, их семантика (смысловое наполнение данных) отличаться не будет, примером этого является дуплет TXF/SXF, используемый в продуктах КБ "Панорама" [17].

Наконец, карта местности может быть представлена в растровом графическом формате (например, TIFF), который несопоставим как с векторными бинарными данными (по семантике), так и с текстовыми форматами данных (как по семантике, так и по синтаксису). Но, вместе с тем, все эти данные объединяет общая сущность — все они являются графическим изображением некоторой местности.

Бинарный вариант файла занимает меньше места, чем текстовый, чтение/запись таких данных в разы быстрее. В свою очередь, текстовый вариант переносим на аппаратно-программную платформу с другим представлением чисел, при этом числа могут иметь произвольное число знаков.

Итого, на основании вышесказанного можно выделить следующие аспекты совместимости данных [5]:

- синтаксическая совместимость — использование одинаковых правил кодирования и разметки данных. Например, символы латинского алфавита в большинстве компьютерных систем задаются одними и теми же битовыми последовательностями в соответствии с кодовой таблицей ASCII, все служебные элементы (теги) веб-страницы в формате HTML заключены в угловые скобки, что позволяет отличить их от основного текста;
- семантическая совместимость — одинаковая смысловая трактовка элементов данных. Например, символ "М" в определенной позиции документа может означать и мужской пол, и единицу измерения длины (метр), и инициал имени.

Согласно [5] имеют место следующие определения:

- синтаксическая совместимость — способность информационных систем к взаимодействию благодаря единству представления и структуры данных;
- семантическая совместимость — способность информационных систем к взаимодействию благодаря единству интерпретации и классификации данных.

Семантическая интероперабельность геоданных заключается в обеспечении их пользователей средствами согласованного понимания семантической составляющей геоданных, полученных в результате обмена, а также в выявлении и исправлении тех ситуаций, когда имеет место несогласованное понимание [6].

Семантическая геоинтероперабельность принципиально отличается от синтаксической, так как в последнем случае речь идет только об обеспечении обмена и совместного использования геоданных потребителями на нескольких уровнях взаимодействия: сетевых протоколах передачи данных, вызовах удаленных процедур, унифицированных запросах к базам геоданных. Обеспечение синтаксической геоинтероперабельности является необходимым условием решения проблемы обеспечения семантической геоинтероперабельности.

5. Схемы преобразования (конвертации) форматов данных

Использование в процессе работы геоданных из ряда источников, использующих различные подходы к представлению данных и, соответственно, форматы данных, предполагает их автоматическое

или автоматизированное преобразование. Поэтому задача обеспечения межсистемного обмена данными предполагает существование определенной схемы преобразования форматов данных между собой.

В работе [2] упоминались три схемы преобразования (конвертации) форматов данных, которые имеет смысл рассмотреть более тщательно.

1. *Сетевая схема* — наиболее очевидный способ обеспечения преобразования данных, который заключается в создании процедур конвертации (трансляторов) для каждой из пар интероперабельных форматов (рис. 3). Таким образом, данные могут быть напрямую (и, соответственно, наиболее оптимальным образом) оттранслированы к другому формату.

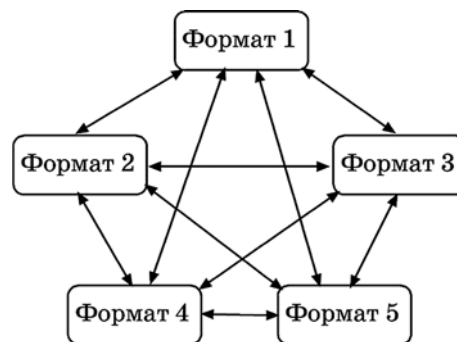


Рис. 3. Сетевая схема преобразования форматов

Недостатком этого метода является то, что трансляторы необходимы между каждой парой форматов данных, что является весьма трудоемкой задачей: для N форматов необходимо создать до

$$K = \sum_{i=1}^N (N - i) \text{ парных процедур трансляции. На}$$

практике их число может быть значительно меньшим, поскольку для некоторых пар форматов обеспечение преобразования не требуется (и даже в принципе невозможно). Кроме того, для некоторых случаев достаточно преобразования в одну сторону, т. е. только для того, чтобы прочитать некоторый специфический формат, сохранять данные в него нет необходимости³.

