

# РАСПРЕДЕЛЕННАЯ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА ДЛЯ ПОИСКА, ОБРАБОТКИ И АНАЛИЗА ПРОСТРАНСТВЕННЫХ ДАННЫХ

Ю. И. Шокин, О. Л. Жижимов, И. А. Пестунов,  
Ю. Н. Синявский, В. В. Смирнов

*Институт вычислительных технологий СО РАН, Новосибирск, Россия*  
e-mail: pestunov@ict.nsc.ru

A new model of a distributed informational-analytical system (with one way to access data via heterogenous spatial distributed metadatabases) and software tools for its visualization and processing using high performance computational systems is proposed.

**Введение.** Создание информационных ресурсов и интеграция их в единую информационную среду являются приоритетными направлениями развития современного общества. Разработка механизмов, обеспечивающих как функционирование общей информационно-аналитической среды, так и доступ к научным ресурсам и их сохранность, имеет первостепенное значение в задачах информационной поддержки научных исследований. Эти вопросы приобретают особую важность при исследованиях в области экологии и природопользования, выполняемых в рамках междисциплинарных интеграционных проектов СО РАН различными группами ученых, разделенными географически, которым необходимо осуществлять совместную работу, обмен данными и координировать свои действия.

Применение информационных технологий в науках о Земле способствует пониманию как глобальных, так и региональных природных процессов, формирующих окружающую среду. Развивающиеся на протяжении уже нескольких десятилетий геоинформационные технологии в настоящее время переходят на новый этап развития — создание распределенных ГИС (РГИС). Предпосылками этого перехода стали внедрение передовых технологий (распределенные базы данных, распределенные вычисления, высокоскоростные линии связи) и доступность веб-технологий широкому кругу пользователей. Кроме того, этому способствует стремительный прогресс в области создания и развития средств и технологий дистанционного зондирования Земли, обеспечивающий миллионам пользователей доступ к данным, получаемым с помощью спутниковых систем нового поколения (QuickBird, IKONOS, OrbView, GeoEye, SPOT, TerraSarX, Ресурс-ДК и др.) [1]. С другой стороны, в последние годы крупнейшие разработчики программного

обеспечения объединяются в консорциумы по стандартизации. В области геоинформатики создан Консорциум открытых ГИС (Open Geospatial Consortium — OGC) [2]. Соответствие стандартам OGC позволяет программным продуктам различных производителей (в том числе коммерческим) взаимодействовать при решении конкретных задач. Переходу к РГИС также способствует внедрение спутниковых систем навигации NAVSTAR GPS и ГЛОНАСС, позволяющих с высокой точностью определять местоположение объекта в любой точке планеты.

Распределенные геоинформационные системы имеют неоспоримые преимущества перед настольными благодаря таким факторам, как:

- распределенный доступ к системе (наличие веб-интерфейса, позволяющего избежать установки дорогостоящего программного обеспечения, простота изменения и обновления программного обеспечения, доступность широкому кругу пользователей);
- распределенное хранение данных (организация доступа к архивам данных, возможность хранения пользовательских данных на сервере);
- распределенная обработка данных (возможность проведения обработки на высокопроизводительных вычислительных системах).

Вследствие этого последние годы проблеме создания РГИС уделяется все большее внимание [3–9]. В настоящей работе предлагается модель распределенной информационно-аналитической системы с единой точкой доступа к геоданным (через разнородные пространственно распределенные базы метаданных), а также к инструментарию для их визуализации и обработки на высокопроизводительных вычислительных системах. Потребность в системах подобного рода особенно остро ощущается при проведении фундаментальных и прикладных исследований в области экологии и рационального природопользования.

**Принципы построения и структура системы.** Предлагаемая система базируется на операционных системах семейства UNIX и наборе стандартных и специализированных программных продуктов с открытым исходным кодом, распространяемых под лицензией GPL (GNU General Public License). Она полностью удовлетворяет требованиям OGC, предъявляемым к геоинформационным системам, и допускает возможность подключения других ГИС.

