

УДК 528.83

ИНТЕГРАЦИЯ В ЕДИНОМ ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ РАЗНОРОДНЫХ ГЕОПРОСТРАНСТВЕННЫХ ДАННЫХ

С. А. Карин,

канд. воен. наук

Военно-космическая академия им. А. Ф. Можайского, г. Санкт-Петербург

Определяется роль и место геопространственных данных в системах информационного обеспечения органов государственного управления. Рассматриваются современные технологии представления и интеграции разнородных геоданных, обосновываются их основные недостатки. Описывается метод интеграции разнородных геопространственных данных на основе концепции метаданных в современных автоматизированных межвидовых системах информационного обеспечения и взаимодействия, преодолевающий указанные недостатки.

Ключевые слова — единое информационное пространство, геоинформационные системы, геопространственные данные, профиль метаданных.

Введение

Анализ современных взглядов на организацию процессов управления экономической, социальной, политической и оборонной деятельностью государства показал, что одним из важнейших условий достижения поставленной цели является своевременное обеспечение органов управления точной и полной информацией. При этом необходимо отметить постоянно возрастающую роль взаимозависимости, взаимовлияния сложных систем, применяемых в различных взаимодействующих органах.

Данное обстоятельство привело к развитию современных автоматизированных межвидовых систем информационного обеспечения и взаимодействия (АМС ИОВ), функционирующих в среде единого информационного пространства, которое представляет собой совокупность взаимодействующих разнородных технических, программно-алгоритмических и информационных средств, использование которых направлено на достижение единой цели.

В ряду организационных, технологических и технических проблем, стоящих перед разработчиками АМС ИОВ, важной составной частью которых являются геоинформационные системы (ГИС), ключевое место занимает задача интеграции и совместного использования разнородных геопространственных и атрибутивных данных, добыва-

емых, хранящихся и обрабатываемых в различных по назначению и способам реализации программно-технических комплексах. Сложность и трудоемкость этой задачи обусловлена следующими особенностями, которые присущи циркулирующей в таких системах информации:

- значительная часть (до 80 и более процентов) информации имеет пространственную и временную привязку (т. е. является геопространственной);

- для представления однотипных данных, особенно пространственных, в разных системах используются разные и несовместимые между собой форматы представления;

- объемы информации, которую необходимо обработать, сохранить и предоставить потребителю, в настоящее время достигают размеров в десятки и сотни терабайт и продолжают возрастать.

Основу любой ГИС составляет реализованная в ней модель представления геопространственных данных (ГПД), которая выступает как система правил цифрового моделирования реальных объектов путем описания типов и свойств элементарных пространственных объектов, их наборов и межобъектных отношений. При этом практически в каждой существующей сегодня прикладной ГИС или специализированном банке пространственных данных используются собственные форматы и модели представления и хранения обрабатываемых в них данных.

Примером реализации таких моделей является академический профиль пространственных метаданных GeoMETA v3.0:2007 [1].

Наличие большого числа различных видов и форматов ГПД обусловлено как объективными обстоятельствами (их природой и разнообразием источников формирования и способов получения), так и субъективным фактором, состоящим в предпочтении разработчиков конкретных систем в отношении математической основы и информационно-логических моделей, используемых для их представления и обработки.