Достоинством этой схемы является то, что добавление в нее нового формата практически не влияет на уже существующие элементы схемы, достаточно только обеспечить необходимые процедуры трансляции.

³ В частности, форматы первичных данных, а также в случае, когда преобразование связано с невозможной утратой информации и обратное преобразование в принципе бессмысленно, например, преобразование двухмерных координат в трехмерные.

2. *Звездообразная схема* — этот подход заключается в использовании некоторого единого (унифицированного) формата данных, в этом случае преобразование происходит в два этапа: сначала данные из исходного специфического формата преобразуются в унифицированный, из которого затем выполняется преобразование в целевой специфический формат (рис. 4).



Рис. 4. Звездообразная схема преобразования форматов

Это позволяет значительно сократить теоретическое число пар трансляторов до $K = N - 1$, однако в данном случае возникает ряд проблем.

Единый формат должен быть достаточно сложным, чтобы нести в себе все важные особенности, присутствующие в специфических форматах. Если специфические особенности в унифицированном формате не предусмотрены, даже если исходный и целевой форматы их предполагают, то такое преобразование приведет к потерям, которые могут оказаться неприемлемыми. Таким образом, при разработке унифицированного формата необходимо

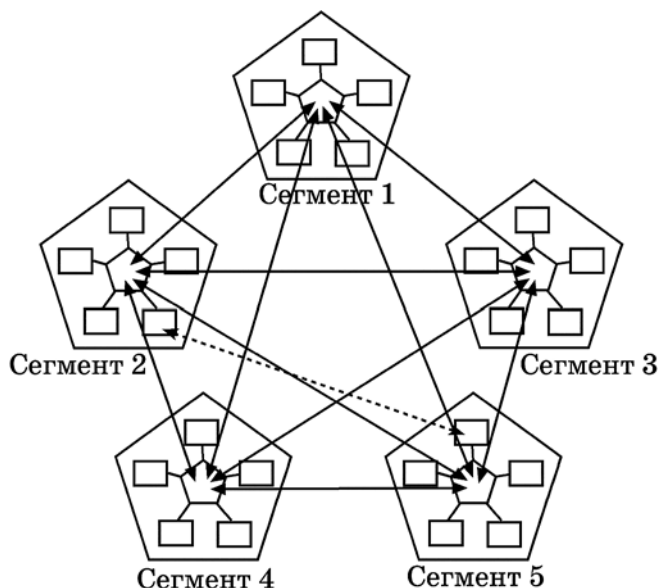


Рис. 5. Иерархическая схема преобразования форматов

сразу учесть все возможные случаи, которые могут возникнуть при эксплуатации такой схемы интероперабельности.

Это же обстоятельство обуславливает проблематичность расширения такой схемы.

Также следует иметь в виду, что многоступенчатое преобразование приводит к более низкой производительности.

3. *Многоуровневая иерархическая схема* преобразования (рис. 5) является разумным компромиссом между приведенными выше радикальными подходами.

В своих общих чертах она подобна сетевой схеме, однако каждый из узлов такой сети представляет собой не конкретный специфический формат, а некую достаточно автономную подсистему (сегмент). Подсистемы строятся уже по звездообразной схеме, т. е. обладают центральным узлом — локальным унифицированным форматом, относительно которого строится окружение из родственных специфических форматов. Аналогичный подход применим и на более локальных уровнях.

Другими словами, основу схемы представляет сеть относительно небольшого числа локальных унифицированных форматов, каждый из которых образует звездообразную схему для семейства родственных специфических форматов или же унифицированных форматов подсистем более низкого ранга.

Такая схема на глобальном уровне обеспечивает простоту и расширяемость сетевой схемы, а на локальных уровнях, где форматы уже более сопоставимы друг с другом, обеспечивает сокращение числа необходимых трансляторов.