Основная цель разработки системы — создание виртуальной информационно-аналитической среды для поиска, обработки и анализа пространственных данных. Она позволит организовать единую точку доступа к различным геоинформационным системам и распределенным разнородным хранилищам геоданных и атрибутивной информации.

Структура предлагаемой системы представлена на рис. 1. Доступ к системе осуществляется через стандартный веб-браузер, что обеспечивает платформенную независимость. Благодаря многоуровневой системе разграничения прав доступа после авторизации происходит генерация “на лету” графического интерфейса пользователя (в соответствии с уровнем доступа). Интерфейс представляет собой виртуальный рабочий стол с привычными пользователю элементами управления (рис. 2). Ядро виртуальной среды состоит из набора Perl/PHP/JavaScript-приложений (с поддержкой технологии AJAX) и интерфейсов к внешним компонентам системы, работающих под управлением веб-сервера Apache. Внешние приложения взаимодействуют с системой через интерфейсы, описанные на языке XML.

Система состоит из следующих функциональных блоков.

В качестве *HTTP-сервера* используется Apache 2.0 (с расширением Tomcat) для платформы UNIX. Веб-сервер Apache разрабатывается и поддерживается открытым

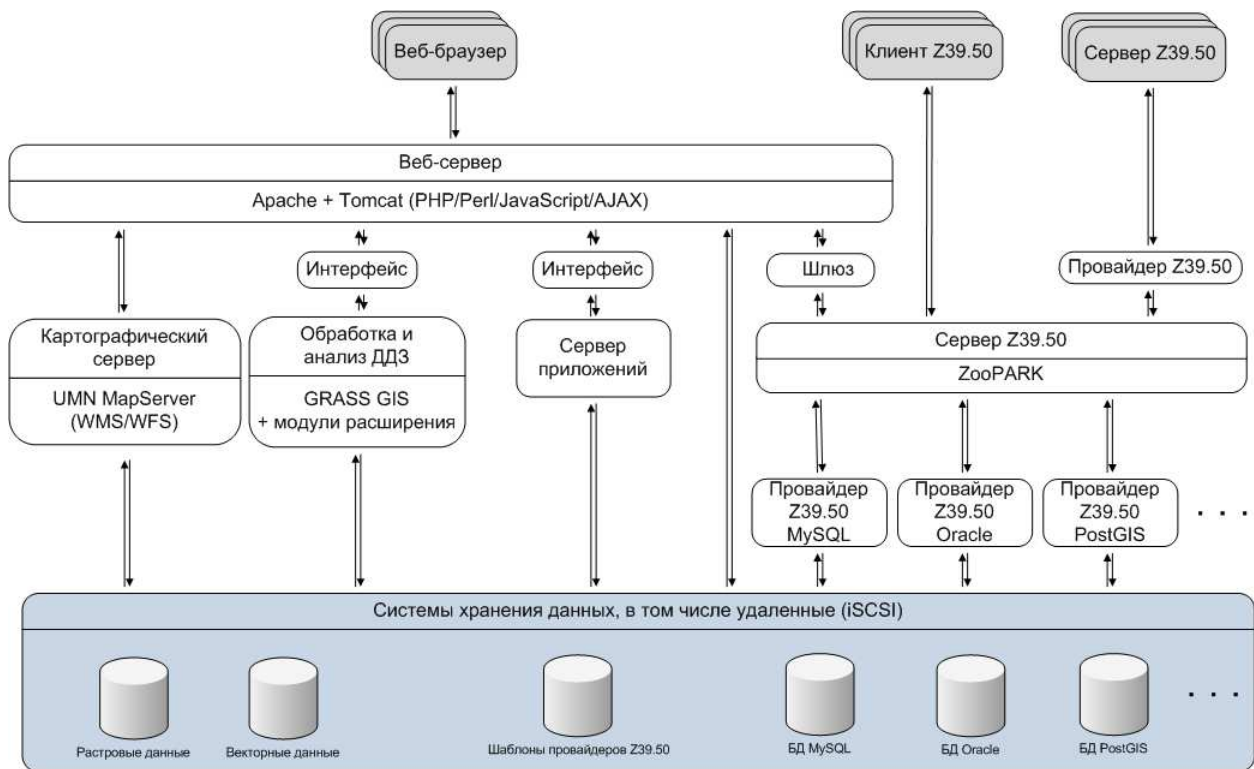


Рис. 1. Структура системы

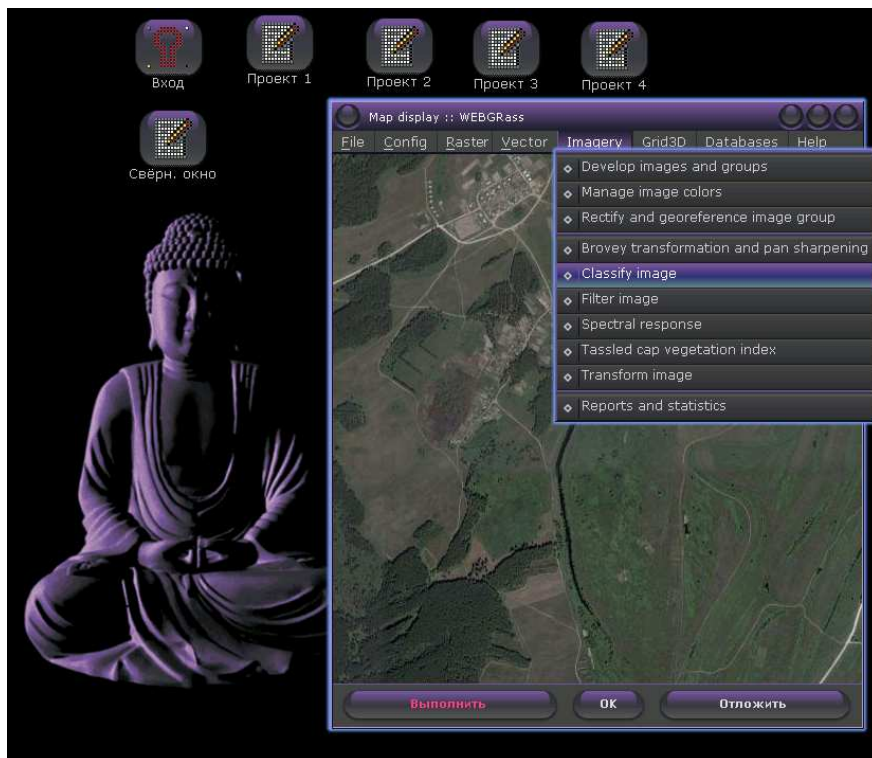


Рис. 2. Интерфейс системы

сообществом разработчиков под эгидой Apache Software Foundation<sup>1</sup> и включен во многие программные продукты, среди которых СУБД Oracle и IBM WebSphere. К основным достоинствам Apache относятся надежность и гибкость конфигурации.

Для публикации пространственно-координированных данных используется *картографический сервер* UMN MapServer<sup>2</sup>. В настоящее время данный продукт становится одним из наиболее популярных инструментов для создания геоинформационных веб-сервисов. По функциональным возможностям он не уступает коммерческим программам, а по ряду параметров превосходит их, в частности по производительности [10]. MapServer включает в себя все необходимое для разработки основных картографических сервисов WMS/WFS, рекомендованных к использованию OGC [11]. Он позволяет пользователю формировать карты с одновременным использованием материалов, хранящихся как в локальных, так и в удаленных архивах. Богатая функциональность, легкость интеграции с различными СУБД и открытость исходных кодов предопределили популярность программы.