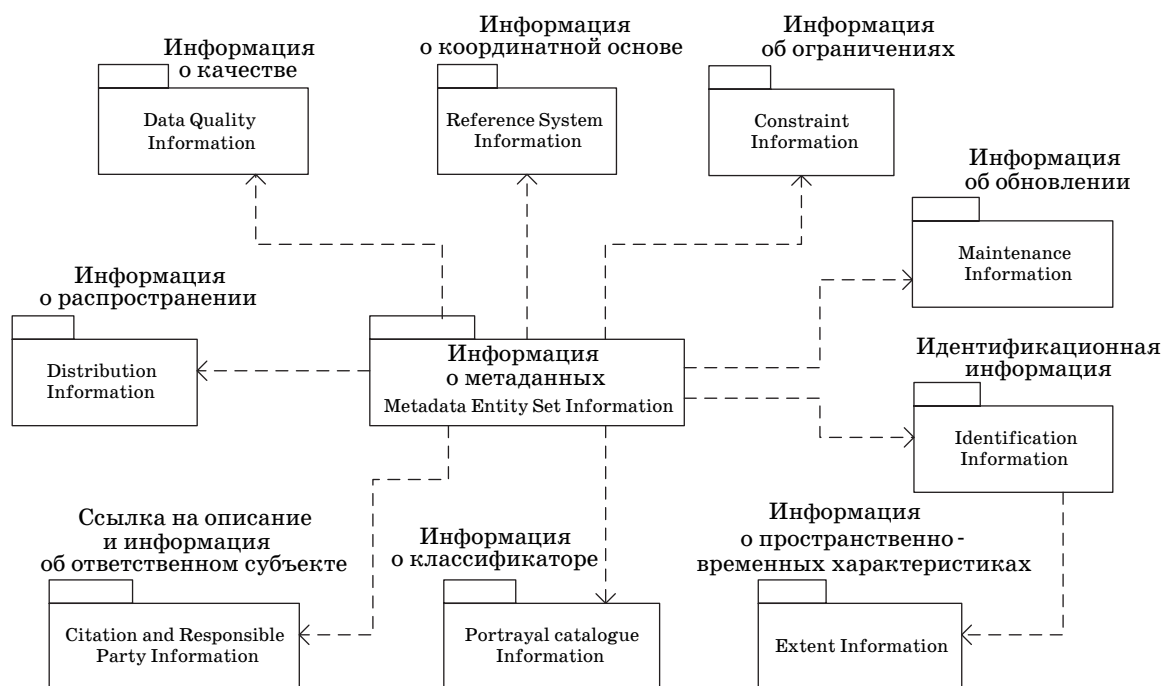
Для обеспечения интеграции и совместного использования разнородных ГПД, поступающих от различных источников и взаимодействующих систем, применяются обменные форматы и конверторы. Как показывает практика, конвертирование (преобразование) таких данных из одного формата в другой, во-первых, не всегда имеет единственное решение и поэтому предполагает обязательное участие в этом процессе человека, а во-вторых, эта операция требует значительных вычислительных и временных ресурсов. Все это отрицательно влияет на точность и оперативность подготовки и доведения необходимой потребителю информации.

Выходом из создавшегося положения является разработка некоторой универсальной (для определенного круга задач и используемых информационных ресурсов) модели представления ГПД и соответствующих методов их автоматического учета, оперативного поиска, обработки и доведения до потребителей.

В настоящей статье описан подход к решению данной задачи применительно к системам информационного обеспечения органов управления. Предложенные модели и методы базируются на концепции метаданных, современных технологиях проектирования и разработки ГИС и баз пространственных данных, а также теории формальных грамматик и концепции единого информационного пространства.

Отечественные и международные стандарты в области создания профилей метаданных

Важную роль в процессе объединения усилий международного сообщества потребителей и разработчиков геоинформации в создании единых стандартов и моделей представления ГПД играет технический комитет ISO/TC 211. Под его руководством разработаны методология, технические регламенты и базовые инструменты для проектирования, разработки и эксплуатации систем сбора, хранения, обработки и распространения ГПД. Основные положения этих материалов изложены в таких стандартах, как ISO 19115:2003. Geographic information — Metadata (Географическая информация. Метаданные) [2], ISO 19118 Geographic information — Encoding (Географическая информация. Кодирование) [3] и ISO 19139:2007. Geographic information — Metadata — XML schema implementation (Географическая информация. Метаданные. Спецификация реализации) [4].