В большинстве случаев выполняются преобразования форматов внутри локальной "звезды", так как она объединяет наиболее однородные (близкие по своей структуре и свойствам) форматы, конвертация между которыми сопряжена с наименьшими потерями.

Если необходимо преобразование между форматами из различных подсистем, то оно происходит через более глобальный формат. Последнее обстоятельство можно рассматривать в качестве недостатка такой схемы, так как в этом случае процесс преобразования может включать достаточно большое число этапов трансляции. Однако, если иметь в виду, что форматы, находящиеся в разных подсистемах, достаточно разнородны, то такие преобразования являются скорее исключением, чем правилом, поэтому для таких случаев могут быть предусмотрены специальные процедуры прямой трансляции (показана на рис. 5 штриховой линией).

Предназначением схемы преобразования форматов данных в структуре совместного использования различных источников геоданных является обеспечение возможности клиентскому приложению или серверу геоданных (в зависимости от того, на чьей стороне выполняется преобразование дан-

ных) получить четкие указания относительно того, какие процедуры конвертации и в какой последовательности должны быть проведены в процессе преобразования данных, получаемых или возвращаемых по запросу. Отсюда вытекает следствие, что такая схема должна носить характер стандарта обмена данными, по крайней мере, на межсистемном уровне.

Заключение

Геоинформационные системы в современных реалиях зачастую представляют собой сложные распределенные системы (в том числе облачные), включающие в себя множества источников геоданных и прикладных геоинформационных сервисов, создаваемых и эксплуатируемых различными организациями-операторами, каждая из которых, в свою очередь, может иметь собственные представления о способах организации геоданных и правилах их предоставления потребителям, использовать различные устройства для сбора геоданных.

Вместе с тем, высокая стоимость как создания и эксплуатации хранилищ геоданных, так и сбора и своевременного обновления собственно геоданных, сопровождаемого необходимостью их согласования с уже существующими данными, обуславливает необходимость обеспечения возможности совместного и многократного использования геоинформационных ресурсов различными заинтересованными пользователями.

Для обеспечения как самой возможности совместной работы, так и ее эффективности геоинформационный контент должен быть рационально организован таким образом, чтобы любой пользователь, независимо от специфики решаемых задач или используемого им аппаратно-программного обеспечения, мог обратиться к тем или иным геоданным через эффективный и удобный метод запроса, чтобы быстро получить точную и исчерпывающую информацию по интересующему его вопросу. При этом должна обеспечиваться возможность предоставления геоданных с различной степенью обобщения (генерализации) как в пространственном, так и в семантическом аспектах.

Актуальность проблемы обеспечения геоинтероперабельности (определяемой как способность к взаимодействию в геопространстве) проявляется наиболее наглядно в тех ситуациях, когда необходимо обеспечить взаимодействие геопорталов, использующих сервисы и приложения, в которых реализованы разные модели представления геоданных или синтаксис их описания, а также различающихся методическими и функциональными подходами к обработке данных. Семантическая интероперабельность геоданных заключается в обеспечении их пользователей средствами согласованного понимания семантической составляющей геоданных,

полученных в результате обмена, а также выявлении и исправлении тех ситуаций, когда имеет место несогласованное понимание. В целях выработки подходов к решению проблемы интероперабельности в данной работе был рассмотрен ряд традиционных моделей геоданных.

Подводя итоги, можно сказать, что основными проблемами интероперабельности пространственных данных являются:

- преобразования между пространствами различной размерности;
- различные требования к точности представления координат, а также к точности аппроксимации контуров геообъектов;
- необратимые потери информации при перепроектировании растровых изображений, несоответствия между различными моделями данных (касающиеся пространственной, семантической и топологической составляющих, а также метаданных) и т. п. Вопрос интероперабельности данных должен решаться как на уровне модели данных, так и на уровнях источника данных (ресурса), а в ряде случаев — и конкретного набора данных.

