

# REMOTE ANALYSIS AND PROCESSING OF REMOTE SENSING DATA PROVIDED BY BIG DISTRIBUTED ARCHIVING SYSTEMS

*Alexandr V. Kashnitskiy, Evgeny A. Loupian*

Space Research Institute, Russian Academy of Sciences, Moscow, Russia

## **Abstract**

This work is about approaches to working with very big remote sensing data archives that were established in the Space Research Institute of the Russian Academy of Sciences. These approaches increase the efficiency of work with satellites data and provide analytical tools to the remote users. The work describes the developed technology for creating tools for analyzing and processing remote sensing information, the used approaches to build up the very big archives of satellite data. Also the work describes several examples to use developed approaches in active remote monitoring informational systems.

*Keywords: remote sensing, satellite data, information systems, satellite data processing, remote analysis of satellite data, web interfaces, ERS information archives, very big archives, data processing technologies*

## УДАЛЕННЫЙ АНАЛИЗ И ОБРАБОТКА ДАННЫХ ДЗЗ, ПРЕДОСТАВЛЯЕМЫХ СВЕРХБОЛЬШИМИ РАСПРЕДЕЛЕННЫМИ СИСТЕМАМИ АРХИВАЦИИ

*Кашицкий А.В., Лупян Е.А.*

Институт космических исследований РАН, Москва

Доклад посвящен созданным в ИКИ РАН подходам к работе с информацией ДЗЗ из сверхбольших распределенных постоянно пополняющихся архивов. Эти подходы повышают эффективность работы со спутниковыми данными и предоставляют удаленным исследователям развитые средства их обработки и анализа. В докладе рассказано о разработанной технологии создания инструментов анализа и обработки информации ДЗЗ, применяемых подходах к построению архивов спутниковых данных, а также приведены примеры использования этих наработок в действующих информационных системах.

*Ключевые слова:* дистанционное зондирование, спутниковые данные, информационные системы, обработка спутниковых данных, удаленный анализ спутниковых данных, web-интерфейсы, архивы информации ДЗЗ, сверхбольшие архивы, технологии работы с данными.

**Введение.** Постоянно возрастающая потребность в данных космической съемки и развитие технологий в последнее десятилетие привели к появлению большого количества высококачественных спутников дистанционного зондирования Земли (ДЗЗ). При этом в последние годы объемы получаемой с них информации увеличиваются намного быстрее возможностей по ее усвоению, обработке и анализу с применением традиционных подходов к работе со спутниковыми данными. Все это требует создания новых информационных технологий и подходов к организации работы с данными дистанционного зондирования. “Традиционные” подходы предполагают сначала выбор необходимых наборов данных из архивов, создание их локальных копий у конкретного пользователя (исследователя), а потом проведение их обработки и анализа. Как показано в работах (Лупян, Мазуров и др., 2011), (Лупян, Саворский и др., 2012), (Лупян, Балашов и др., 2015) такой путь в условиях быстро развивающихся спутниковых систем наблюдения становится все менее эффективным, а в перспективе и нереализуемым, из-за необходимости использования в различных проектах значительных вычислительных и коммуникационных ресурсов. Поэтому в последние годы стали создаваться и развиваться системы, предоставляющие возможности доступа к архивам спутниковых данных и результатам их обработки одновременно со средствами, обеспечивающими проведение их обработки и анализа (Лупян, Саворский и др., 2012). Данные в таких системах могут находиться в территориально распределенных архивах, а их обработка и анализ, которыми управляет пользователь, обычно производится на вычислительных средствах, расположенных в тех же центрах, в которых физически осуществляется хранение данных. Это позволяет избежать передачи больших объемов информации и обеспечивает возможность эффективной работы со сверхбольшими объемами данных. С развитием информационных технологий в последние годы стало возможным создание web-интерфейсов, обеспечивающих не только доступ к данным и их визуализацию, но и управление процессами их обработки. Во многих случаях такие системы по функциональности могут не уступать настольным приложениям, ориентированным на обработку спутниковой информации, обеспечивая при этом возможность непосредственной работы с огромными распределенными массивами информации с любого компьютера без использования сложных локальных приложений, требующих значительных вычислительных ресурсов и мощностей, обеспечивающих хранение и работу с данными. Настоящий доклад посвящен разработке методов и подходов, с помощью которых возможна реализация подобных систем. Аналогичные работы сейчас выполняются во многих организациях, как в России (Шокин и др., 2012), (Шокин и др., 2013) (Шокин и др., 2015), (Левин и др., 2008), (Бабяк и др., 2012), (Недолужко и др., 2012), так и за рубежом (Moore, Hansen, 2011), (Gorelick, 2013), (Acker J. G., Leptoukh G., 2007), (Koubarakis M. et al., 2012). Разработка предлагаемых методов и под-