■ Рис. 1. Пакеты метаданных ГОСТ Р 52537–2006

Стандарт ISO 19115:2003 описывает состав и содержание метаданных, а также взаимоотношения между элементами метаданных. Набор метаданных ISO 19115 содержит более 400 элементов и атрибутов, сгруппированных в 95 классов, при этом 22 элемента составляют ядро, включающее 7 обязательных, 4 условных и 11 необязательных элементов. Стандарт 19115 представляет собой концептуальную UML-модель метаданных. Правила и методы превращения логических UML-абстракций в кодированные описания в виде XML-схем описаны в стандарте ISO 19118. Детально же описанию методов кодирования и специфических приемов их применения для UML-моделей метаданных посвящена техническая спецификация ISO 19139.

В нашей стране принят ГОСТ Р 52573–2006 «Географическая информация. Метаданные» [5], который является адаптированным для нашей страны аналогом стандарта ISO 19115:2003 и в соответствии с ним разработан на основе унифицированного языка моделирования (Unified Modeling Language — UML). Все типы метаданных в стандарте ГОСТ Р 52573–2006 представлены в виде совокупности UML-пакетов. Пакет состоит из одной или более сущностей, связанных отношениями обобщения или агрегирования. Каждая сущность представляет собой совокупность элементов (атрибутов), характеризующих тот или

иной аспект метаданных. Основные пакеты метаданных, связанные между собой отношением зависимости (данное отношение описывает ситуацию, при которой изменения в независимом элементе модели требуют изменений в зависимом элементе), представлены на рис. 1.

Метод интеграции разнородных ГПД в автоматизированных системах органов управления

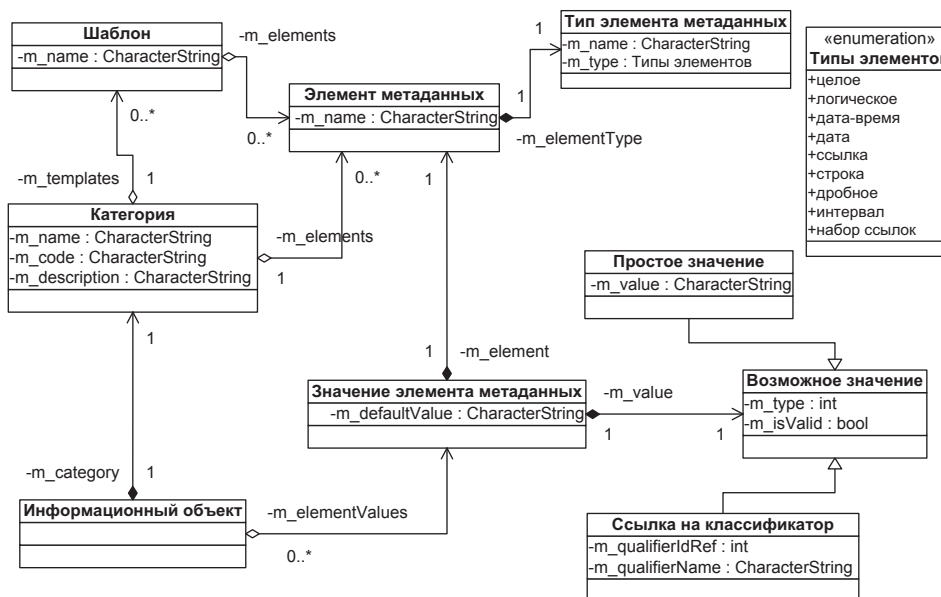
Подход, основанный на создании профилей метаданных, активно применяется и при автоматизации процессов, происходящих в органах государственного и муниципального управления.

В целях разработки онтологии унифицированного профиля метаданных ГПД для перспективной АМС ИОВ органов управления был выполнен анализ состава и особенностей решаемых с ее помощью прикладных задач, источников и форматов используемых при этом исходных данных, требований к содержанию и форме выходных документов. По результатам анализа были определены состав и содержание набора метаданных разрабатываемого профиля.