Решение проблемы интероперабельности в геоинформационных технологиях сосредотачивается на обеспечении адекватного обмена данными между геоинформационными системами с различной архитектурой и обеспечении возможности совместного использования как данных, так и функциональных методов их обработки. Разработка и принятие стандартов является важным средством для формирования и развития взаимодействующих геопространственных данных и служб. Вместе с тем, задача обеспечения межсистемного обмена геоданными предполагает существование определенной схемы преобразования различных моделей и форматов данных между собой. Из трех рассмотренных в статье схем наиболее оптимальной представляется многоуровневая иерархическая схема преобразования, она обеспечивает гибкость и расширяемость (масштабируемость) на глобальном уровне и сокращение числа видов преобразований и процедур трансляции на локальных уровнях.

Список литературы

1. **Lemmens R.** Semantic interoperability of distributed geoservices // Publications on Geodesy "63 Netherlands Geodetic Commission (NCG). — Rotterdam: Optima Graphic Communication, 2006. 312 p.
2. **Liu Q., Zhou L. Z., Chen J.** Data Models for Interoperability // Towards Digital Earth — Proceedings of the International Symposium on Digital Earth Beijing: Science Press, 1999. P. 1—6.
3. **Комков А. М., Николаев С. А., Шилов Н. И.** Составление и редактирование карт. М.: Изд-во ВИА, 1958. 248 с.
4. **Дулин С. К., Дулина Н. Г., Никишин Д. А.** О проблемах реализации семантической геоинтероперабельности в Semantic WEB // Системы и средства информатики. 2014. Вып. 24. № 2. С. 143—165.
5. **Интероперабельность** информационных систем. Сб. матер. (сост. М. Брауде-Золотарев). М.: INFO-FOSS.RU, 2008. 128 с.

6. **Zatsman I. M.** Pictorial Signs for Geoimages in Digital Libraries. URL: http://www.researchgate.net/publication/257150374_Pictorial_Signs_for_Geoimages_in_Digital_Libraries.

7. **Шекхар Ш., Чаула С.** Основы пространственных баз данных / Пер. с англ. М.: Кудлиц-Образ, 2004. 336 с.

8. **Берлянт А. М., Кошкарёв А. В., Серапинас Б. Б.** и др. Толковый словарь по геоинформатике. М.: ГИС-обозрение, 1998.

9. **ГОСТ Р 51353—99.** Геоинформационное картографирование. Метаданные электронных карт. Состав и содержание. М.: ИПК Изд-во стандартов, 2000. 8 с.

10. **ISO 19115:2003.** Geographic information — Metadata.

11. **ISO 19139:2007.** Geographic information — Metadata — XML schema implementation.

12. **ISO 19115-2:2009.** Geographic information — Metadata. Part 2: Extensions for imagery and gridded data.

13. **Профиль** метаданных ЕБГД на основе стандартов ISO 19115, ISO 19115-2 и ISO 19139. М.: НИИТП, 2010. 135 с.

14. **Мартыненко А. И., Никишин А. Н., Никишин Д. А.** Единая географическая информационная система: проблемы и стратегии формирования // Системы и средства информатики. М.: Наука, 2007. Т. 17. № 1. С. 355—390.

15. **Скворцов А. В., Мирза Н. С.** Алгоритмы построения и анализа триангуляции. Томск: Изд-во ТГУ, 2006. 168 с.

16. **Формат** цифрового классификатора RSC и библиотеки условных знаков. М.: Панорама, 2009. 34 с.

17. **Открытый** формат цифровой информации о местности SXF. Редакция 4.0. Ногинск: Панорама, 2011. 39 с.