ходов распределенной работы со сверхбольшими архивами данных дистанционных наблюдений Земли позволит существенно расширить возможности использования технологий дистанционного зондирования для решения различных научных и прикладных задач.

**Архивы данных, поддерживаемые в ИКИ РАН.** В различных центрах и системах хранения на сегодня уже накоплены гигантские объемы спутниковых данных. Любая вновь создаваемая технология, связанная с обработкой этих данных, должна учитывать существующее состояние дел и особенности такого хранения. В частности, подходы, разработанные в данной работе, в первую очередь ориентированы на существующую инфраструктуру центров и систем хранения, поддерживаемых в ИКИ РАН, в том числе архивы спутниковых данных, накопленных в центре коллективного пользования "ИКИ-Мониторинг" (далее ЦКП "ИКИ-Мониторинг") (Лупян, Прошин и др., 2015), Объединенной системе архивов данных центров НИЦ "Планета" (Бурцев и др., 2012), (Лупян, Милехин и др., 2014), а также архивы данных, поддерживаемых в рамках отдельных проектов и систем дистанционного мониторинга, например, Информационной системы дистанционного мониторинга лесных пожаров Федерального агентства лесного хозяйства "ИСДМ-Рослесхоз" (Барталев С. А. и др., 2010), системы мониторинга активности вулканов Камчатки и Курил "VolSatView" (Гордеев и др., 2016) и других. В этих архивах имеется уникальное для отечественных ресурсов пространственное и временное покрытие спутниковой информацией. Суммарно область покрытия включает всю территорию Северной Евразии и ряд локальных областей в Америке, Африке и Азии. Важным моментом является наличие долговременных однородных рядов данных за последние десятилетия по сегодняшний день, например, рядов снимков группировки LANDSAT, начиная с 1984 года, информации с прибора MODIS с 2000 года и так далее. Имеются радарные данные, гиперспектральные данные, данные оптических систем самого разного назначения от полученных с геостационарных спутников и аппаратов метеорологического назначения до снимков высокого разрешения. Общий объем архивов уже сегодня заметно превышает один петабайт, причем общий объем продолжает стремительно увеличиваться. Это увеличение можно оценить на момент 2017 года в цифрах порядка 1 Тб в сутки. Информация из описываемых архивов высоко востребована как при решении научных задач различной направленности, так и в разнообразных коммерческих проектах. Например, в работе (Лупян, Прошин и др., 2015) указывается, что информация из этих архивов используется более чем в двадцати научных организациях и группах. Имеющиеся архивы данных являются распределенными с центрами хранения, физически находящимися в разных городах, от центра во Франкфурте в Германии до центра в Петропавловске-Камчатском в России. Таким образом, архивы данных ЦКП "ИКИ-Мониторинг" и объединенной системы архивов центров НИЦ "Планета" выступают в качестве хорошей площадки для создания и отработки новых технологий работы со спутниковыми данными из сверхбольших распределенных архивов.

**Технология создания инструментов для удаленного анализа сверхбольших объемов данных дистанционного зондирования Земли.** В докладе рассказывается про разработанную архитектуру системы удаленного анализа и обработки спутниковых данных дистанционного зондирования Земли, предоставляемых большими распределенными системами архивации. Основные ее особенности были опубликованы в работах (Кашницкий, Лупян и др. 2016), (Кашницкий, Балашов и др., 2015). На базе такой архитектуры возможно создание различных инструментов анализа и обработки данных ДЗЗ из больших распределенных архивов. Таким образом, авторами была разработана технология создания инструментов для удаленного анализа сверхбольших объемов данных дистанционного зондирования Земли. Созданная технология обеспечивает быструю и удобную обработку и анализ спутниковых данных из сверхбольших распределенных архивов и позволяет создавать "on-line" инструменты. В соответствии с данным подходом удаленные пользователи управляют процедурами обработки через web-браузер, без установки сложного программного обеспечения. При работе с такими инструментами важно отсутствие жесткой привязки к месту и времени работы, так как создаваемые на базе технологии инструменты позволяют проводить все операции по обработке и анализу на вычислительных ресурсах самих центров хранения. Все это снимает требования к

