

О распределённой инфраструктуре системы оперативного спутникового мониторинга ЦКП ДДЗ СО РАН*

Ю. И. Шокин¹, Н. Н. Добрецов^{1,2}, В. А. Кихтенко^{1,3},
В. В. Смирнов¹, Д. Л. Чубаров^{1,3}, Л. Б. Чубаров^{1,3}

¹Институт вычислительных технологий СО РАН, Новосибирск, Россия

²Институт геологии и минералогии им. В.С. Соболева СО РАН, Новосибирск, Россия

³Новосибирский государственный университет, Россия

e-mail: geosphere@ict.nsc.ru

Представлено современное состояние распределённой инфраструктуры системы оперативного спутникового мониторинга территории Сибири и Дальнего Востока. Активный архив спутниковых данных, а также его аппаратное и программное обеспечение, система взаимодействия с приёмными центрами составляют ядро информационно-вычислительной инфраструктуры Центра коллективного пользования данными дистанционного зондирования СО РАН.

Ключевые слова: дистанционное зондирование Земли.

Введение

За 50 лет, прошедших с момента запуска первых спутников дистанционного зондирования Земли, передаваемая ими информация изменила некоторые традиционные отрасли науки и стала основой для возникновения нескольких новых научных направлений. Благодаря передаваемым с орбиты данным о процессах, протекающих на Земле, удалось впервые создать прогностические модели климата, повысить точность прогноза опасных метеорологических явлений, построить глобальную систему мониторинга лесных пожаров [1]. На основе данных оперативного спутникового мониторинга функционируют информационные системы государственных служб, чья деятельность связана с учётом и сохранением природных ресурсов [2].

В настоящее время на орбите находится свыше сотни спутников, оснащённых аппаратурой дистанционного зондирования Земли. В планах только Российского космического агентства запуск около 20 космических аппаратов (КА) дистанционного зондирования до 2020 г. Объём данных, передаваемых на Землю всеми этими аппаратами, продолжает расти со скоростью в несколько терабайт за сутки [3, 4].

Задача обработки и анализа такого объёма данных требует создания соответствующей инфраструктуры для приёма, обработки и архивного хранения данных дистанционного зондирования и информационных продуктов, полученных на их основе. Работы

*Работа выполнена при поддержке Президентской программы “Ведущие научные школы РФ” (грант № НШ-6293.2012.9), Программы базовых фундаментальных исследований СО РАН (проект № IV.38.2.1), Программы интеграционных фундаментальных исследований СО РАН (проект № 131а), Программы Президиума РАН (проект № 4.10), РФФИ (гранты № 12-07-00545-а, № 13-07-12105-офи_м).

по созданию централизованного архива данных дистанционного зондирования Земли (ДДЗ) в Сибирском отделении Российской академии наук (СО РАН) были начаты в 2007 г. К этому времени работа с такими данными велась в институтах СО РАН независимо друг от друга. Первый централизованный архив ДДЗ в Новосибирске был создан на базе Института вычислительных технологий (ИВТ СО РАН).

Первоначальный план предполагал на основе имеющихся в архиве данных решение широкого круга задач, связанных с мониторингом как природных, так и социально-экономических процессов [5], построение распределённой системы каталогов для поиска файлов в каталоге по широкому перечню атрибутов [6, 7], а также интеграцию с аналогичными системами, развиваемыми в Дальневосточном отделении РАН [8]. В 2012 г. все созданные ресурсы были объединены в рамках Центра коллективного пользования данными дистанционного зондирования Земли (ЦКП ДДЗ СО РАН).