В качестве базового *инструментария для обработки и анализа данных дистанционного зондирования* используется пакет программ GRASS GIS (Geographic Resources Analysis Support System<sup>3</sup>). Отличительные особенности пакета — его интеграция в среду UNIX, поддержка основных типов пространственных данных, мощный процессор обработки растровых данных, модульность и открытый инструментарий для быстрой и эффективной разработки модулей расширения. Использование в пакете библиотек GDAL и PROJ4 обеспечивает поддержку всех современных стандартов геоданных и большой набор функций для трансформации и перепроецирования изображений. По функциональности GRASS GIS ни в чем не уступает коммерческим системам. Он позволяет разрабатывать модули расширения практически на всех языках программирования, для которых есть компилятор под UNIX (Perl, sh, C/C++, Fortran и т. д.). Пакет позволяет выполнять ресурсоемкие алгоритмы на высокопроизводительных вычислительных системах. Он включает библиотеки для работы практически со всеми современными СУБД. Внутреннее представление растровых данных базируется на вокселях (3D-пикселях), что существенно повышает возможности по обработке.

В последние годы во многих университетах США и Европы GRASS GIS используется как базовый прикладной пакет для обучения студентов и проведения исследований. Так, например, в университете Ганновера<sup>4</sup> курс изучения основ пакета включен в программу обучения. Интеграция пакета GRASS GIS в разрабатываемую систему позволяет с минимальными временными затратами обеспечить пользователя доступом к полнофункциональной ГИС, расширенной специализированным математическим аппаратом.

Большое внимание в системе уделено статистической обработке данных и визуализации результатов. Для этих целей в нее интегрирован статистический пакет R.Stat<sup>5</sup>, распространяемый на условиях лицензии GPL.

Для расширения функциональности системы используется *сервер приложений*. Он содержит интерфейсы для взаимодействия с внешними приложениями, описанные на языке XML.

---

<sup>1</sup><http://www.apache.org/>

<sup>2</sup><http://mapserver.gis.umn.edu>

<sup>3</sup><http://grass.itc.it>

<sup>4</sup><http://www.gdf-hannover.de>

<sup>5</sup><http://www.r-project.org>



Рис. 3. Основные блоки сервера ZooPARK

Для обеспечения функционирования предлагаемой системы в распределенном режиме и интероперабельности по протоколам доступа к метаданным и их представлению описываемая система предусматривает включение *модулей поддержки протокола Z39.50* [12, 13] (и CIP [14, 15] как одного из его профилей). Для этого в систему включены следующие программные модули:

— сервер Z39.50 (ZooPARK), обеспечивающий базовую функциональность сервисов Z39.50 в соответствии с различными прикладными профилями (рис. 3);

— шлюз Z39.50-HTTP, обеспечивающий простые пользовательские интерфейсы для доступа к ресурсам Z39.50 по протоколу HTTP;

— набор динамических провайдеров данных, каждый из которых описывает условия и протокол взаимодействия с конкретной СУБД, в которой хранятся метаданные. Вся логика работы с конкретной СУБД локализована в соответствующем провайдере данных. Взаимодействие базового сервера с провайдерами данных осуществляется через единый интерфейс.

Для обеспечения логической интероперабельности метаданных предусмотрено применение прикладного профиля Z39.50 GEO [16], в основе которого лежит стандарт CSDGM (Content Standard for Digital Geospatial Metadata, [17]) американского федерального комитета по географическим данным (Federal Geographic Data Committee<sup>6</sup>).