UML-диаграмма пакетов данного профиля представлена на рис. 2. В отличие от базового набора метаданных, определенного в ГОСТ 52573–2006, в этот профиль были добавлены два новых



■ Рис. 2. UML-диаграмма пакетов унифицированного профиля метаданных для использования в АМС ИОВ органов управления



■ Рис. 3. UML-диаграмма пакета «Информация специального назначения»

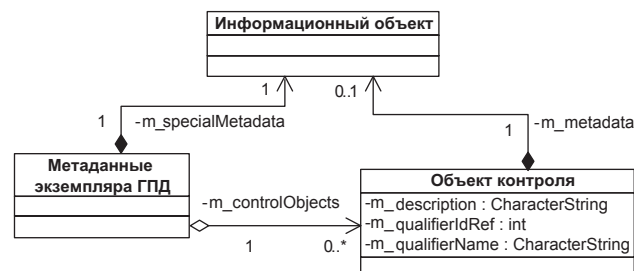
пакета, исключены три пакета и внесены изменения в два пакета.

Проведенный анализ показал, что основными типами используемых и разрабатываемых в АМС ИОВ органов управления ГПД являются цифровые векторные и растровые карты, растровые геокодированные изображения, цифровые данные о рельефе местности (векторные и растровые), тематические слои полной и типовой векторизации, тематические трехмерные модели местности и объектов, фоноцелевая информация, слои данных об оперативной обстановке, слои данных о метеобстановке, электронные фотосхемы (ЭФС), слои с результатами решения специальных расчетно-аналитических задач. Для обеспечения возможности автоматизированного или автоматического построения описаний этих достаточно разнородных ГПД в состав набора метаданных были включены следующие сущности (UML-классы): информационный объект, категория информационного объекта, тип элемента метаданных, шаблон представления информационного объекта, ссылка на классификатор и некоторые другие. Все эти сущности собраны в пакет «Информация специального назначения», диаграмма классов которого представлена на рис. 3.

В связи с тем, что задачи информационного обеспечения пользователей АМС ИОВ органов управления всегда связаны с конкретными объектами контроля, управления и воздействия, в состав набора метаданных разработанного профиля ГПД предложено ввести еще один пакет — «Информация об объективном составе». Этот пакет содержит классы, предназначенные для опи-

сания объектового состава конкретных экземпляров ГПД. Примерами таких объектов являются места расположения объектов воздействия, пункты управления, узлы связи, склады, транспортные системы и узлы, аэродромы, линии электропередач, объекты инфраструктуры электроэнергетики и транспорта и др. Основной класс данного пакета метаданных «Объект контроля» и его взаимосвязи с другими пакетами разработанного профиля представлены на рис. 4.

Примером информационного объекта может служить ЭФС, которая характеризуется определенным набором элементов метаданных (название, масштаб, автор, время исполнения, объект и т. д.) и включает несколько прикрепленных файлов. В этих файлах содержатся слои, из которых, собственно, и состоит ЭФС. Например, она может включать слой с растровым геокодированным изображением, тематические слои полной и тематической векторизации, цифровую картографическую основу в виде электронной топокарты и векторную координатную сетку.



■ Рис. 4. Пакет метаданных «Информация об объективном составе» и его взаимосвязи

При этом ЭФС может описывать один или несколько объектов контроля, которые имеют свой собственный набор элементов метаданных (наименование объекта контроля, состояние, код объекта, принадлежность, координатная и временная привязка т. д.).

Данный пример показывает, что, с одной стороны, ЭФС является информационным объектом, а с другой стороны, каждый из присутствующих на ней объектов управления и воздействия также можно рассматривать в качестве информационного объекта.

Важно отметить, что представленная на рис. 3 диаграмма классов содержит все необходимые элементы для описания используемых в АМС ИОВ органов управления ГПД. В связи с этим она была положена в основу программной реализации метода формирования метаданных ГПД. Данный метод обеспечивает преобразование концептуальной UML-модели профиля в конкретную XML-схему и XML-описания в строгом соответствии с требованиями стандартов ISO 19118 и 19139.