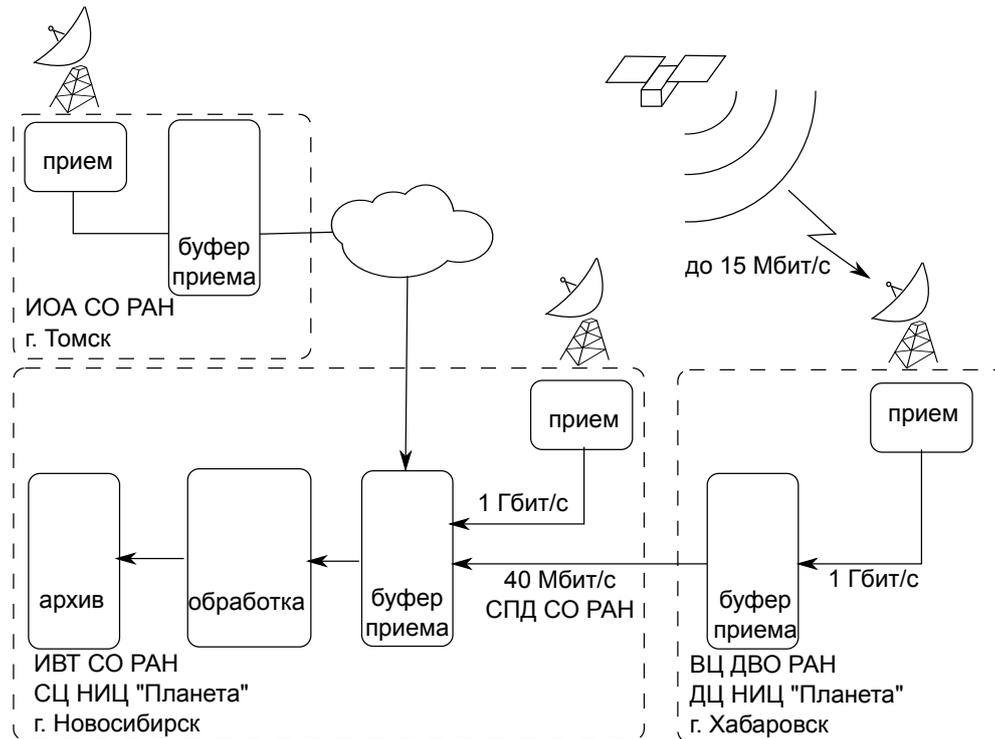
Сами по себе файлы, содержащие восстановленные в соответствии с известными методиками характеристики атмосферы и земной поверхности, могут служить лишь одним из источников информации для экспертного сообщества. Для вовлечения всего объёма накопленных данных в исследовательскую работу необходимы новые информационные системы, способные извлекать информацию из множества файлов каталога. Обеспечение необходимых условий для функционирования таких систем является одной из основных целей развития представленной в настоящей работе инфраструктуры.

1. Организация работы

Создание Центра коллективного пользования позволяет избежать выделения технических и организационных ресурсов, необходимых для поддержания постоянно пополняющегося в автономном режиме архива спутниковых данных, в каждом институте Сибирского отделения и вузе, где есть исследователи, заинтересованные в работе с архивом. В нынешней форме работа архива обеспечивается благодаря участию нескольких организаций.

Приём данных с ряда космических аппаратов, передающих их в режиме прямой трансляции, осуществляется в Сибирском и Дальневосточном центрах НИЦ “Планета”, входящих в структуру Российской службы по гидрометеорологии и мониторингу окружающей среды (Росгидромет), а также в Институте оптики атмосферы СО РАН (ИОА СО РАН, г. Томск). Данные, принятые в Дальневосточном центре НИЦ “Планета” (г. Хабаровск), передаются в Вычислительный центр Дальневосточного отделения РАН (ВЦ ДВО РАН). Оттуда по каналу Сети передачи данных СО РАН данные копируются в Новосибирск, где в ИВТ СО РАН осуществляется их обработка и хранение (см. рисунок).

Организации, поставляющие исходные данные для архива ЦКП, поддерживают также свои собственные системы обработки и архивации спутниковых данных, ориентированные, как правило, на свой круг решаемых задач. Так, ВЦ ДВО РАН совместно с Институтом космических исследований РАН и НИЦ “Планета” реализует проект по созданию информационной системы для работы со спутниковыми данными по территории Дальнего Востока [9, 10]. В ИОА СО РАН поддерживается региональная база данных информационных продуктов, произведённых на основе спутниковых данных, и программное обеспечение для их обработки [11]. Одновременно с использованием данных в своих системах эти организации передают часть поступающих к ним данных для обработки в архив ЦКП сразу после их получения с приёмных комплексов.



Распределённая инфраструктура приёма, обработки и хранения данных

Поступающие в ЦКП исходные данные и продукты их обработки в короткий срок публикуются в открытом каталоге, доступ к которому осуществляется по протоколу FTP. В тех случаях, когда доступ к данным ограничен лицензионными соглашениями, действует система заказов, позволяющая авторизованным пользователям получать необходимые продукты [8].

Возможности ЦКП не должны ограничиваться только предоставлением доступа к открытому каталогу генерируемых продуктов. Перспективным направлением развития является предоставление исследователям возможности выполнять собственные алгоритмы обработки над всем объёмом данных, имеющихся в архиве.