Заметим, что описание и комментарий профиля GEO выходят за рамки настоящей работы, так же как и описание более специализированного профиля CIP [14, 15], поддержка которого в настоящее время в программном комплексе на основе сервера ZooPARK реализована только частично. Что же касается профиля GEO, то на данный момент сервер ZooPARK позволяет

- обрабатывать запросы по протоколу Z39.50 версии 3 над TCP/IP с поддержкой:
  - поиска, в том числе многопоточного, в многобазовом окружении;
  - представления данных в различных форматах (SUTRS, GRS-1, USmarc, UNIMARC, RUSmarc, HTML, RTF, XML);

<sup>6</sup><http://www.fgdc.gov/>

- различных схем данных (GILS, GEO, CIMI, ZTHES и др.);
  - работы со словарем;
  - создания и удаления именованных сеансовых наборов данных на сервере;
  - работы с логическими группами баз данных;
  - контроля и ограничения доступа к базам данных по IP-адресам;
  - аутентификации пользователей, в том числе в каталогах LDAP;
  - различных кодовых таблиц (UTF8, DOS, WIN, ISO, MAC) для кириллических текстов;
  - специализированного профиля Explain (представление конфигурационной информации сервера);
  - связи с любыми СУБД через специализированные динамически загружаемые модули (провайдеры данных);
- обрабатывать запросы по протоколу HTTP 1.0/1.1 на одном порту с Z39.50 с поддержкой:
- функций веб-сервера с интерпретаторами PHP и Perl;