Полученная в результате этого совокупность XML-описаний представляет собой описание профиля метаданных ГПД на некотором формальном языке $L(G)$, который в соответствии с теорией формальных грамматик [6–8] задается грамматикой L , т. е.

$$L(G) = \{x \in T^* \mid S \Rightarrow_G^* x\},$$

$$G = (N, T, P_G, S),$$

где N — множество нетерминальных символов; T — множество терминальных символов таких, что $T \cap N = \emptyset$; P_G — множество правил вывода (продукций) вида $A \rightarrow \alpha$: $\alpha = aAa'$, $\alpha = aBa'$, $\alpha = a$, $A \in N$, $a \in T$, $a' \in T$; S — начальный символ (аксиома грамматики).

Очевидно, что для ГПД, представленных в других форматах, также возможно построить грамматики, порождающие соответствующий им язык. В ходе проведенных исследований были проанализированы следующие форматы представления метаданных пространственных данных:

- формат паспорта карты ГИС «Панорама»;
- профиль метаданных ESRI;
- модель пространственных данных ГИС MapInfo.

По результатам проведенного анализа определены правила вывода порождающей грамматики:

- для формата паспорта карты ГИС «Панорама»

$$\alpha = \beta, \beta \in (T)^*,$$

$$\alpha = \beta A, \beta \in (T)^*, A \in N;$$

- для форматов профиля метаданных ESRI и MapInfo

$$\alpha = \beta A \gamma \mid \beta, \gamma \in (T)^*, A \in N,$$

$$A \rightarrow \xi_H \psi \xi_K \mid \xi_H, \xi_K \in T, \psi \in (T \cup N).$$

Последнее правило означает, что множество терминальных символов разбито на пары, и всегда можно выделить терминальный символ, открывающий, а также закрывающий правило вывода. При этом открывающий терминальный символ не может являться закрывающим и наоборот.

С учетом вышеизложенного задача преобразования метаданных ГПД, представленных в форматах, отличных от унифицированного профиля, в формат унифицированного профиля состоит в разработке двух следующих объектов:

- 1) трансляционной грамматики G_T вида

$$G_T = \langle N_T, T_T, P_{G_T}, S_T \rangle,$$

$$N_T = N_B \cup T_B; T_T = T \cup T'; P_{G_T} = P_{G_B} \cup P_G \cup P';$$

$$P': \phi \rightarrow \psi \mid \phi \in (T_B \cup N_B)^*, \psi \in (T \cup N)^*; S_T = S;$$

- 2) преобразователя с магазинной памятью M (МП-преобразователя), который производит синтаксический анализ входных данных и осуществляет их перевод с исходного языка $L_B(G_B)$ на язык $L(G)$ вида

$$M = \langle Q, A, \Delta, \Gamma, \delta, q_0, z_0, F \rangle,$$

где $Q = \{q_0\}$ — множество состояний МП-преобразователя; $A = T_B$ — входной алфавит; $\Delta = T$ — выходной алфавит; $\Gamma = T_B \cup N_B$ — магазинный алфавит; δ — отображение вида $Q \times \{A \cup \{\epsilon\}\} \times \Gamma \rightarrow Q \times \Gamma^* \times \Delta^*$; q_0 — начальное состояние; $z_0 = \emptyset$ — начальный символ в магазинной памяти; $F = \emptyset$ — множество заключительных состояний.

Заключение

Описанный выше метод решения задачи интеграции и совместного использования разнородных пространственных данных на основе унифицированного профиля метаданных является результатом научной и практической деятельности автора на протяжении 2007–2011 гг. Метод лег в основу ряда автоматизированных информационных систем, разрабатываемых специалистами Военно-космической академии им. А. Ф. Можайского для учреждений Министерства обороны Российской Федерации, а также органов муниципального и государственного управления г. Санкт-Петербурга и Ленинградской области. Кроме того, предложенный метод использован при создании единого информационного пространства ВКА им. А. Ф. Можайского, в частности геопортала кафедры систем дистанционного зондирования Земли.