2. Аппаратное обеспечение

Основными компонентами наземных систем для работы со спутниковыми данными являются комплекс приёма и декодирования данных, передаваемых с космического аппарата, вычислительная система, позволяющая осуществлять стандартные этапы обработки данных, такие как радиометрическая и геометрическая коррекция, уточнение географической привязки данных, коррекция данных с учётом модели рассеяния излучения в атмосфере, а также генерацию тематических информационных продуктов, основанных на взаимосвязи между соотношениями яркости отражённого излучения в различных спектральных каналах и физическими свойствами отражающей поверхности.

В настоящее время в ЦКП поступают данные из трёх центров, оснащённых необходимым оборудованием для приёма данных с большинства находящихся на орбите КА. С ростом их числа чаще возникают ситуации, когда сеансы связи накладываются во времени. В таких случаях сбор данных из нескольких центров приёма, расположенных

в относительной близости друг от друга, позволяет согласовать расписание приёма так, чтобы вся необходимая информация могла быть собрана без потерь и дублирования. Эта задача могла бы быть решена за счёт создания специализированного протокола автоматизированного обмена информацией о принимаемых данных между станциями приёма. Работа по внедрению такого протокола очень перспективна, но потребует межведомственных согласований.

Для передачи данных в ЦКП используются каналы связи сетей передачи данных Сибирского и Дальневосточного отделений Российской академии наук [12]. Так, данные, принятые в Хабаровске, через три-пять минут после окончания сеанса связи становятся доступны в Новосибирске благодаря наличию арендованного канала связи пропускной способностью 40 Мбит/с. При передаче данных между Томском и Новосибирском маршрутизация осуществляется сторонними операторами связи.

Современные спутниковые сенсоры порождают большой объём информации, которая передаётся на Землю по стандартным радиоканалам со скоростью до 15 Мбит/с. За один сеанс связи продолжительностью около 15 мин может быть передано около 1 ГБ. Ещё десять лет назад передача такого объёма данных по наземным каналам связи была трудноосуществима. В настоящее время благодаря повсеместному развитию Интернет стоимость каналов связи значительно снизилась, что сделало передачу таких объёмов данных возможной. Ускоренное развитие технологий “наземной” передачи информации позволяет пересмотреть концепцию архитектуры пространственно распределённой сети приёмных станций, популярную ещё сравнительно недавно в России.

Следующим элементом инфраструктуры ЦКП являются системы хранения данных. Со времени создания архива в 2007 г. сменилось уже три поколения этих систем. На первом этапе использовались дисковые массивы, присоединяемые по протоколу iSCSI непосредственно к серверам обработки. Затем была введена в строй система хранения на основе сети хранения данных с выделенными серверами, экспортирующими дисковое пространство как сетевую файловую систему NFS. Наконец, на третьем этапе был осуществлён переход к использованию масштабируемой сетевой файловой системы на основе коммерческой распределённой системы OneFS. В настоящее время объём накопленных в архиве данных составляет около 129 ТБ. Общее число файлов в архиве — около 33 млн. Объём данных прирастает в среднем со скоростью 80 ГБ в день.

3. Архив спутниковых данных

Архив ЦКП ДДЗ СО РАН содержит информационные продукты, получаемые на основе данных, поступающих от пяти спутниковых систем: Landsat, SPOT, Terra, Aqua и NPP. Данные в архиве собираются для всей территории Сибири.

Спутники серии Landsat обеспечивают непрерывный мониторинг Земли на протяжении 40 лет. В настоящее время на орбите находится восьмой космический аппарат этой серии — Landsat 8, обеспечивающий съёмку в 11 спектральных диапазонах видимого и инфракрасного излучения с разрешением 30 м в видимой части спектра и 100 м в инфракрасной. Данные поступают из архива геологической службы США после геометрической коррекции и интерполяции на единую сетку с разрешением 30 м. Данные Landsat, относящиеся как к архивным, так и к активным КА, распространяются свободно.

Серия спутников SPOT администрировалась компанией SPOT Image, созданной Космическим агентством Франции. Эти спутники передавали данные среднего разреше-

Объёмы данных в разделах архива ЦКП ДДЗ СО РАН по состоянию на конец 2013 г.