наличию у исследователей в конкретном проекте каких-либо ресурсов и позволяет им получить максимально мощные средства для работы со спутниковыми данными только с использования обычного компьютера. Также обеспечивается возможность реализовывать инструменты максимально простыми, предъявляющими минимальные требования к знаниям и навыкам пользователей в области обработки спутниковых данных и геоинформационных технологий. То есть, при их использовании, прежде всего, необходимо понимание задачи и наличие знаний в предметной области. Технология учитывает инфраструктуру и особенности существующих центров и систем хранения, в частности, поддерживаемых в ИКИ РАН центра коллективного пользования "ИКИ-Мониторинг", Объединенной системы архивов данных центров НИЦ "Планета", а также архивы данных, поддерживаемых в ИКИ РАН в рамках отдельных проектов и систем дистанционного мониторинга. Разработанная технология и инструменты на ее основе удалось легко внедрить в различные действующие информационные системы, поддерживаемые в ИКИ РАН. В докладе также приводятся примеры созданных инструментов обработки данных.

**Подходы к построению сверхбольших распределенных архивов спутниковых данных.** В докладе также рассказывается, как для обеспечения возможности работы инструментов удаленного анализа в соответствии с предложенной архитектурой необходимо организовать хранение информации ДЗЗ. Реализация предложенных в данном разделе подходов применительно к описанным выше архивам, поддерживаемым в ИКИ РАН, была описана в статье (Прошин и др., 2016). Разработанный подход к построению системы ведения архивов спутниковых данных заключается в разделении базы данных на часть, хранящую сами данные, и часть, хранящую всю вспомогательную информацию. В докладе описываются связанные с таким подходом блоки получения полной информации, необходимой для унифицированного проведения процедур обработки и анализа. Это позволяет однотипно использовать в процедурах обработки совершенно разнородные данные из архивов. Также в докладе рассказано о получаемой на основе предложенного подхода возможности построения виртуальных продуктов в момент их запроса. Под термином виртуальный продукт подразумевается тематический информационный продукт, являющийся результатом преобразования спутниковых данных, и получаемый только в момент его запроса "на лету" программными средствами путем комбинирования, вычисления или применения различных правил преобразования физически хранящихся данных. Такое получение продуктов "на лету" имеет целый ряд преимуществ, в частности, позволяет легко описывать и модифицировать предоставляемые исследователям виртуальные информационные продукты на основе спутниковых данных без проведения предварительных сложных процедур обработки и переобработки архива. Также существенным плюсом является то, что предложенный подход позволяет избежать хранения в архивах всех требуемых производных продуктов и тем самым существенно уменьшить объем физически хранящихся данных возможно создание виртуальных продуктов, требующих проведение различных вычислений, в том числе над разновременными и разнородными данными. В докладе приводятся примеры различных виртуальных информационных продуктов и схем их формирования.

**Заключение.** Изложенные подходы и инструменты высоко востребованы при изучении явлений и процессов, происходящих в самых разных областях жизнедеятельности. Примеры их применения были описаны в большом количестве работ, например (Кашницкий, Лупян и др., 2015), (Рыбалко и др., 2016), (Саворский и др., 2016), (Гордеев и др., 2016). В докладе приводятся несколько примеров практического применения разработанных подходов при построении блоков обработки и анализа данных в информационных системах дистанционного мониторинга различного назначения. Примеры демонстрируют использование созданных инструментов анализа данных для изучения динамики растительного покрова, мониторинга загрязнений и опасных природных явлений, в том числе изучения вулканической активности.