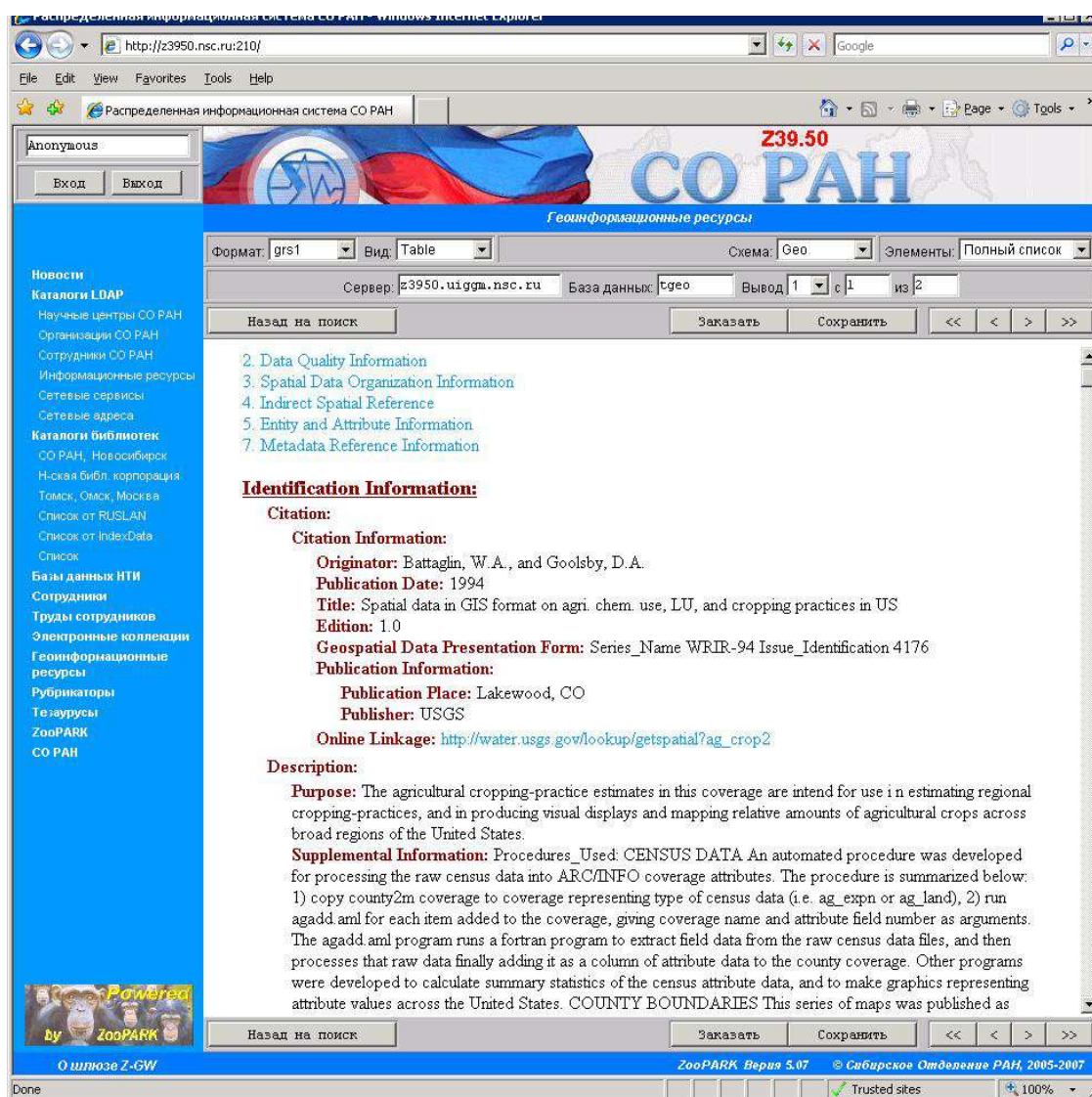


Рис. 4. Представление данных GEO через шлюз сервера ZooPARK

- SOAP (SRW/SRU);
- встроенного шлюза Z39.50-HTTP.

Для иллюстрации на рис. 4 приведен пример представления данных GEO через шлюз сервера ZooPARK. Более подробная информация о сервере ZooPARK доступна по адресу: <http://z3950.uiggm.nsc.ru:210/index.htm> (раздел “Документация”).

Таким образом, на основе сервера ZooPARK можно создавать распределенные информационные системы, поддерживающие профиль GEO. Они могут взаимодействовать с другими системами как в России (см., например, [18]), так и за рубежом.

Для добавления данных дистанционного зондирования в хранилище системы планируется создать специализированный сервис, отвечающий за первоначальную обработку и индексирование поступающих данных. После индексирования данные перемещаются для долгосрочного хранения в систему хранения данных, а полученные метаданные приводятся к стандартному виду и размещаются в соответствующих разделах поисковой системы. Только после этого данные становятся общедоступными.

Поисковая система позволяет не только находить данные по метаданным, но и выполнять комплексные запросы (содержащие географические координаты, номера трека и кадра, дату и время, параметры облачности и др.).

Разрабатываемая система использует стандартизированные внешние протоколы, что позволяет легко обмениваться данными с другими системами (например, с европейской информационной системой INFEO).

**Заключение.** В работе описана модель распределенной информационно-аналитической системы, обеспечивающей поддержку исследований, выполняемых в рамках междисциплинарных исследований, связанных с обработкой и анализом пространственно распределенных данных. Создаваемая в Институте вычислительных технологий СО РАН система взаимодействует с вычислительным кластером с пиковой производительностью порядка 200 Гфлопс и системой хранения данных общим объемом памяти более 40 Тбайт. Обмен данными осуществляется через сеть передачи данных СО РАН, созданную и поддерживаемую ИВТ СО РАН. Сеть обслуживает более 150 научно-образовательных учреждений сибирского региона и насчитывает более 50 тысяч активных пользователей.

В настоящее время функциональные блоки объединяются в единую систему и производятся разработка графического интерфейса пользователя и подключение баз данных, созданных в ходе выполнения интеграционных проектов СО РАН.

## Список литературы

- [1] МАТЕРИАЛЫ Междунар. конф. “Космическая съемка — на пике высоких технологий”. Россия, Москва, 18–20 апреля 2007 г.  
<http://www.sovzondconference.ru/archive2007/rus/agenda.html>
- [2] The Open Geospatial Consortium, 2007. <http://www.opengeospatial.org>
- [3] Кудашев Е.Б., Филонов А.Н. Технологии и стандарты интеграции сервисов, каталогов и баз данных дистанционного исследования Земли из космоса // Тр. 9-й Всерос. научн. конф. “Электронные библиотеки: перспективные методы и технологии, электронные библиотеки” (RCDL’2007). Переславль-Залесский, 2007. С. 273–279.
- [4] Якубайлик О.Э., Попов В.Г. Построение ГИС-портала на платформе Mapguide и 1С-Битрикс // Тез. IX Всерос. конф. “Современные методы математического модели-