Источник данных	Объём, ГБ	Примечание
Landsat	4 500	Пополняющийся архив Landsat (Landsat 5, 7, 8)
SPOT	24 810	Архив данных со спутников SPOT 2/4 за 2008–2012 гг.
Terra	41 900	Архив информационных продуктов, полученных на основе данных, поступающих со спектрорадиометра MODIS на КА Terra за 2011–2013 гг.
Aqua	16 820	Архив исходных данных, поступающих со спектрорадиометра MODIS и инфракрасного сенсора AIRS на КА Aqua, и информационных продуктов, полученных на их основе за 2012–2013 гг.
NPP	18 700	Архив исходных данных, поступающих с КА NPP “Suomi”, и информационных продуктов, полученных на их основе
Другие источники	1 040	Данные реанализа метеорологических наблюдений и численного прогноза погоды, определенные на сетке с разрешением 1°, и данные, поступающие с автоматических метеостанций СО РАН
Прочее	24 330	Оперативные данные, данные в форматах, оптимизированных для визуализации, продукты альтернативных веток обработки и другие материалы

Итого 129 ТБ

ния, используемые в первую очередь для исследований почв и растительного покрова. К данным SPOT предоставляется ограниченный доступ в соответствии с требованиями лицензионного соглашения¹. До окончания срока службы КА в декабре 2012 г. данные поступали с приёмного комплекса Сибирского центра НИЦ “Планета”.

Большая часть объёма архива относится к данным, поступающим ежедневно с сенсоров MODIS, установленных на спутниках Terra и Aqua, а также с сенсора VIIRS, установленного на спутнике NPP “Suomi”. Данные поступают с приёмных комплексов в течение нескольких минут после окончания сеанса приёма и передаются на обработку. Основной массив информационных продуктов становится доступным через 10–15 мин по окончании обработки. В дальнейшем по заранее определённому расписанию генерируются информационные продукты, использующие все данные за один или несколько дней.

Также в архиве содержатся данные о состоянии атмосферы, полученные в результате численного моделирования и реанализа метеорологических наблюдений, которые публикуются метеорологическими службами в рамках протоколов Всемирной службы погоды, и непосредственно данные, поступающие с сети наземных наблюдений СО РАН. Структура архива и распределение объёмов данных приведены в таблице.

4. Система оперативной обработки

Оригинальная система оперативной обработки спутниковых данных, созданная и эксплуатируемая в ЦКП ДДЗ СО РАН [13], — это ключевой компонент инфраструктуры. Система предназначена для получения широкого ассортимента информационных

¹Данные Spot Image® используются по соглашению с компанией Spot Image.

продуктов. Основной объём продуктов генерируется на основе данных, поступающих со спектрорадиометра MODIS² [14]. Эта система является адаптацией методик и программного обеспечения, разрабатываемых в рамках поддержки миссий Terra и Aqua (Product Generation Executable, PGE). Доступ к исходным текстам таких пакетов предоставляется по запросу службой обработки данных MODIS в Национальном управлении по авионавигации и исследованию космического пространства (NASA). Каждый пакет PGE в качестве входных данных получает набор файлов, содержащих информацию о состоянии атмосферы, рельефе, а также ряд других продуктов MODIS, и создает новый продукт, имеющий значения определяемых параметров и оценки достигнутой точности.

Запуск PGE осуществляется по мере готовности входных наборов данных. Зависимости между PGE задают потоки обработки, предполагающие запуск нескольких десятков PGE для получения одного продукта. Для синхронизации исполнения таких потоков был привлечен диспетчер вычислительных процессов Taverna [15], дополненный необходимыми функциями. Использование диспетчера вычислительных процессов позволило реализовать обработку данных на вычислительном кластере с общей файловой системой. Тем самым значительно сократилось время генерации продуктов, повысились устойчивость и эластичность системы. В перечне ежедневно производимых информационных продуктов: альbedo поверхности, характеристики атмосферы, растительного покрова, карты снежного покрова и другая оперативная информация, представленная на сайте ЦКП³.

5. Перспективы развития информационных систем на основе инфраструктуры ЦКП

В целом инфраструктура хранения и обработки ЦКП ДДЗ СО РАН представляет собой основу для построения новых информационных систем, использующих извлекаемые из архива данные для решения междисциплинарных фундаментальных и прикладных задач мониторинга природной среды и социально-экономических процессов. Такого рода информационные системы берут на себя, в первую очередь, решение задач извлечения временных рядов с ограничениями, зависящими как от метаданных, так и от самих данных, а также выполнения алгоритмов анализа данных непосредственно в разделах архива. Можно представить себе несколько примеров подобных задач. Например, с помощью имеющихся в архиве данных можно определить, как долго заданная территория находилась под снежным покровом за всю историю наблюдений. В случае труднодоступных территорий решение такой задачи без использования спутниковых данных невозможно.