S. K. Dulin^{1, 3}, Chief Researcher, e-mail: s.dulin@ccas.ru,

N. G. Dulina², Senior Researcher, e-mail: ngdulina@mail.ru, **D. A. Nikishin**³, Head of Dextor

¹ Research & Design Institute for Information Technology,
Signalling and Telecommunications on Railway Transport (JSC NIIAS)

² Dorochnyyn Computing Center, Russian Academy of Sciences

³ Institute of Informatics Problems, Russian Academy of Sciences

Features of Models of Geodata and Methods of their Processing in Aspect of Maintenance Semantic Geointeroperability

The Key problem of maintenance of semantic geointeroperability is creation of uniform conceptual model of representation of geodata on the basis of integration of the spatially-distributed information. Such model should be create on the basis of existing metaschemes of bases of geodata with the account multyfactority interactions of users and semantics which has been put in pawn in spatial ontologies and-or geothesauruses and qualifiers. Work is devoted research and the analysis of existing models and methods of representation of geodata and has for an object an estimation of relevance of existing structures geogiven to methods and approaches of maintenance of their semantic geointeroperability at the decision of problems of conceptual search of geodata. For the analysis those structures of geodata which combine in itself means of display of all three groups of the spatially-distributed information have been chosen: semantic, metric and topological. Various levels of representation of a geoinformation content which demand essentially various methods of processing are considered.

Keywords: semantic geointeroperability, modelling of geodata, ontologies, conceptual search

References

1. **Lemmens R.** Semantic interoperability of distributed geo-services. *Publications on Geodesy 63 Netherlands Geodetic Commission (NCG)*. Netherlands, Rotterdam: Optima Graphic Communication, 2006. 312 p.

2. **Liu Q., Zhou L. Z., Chen J.** Data Models for Interoperability. *Towards Digital Earth — Proceedings of the International Symposium on Digital Earth*. Beijing: Science Press, 1999.

3. **Komkov A. M., Nikolaev S. A., Shilov N. I.** *Sostavlenie i redaktirovanie kart*. М.: Изд-во VIA, 1958. 248 p.

4. **Dulin S. K., Dulina N. G., Nikishin D. A.** O problemah realizacii semanticheskoy geointeroperabel'nosti v Semantic WEB. *Sistemy i sredstva informatiki*. 2014. V. 24. N. 2. P. 143—165.

5. **Interoperabel'nost' informacionnyh sistem**. Sbornik materialov (sost. M. Braude-Zolotarev). М.: INFO-FOSS.RU, 2008. 128 p.

6. **Zatsman I. M.** Pictorial Signs for Geoimages in Digital Libraries. URL: http://www.researchgate.net/publication/257150374_Pictorial_Signs_for_Geoimages_in_Digital_Libraries.

7. **Shekhar S., Chaula S.** *Osnovy prostranstvennyh baz dannyh*. Per. s angl. М.: Kudic-Obraz, 2004. 336 p.

8. **Berljant A. M., Koshkarev A. V., Serapinas B. B.** i dr. *Tolkovyj slovar' po geoinformatike*. М.: GIS-obozrenie, 1998.

9. **ГОСТ Р 51353—99.** *Geoinformacionnoe kartografirovanie. Metadannye jelektronnyh kart. Sostav i sodержanie*. М.: ИПК Изд-во стандартов, 2000. 8 p.

10. **ISO 19115:2003.** *Geographic information — Metadata*.

11. **ISO 19139:2007.** *Geographic information — Metadata — XML schema implementation*.

12. **ISO 19115-2:2009.** *Geographic information — Metadata. Part 2: Extensions for imagery and gridded data*.

13. **Profil' metadannyh EBGD na osnove standartov ISO 19115, ISO 19115-2 i ISO 19139**. М.: НИИТП, 2010. 135 p.

14. **Martynenko A. I., Nikishin A. N., Nikishin D. A.** Edinaja geograficheskaja informacionnaja sistema: problemy i strategii formirovanija. *Sistemy i sredstva informatiki*. М.: Наука, 2007. V. 17. P. 355—390.

15. **Skvorcov A. V., Mirza N. S.** *Algoritmy postroenijai analiza trianguljacii*. Tomsk: Изд-во ТГУ, 2006. 168 p.

16. **Format cifrovogo klassifikatora RSC i biblioteki uslovnyh znakov**. М.: Panorama, 2009. 34 p.

17. **Otkrytyj format cifrovoj informacii o mestnosti SXF. Redakcija 4.0**. Nогинск: Panorama, 2011. 39 p.