- рования природных и антропогенных катастроф". Барнаул, 17–22 сент. 2007 г. Барнаул: Изд-во Алт. гос. ун-та, 2007. С. 122.
- [5] ЖИЖИМОВ О.Л., ФЕДОТОВ А.М. Модели управления доступом к распределенным информационным ресурсам // Тр. 9-й Всерос. научн. конф. "Электронные библиотеки: перспективные методы и технологии, электронные библиотеки" (RCDL'2007). Переславль-Залесский, 2007. С. 296–299.
- [6] АДРИАНОВ В. Инфраструктура пространственных данных // ArcReview. 2006. № 2(37). С. 1–2.
- [7] АДРИАНОВ В. Тенденции развития ПО ГИС на примере продуктов ESRI // ArcReview. 2006. № 2(37). С. 2–4.
- [8] ФИРСОВ К.М., ФАЗЛИЕВ А.З., САКЕРИН С.М. и др. Информационно-вычислительная система "Атмосферная радиация". Современное состояние, перспективы развития // Тр. 9-й Всерос. научн. конф. "Электронные библиотеки: перспективные методы и технологии, электронные библиотеки" (RCDL'2007). Переславль-Залесский, 2007. С. 62–66.
- [9] ЛЕВИН В.А., АЛЕКСАНИН А.И., АЛЕКСАНИНА М.Г. Задачи и проблемы спутникового информационного обеспечения мониторинга океана и атмосферы на Дальнем Востоке // Вестн. ДВО РАН. 2007. № 4. С. 95–110.
- [10] СОЗДАНИЕ картографических сервисов с использованием ArcIMS.  
<http://gis-lab.info/qa/arcims.html>
- [11] OPENGIS Web Map Service Implementation Specification.  
[http://portal.opengeospatial.org/files/?artifact\\_id=5316](http://portal.opengeospatial.org/files/?artifact_id=5316)
- [12] ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Z39.50 Maintenance Agency Official Text for Z39.50. 1995.
- [13] ЖИЖИМОВ О.Л., МАЗОВ Н.А. Принципы построения распределенных информационных систем на основе протокола Z39.50. Новосибирск: ОИГГМ СО РАН; ИВТ СО РАН, 2004. 361 с.
- [14] CATALOGUE Interoperability Protocol (CIP) Specification — Release B 2.4. CEOS/WGISS/PTT. [http://www.dfd.dlr.de/ftp/pub/CIP\\_documents/cip2.4/S\\_cover.pdf](http://www.dfd.dlr.de/ftp/pub/CIP_documents/cip2.4/S_cover.pdf)
- [15] BEST C., HAYES L., NEBERT D. ET AL. Alignment of CIP and GEO profiles // Proc. of the Earth Observation & Geo-Spatial Web and Internet Workshop '98 (EOGEO'98). Salzburger Geographische Materialien. Vol. 27. Instituts fur Geographie der Universitdt Salzburg. 1998.
- [16] DOUGLAS D.N. Z39.50 Application Profile for Geospatial Metadata or "GEO" Version 2.2. U.S. Federal Geographic Data Committee.  
<http://www.blueangeltech.com/Standards/GeoProfile/geo22.htm>
- [17] CONTENT Standard for Digital Geospatial Metadata.  
<http://www.fgdc.gov/metadata/contstan.html>
- [18] КУЗНЕЦОВ Е.Б., БАЛАШОВ А.Д. Интеграция электронной библиотеки спутниковых данных в международную систему космической информации // Тр. 5-й Всерос. науч. конф. "Электронные библиотеки: перспективные методы и технологии, электронные коллекции" (RCDL-2003). Санкт-Петербург, 29–31 окт. 2003 г. С. 88–95.

*Поступила в редакцию 30 ноября 2007 г.*