В качестве другого примера можно привести задачу выбора фрагментов из архива данных среднего разрешения, попадающих на территорию, где были зарегистрированы признаки лесного пожара или наводнения. Собственно сопоставление годовой динамики спектральной яркости или их комбинаций за несколько лет может дать новую информацию об изменениях, происходящих на определённой территории, которые невозможно было бы выявить на отдельных снимках.

²MODIS — MODerate resolution Imaging Spectroradiometer.

³Адрес сайта ЦКП ДДЗ СО РАН <http://sdc.ict.nsc.ru>

6. Сопоставление с аналогами

Инфраструктуры хранения и обработки спутниковых данных существуют в различных масштабах. Пожалуй, наиболее известной и активно развивающейся структурой является подсистема хранения и обработки данных в Системе наблюдения Земли (Earth Observing System Data and Information System — EOSDIS) [4]. Инфраструктура EOSDIS формировалась на протяжении многих лет в тесном контакте с пользователями. Необходимость развития региональной информационной инфраструктуры, примером которой является ЦКП ДДЗ СО РАН, обусловлена потребностью в максимально быстром получении информационных продуктов для определённой территории, а также необходимостью интеграции данных, поступающих с сети наземных наблюдений и перспективных КА Российского космического агентства.

Система хранения и обработки данных для Российского космического агентства в настоящее время создаётся на основе Национального центра оперативного мониторинга Земли. Эта система предназначена в первую очередь для обеспечения данными административных служб [16]. Важно отметить, что задачи предоставления результатов спутникового мониторинга коммерческим потребителям и государственным службам во многом отличаются от научно-исследовательских задач.

Для решения проблемы предоставления доступа к данным дистанционного зондирования для научно-исследовательских задач на протяжении многих лет развивается система обработки и хранения данных в Институте автоматизации и проблем управления Дальневосточного отделения Российской академии наук (ИАПУ ДВО РАН) [17]. В рамках этой работы решены задачи привлечения внешних вычислительных ресурсов для обработки данных. Система направлена на генерацию информационных продуктов по запросу и не предполагает хранения всего объёма информационных продуктов на дисковом архиве. Такой подход не даёт возможности эффективно извлекать ряды значений вычисляемых величин за продолжительные временные промежутки.

Заключение

На протяжении шести лет совместными усилиями институтов СО РАН, вузов сибирского региона и центров НИЦ “Планета” создаётся архив данных дистанционного зондирования Земли. В настоящее время объём этого архива превышает 100 ТБ и продолжает пополняться со скоростью около 80 ГБ в день. На основе такого активного дискового архива могут выполняться различные количественные исследования изменений природной среды. Архив, а также его аппаратное и программное обеспечение, система взаимодействия с центрами приёма данных составляют ядро информационно-вычислительной инфраструктуры Центра коллективного пользования данными дистанционного зондирования СО РАН.

Список литературы

- [1] MICHAEL K., MURPHY K., LOWE D. ET AL. Implementation of the land, atmosphere near real-time capability for EOS (LANCE) // IGARSS-2010. Honolulu, 2010. P. 1442–1445.
- [2] АНДРЕЕВ М.В., ГАЛЕЕВ А.А., ЕФРЕМОВ В.Ю. и др. Построение автоматизированных систем сбора, хранения, обработки и представления спутниковых данных для решения задач мониторинга окружающей среды // Солн.-земн. физика. 2004. Т. 5. С. 8–11.