Литература

1. **Пространственные** метаданные в системе «Гео-МЕТА». <http://www.gisa.ru/47191.html> (дата обращения: 02.12.2011).
2. **International Standart ISO 19115 Geographic information — Metadata.** — ISO, 2003. — 140 p.
3. **International Standart ISO 19118 Geographic information — Encoding.** — ISO, 2005. — 104 p.
4. **International Standart ISO/TS 19139:2007 Geographic information — Metadata — XML-schema implementation.** — ISO, 2007. — 111 p.
5. **ГОСТ Р 52573–2006.** Географическая информация. Метаданные. — М.: Стандартинформ, 2006. — 59 с.
6. **Алферова З. В.** Теория алгоритмов. — М.: Статистика, 1973. — 165 с.
7. **Пентус А. Е., Пентус М. Р.** Теория формальных языков: учеб. пособие. — М.: Изд-во ЦПИ при мех.-мат. ф-те МГУ, 2004. — 80 с.
8. **Ахо А., Ульман Дж.** Теория синтаксического анализа, перевода и компиляции. Т. 1: Синтаксический анализ. — М.: Мир, 1978. — 612 с.

УВАЖАЕМЫЕ АВТОРЫ!

При подготовке рукописей статей редакция просит Вас руководствоваться следующими рекомендациями.

Объем статьи (текст, таблицы, иллюстрации и библиография) не должен превышать эквивалента в 20 страниц, напечатанных на бумаге формата А4 на одной стороне через 1,5 интервала в Word шрифтом Times New Roman размером 13, поля: слева три сантиметра, остальные не менее двух.

Обязательными элементами оформления статьи являются: индекс УДК, заглавие, инициалы и фамилия автора (авторов), ученая степень, звание, полное название организации, аннотация (7–10 строк) и ключевые слова на русском и английском языках, подрисуночные подписи.

Формулы в текстовой строке набирайте в Word, не используя формульный редактор (Mathtype или Equation), только в том случае, если средства Word не позволяют набрать формулу или символ (например, простая дробь, символы с «крышками» и т. д.), используйте имеющийся в Word формульный редактор Mathtype или Equation; формулы, стоящие в отдельной строке, могут быть набраны как угодно; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта никогда не пользуйтесь вкладкой Other..., используйте вкладку Define; в формулах не отделяйте пробелами знаки: + = –; не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст.

При наборе символов в тексте помните, что символы, обозначаемые латинскими буквами, набираются светлым курсивом, русскими и греческими — светлым прямым, векторы и матрицы — прямым полужирным шрифтом.

Иллюстрации:

— рисунки, графики, диаграммы, блок-схемы предоставляйте в виде отдельных исходных векторных файлов, поддающихся редактированию: *.vsd, *.cdr, *.xls, *.doc, *.ai, *.dxf;

— при наличии надписей на рисунке используйте тот же шрифт, что и в основном тексте (Times New Roman), размер шрифта не более 10 pt, но не менее 8 pt;

— если при изготовлении рисунка Вы используете стрелочки, руководствуйтесь принципом единообразия;

— фото и растровые — в формате *.tif, *.png с максимальным разрешением (не менее 300 pixels/inch).

В редакцию предоставляются:

— сведения об авторе (фамилия, имя, отчество, место работы, должность, ученое звание, учебное заведение и год его окончания, ученая степень и год защиты диссертации, область научных интересов, количество научных публикаций, домашний и служебный адреса и телефоны, факс, эл. адрес), фото авторов: анфас, в темной одежде на белом фоне, должны быть видны плечи и грудь, высокая степень четкости изображения без теней и отблесков на лице, фото можно представить в электронном виде в формате *.tif, *.png с максимальным разрешением — не менее 300 pixels/inch при минимальном размере фото 40 × 55 мм;

— экспертное заключение.

Список литературы составляется по порядку ссылок в тексте и оформляется следующим образом:

— для книг и сборников — фамилия и инициалы авторов, полное название книги (сборника), город, издательство, год, общее количество страниц;

— для журнальных статей — фамилия и инициалы авторов, полное название статьи, название журнала, год издания, номер журнала, номера страниц;

— ссылки на иностранную литературу следует давать на языке оригинала без сокращений;

— при использовании web-материалов указывайте адрес сайта и дату обращения.

Более подробную информацию см. на сайте: www.i-us.ru