- [3] Лупян Е.А., САВОРСКИЙ В.П., ШОКИН Ю.И. и др. Современные подходы и технологии организации работы с данными дистанционного зондирования Земли для решения научных задач // Совр. пробл. дистанц. зондирования Земли из космоса. 2012. № 5. С. 21–44.
- [4] RAMAPRIYAN H.K., BENNKE J., SOFINOWSKI E. ET AL. Evolution of the earth observing system (EOS) data and information system (EOSDIS) // Standard-Based Data and Information Systems for Earth Observation / Eds Di Liping, H.K. Ramapriyan. 2010. P. 63–92.
- [5] ДОБРЕЦОВ Н.Н., ПОТАТУРКИН О.И., ЧУВАРОВ Л.Б., ШОКИН Ю.И. О проекте распределённой информационно-вычислительной системы сбора, хранения и обработки данных дистанционного зондирования Земли для регионов Сибири и Дальнего Востока // Вестн. КазНУ им. аль-Фараби. Сер. Математика, механика, информатика. 2008. № 3 (58), ч. 1. С. 492–496.
- [6] ЯКУБАЙЛИК О.Э. Формирование программно-технологического обеспечения системы мониторинга состояния природной среды и ресурсов // Актуальные вопр. совр. техники и технол.: Матер. Междунар. конф. Липецк, 2013. С. 51–55.
- [7] МАТВЕЕВ А.Г., ЯКУБАЙЛИК О.Э. Система управления контентом геоинформационного веб-портала “Геофреймворк” // Совр. проблемы науки и образования. 2013. (в печати)
- [8] ШОКИН Ю.И., ПЕСТУНОВ И.А., СМИРНОВ В.В. и др. Распределённая информационная система сбора, хранения и обработки спутниковых данных для мониторинга территорий Сибири и Дальнего Востока // Журн. Сиб. федер. ун-та. Сер. Техника и технол. 2008. Т. 1, № 4. С. 291–314.
- [9] Лупян Е.А., СОРОКИН А.А., КРАМАРЕВА Л.С., ГИРИНА О.А. Разработка информационной системы для работы с данными дистанционного спутникового мониторинга Дальнего Востока России с целью обеспечения научной и образовательной деятельности в области исследования и контроля состояния окружающей среды и опасных природных явлений // Информ. технологии и высокопроизвод. вычисления: Матер. Междунар. конф. Хабаровск, 2011. С. 213–217.
- [10] СМАГИН С.И., Лупян Е.А., СОРОКИН А.А. и др. Информационная система работы с данными спутниковых наблюдений региона Дальнего Востока России для проведения научных исследований в различных областях знаний // Совр. пробл. дистанц. зондирования Земли из космоса. 2013. Т. 10, № 1. С. 277–291.
- [11] AFONIN S.V., BELOV V.V., ENGEL M.V., КОКН А.М. Development of a database of regional satellite information and software for processing of this information in the Institute of atmospheric optics SB RAS // Proc. SPIE, 11th Intern. Symp. on Atmos. and Ocean Optics, Atmos. Phys. Tomsk, 2004. P. 267–276.
- [12] ШОКИН Ю.И. Построение системы регионального спутникового мониторинга на основе корпоративной сети передачи данных СО РАН // Междунар. конф. вычисл. и информ. технологий в науке, технике и образовании. Усть-Каменогорск, 2013. Ч. 2. С. 286–292.
- [13] ШОКИН Ю.И., АНТОНОВ В.Н., ДОБРЕЦОВ Н.Н. и др. Распределённая система приёма и обработки спутниковых данных Сибири и Дальнего Востока. Текущее состояние и перспективы развития // Совр. пробл. дистанц. зондирования Земли из космоса. 2012. Т. 9, № 5. С. 45–54.
- [14] ЛАГУТИН А.А., НИКУЛИН Ю.А., ЖУКОВ А.П. и др. Математические технологии оперативного регионального спутникового мониторинга характеристик атмосферы и подстилающей поверхности. Ч. 1. MODIS // Вычисл. технологии. 2009. Т. 12, № 2. С. 67–89.

- [15] HULL D., WOLSFENCROFT K., STEVENS R. ET AL. Taverna: a tool for building and running workflows of services // *Nucleic Acids Research*. 2006. Vol. 34. P. W729–W732.
- [16] БУРЦЕВ М.А., ЕМЕЛЬЯНОВ К.С., ЕФРЕМОВ В.Ю. и др. Построение информационной системы удалённой работы с каталогами данных НЦ ОМЗ // *Совр. пробл. дистанц. зондирования Земли из космоса*. 2010. Т. 7, № 4. С. 64–71.
- [17] НЕДОЛУЖКО И.В., БАВЯК П.В., ТАРАСОВ Г.В., ЕРЁМЕНКО В.С. Инфраструктура приёма, распределённой обработки и поставки спутниковых данных в центре коллективного пользования регионального спутникового мониторинга ДВО РАН // *Совр. пробл. дистанц. зондирования Земли из космоса*. 2012. Т. 9, № 3. С. 324–331.

Поступила в редакцию 29 ноября 2013 г.