*Работа выполнена при финансовой поддержке Минобрнауки России (контракт 14.607.21.0122, уникальный идентификатор ПНИЭР RFMEFI60715X0122).*

## ЛИТЕРАТУРА

- [1] Кашницкий А.В., Балашов И.В., Лупян Е.А., Толпин В.А., Уваров И.А. Создание инструментов для удаленной обработки спутниковых данных в современных информационных системах // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 1. С. 156-170.
- [2] Кашницкий А.В., Лупян Е.А., Балашов И.В., Константинова А.М. Технология создания инструментов обработки и анализа данных сверхбольших распределенных спутниковых архивов // Оптика атмосферы и океана. 2016. Т. 29. № 9. С. 772-777. DOI: 10.15372/AOO20160908.
- [3] Прошин А.А., Лупян Е.А., Балашов И.В., Кашницкий А.В., Бурцев М.А. Создание унифицированной системы ведения архивов спутниковых данных, предназначенной для построения современных систем дистанционного мониторинга // Современные проблемы дистанционного зондирования Земли из космоса. 2016. Т. 13. № 3. С. 9-27. DOI: 10.21046/2070-7401-2016-13-3-9-27.
- [4] Гордеев Е.И., Гирина О.А., Лупян Е.А., Сорокин А.А., Крамарева Л.С., Ефремов В.Ю., Кашницкий А.В., Уваров И.А., Бурцев М.А., Романова И.М., Мельников Д.В., Маневич А.Г., Королев С.П., Верхотуров А.Л. Информационная система VOLSATVIEW для решения задач мониторинга вулканической активности Камчатки и Курил // Вулканология и сейсмология. 2016. № 6. С. 1-16. DOI: 10.7868/S0203030616060043. Кашницким А.В. созданы инструменты обработки и анализа спутниковых данных, применявшиеся для дистанционного мониторинга вулканической активности.
- [5] Лупян Е.А., Прошин А.А., Бурцев М.А., Балашов И.В., Барталев С.А., Ефремов В.Ю., Кашницкий А.В., Мазуров А.А., Матвеев А.М., Суднева О.А., Сычугов И.Г., Толпин В.А., Уваров И.А. Центр коллективного пользования системами архивации, обработки и анализа спутниковых данных ИКИ РАН для решения задач изучения и мониторинга окружающей среды // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 5. С. 263-284.
- [6] Лупян Е.А., Балашов И.В., Бурцев М.А., Ефремов В.Ю., Кашницкий А.В., Кобец Д.А., Крашенинникова Ю.С., Мазуров А.А., Назиров Р.Р., Прошин А.А., Сычугов И.Г., Толпин В.А., Уваров И.А., Флитман Е.В. Создание технологий построения информационных систем дистанционного мониторинга // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 5. С. 53-75.
- [7] Бурцев М.А., Антонов В.Н., Ефремов В.Ю., Кашницкий А.В., Крамарева Л.С., Лупян Е.А., Мазуров А.А., Матвеев А.М., Милехин О.Е., Прошин А.А., Соловьев В.И. Система работы с распределенными архивами результатов обработки спутниковых данных центров приема НИЦ "Планета" // Современные проблемы дистанционного зондирования Земли из космоса. 2012. Т.9. № 5. С. 55-76. Кашницким А.В. разработаны подходы к организации сверхбольших распределенных архивов спутниковых данных, применяемые в архивах центров приема НИЦ "Планета".
- [8] Лупян Е.А., Милехин О.Е., Антонов В.Н., Крамарева Л.С., Бурцев М.А., Балашов И.В., Толпин В.А., Соловьев В.И. Система работы с объединенными информационными ресурсами, получаемыми на основе спутниковых данных в центрах НИЦ "ПЛАНЕТА" // Метеорология и гидрология, 2014. № 12. С.89-97.
- [9] Барталев С.А., Ершов Д.В., Коровин Г.Н., Котельников Р.В., Лупян Е.А., Щетинский В.Е. Основные возможности и структура информационной системы дистанционного мониторинга лесных пожаров Федерального агентства лесного хозяйства (ИСДМ Рослесхоз) // Современные проблемы дистанционного зондирования Земли из космоса, 2010. Т.7. № 2. С. 97-105.
- [10] Лупян Е.А., Мазуров А.А., Назиров Р.Р., Прошин А.А., Флитман Е.В., Крашенинникова Ю.С. Технологии построения информационных систем дистанционного мониторинга // Современные проблемы дистанционного зондирования Земли из космоса, 2011. Т.8. № 1. С.26-43.
- [11] Лупян Е.А., Саворский В.П., Шокин Ю.И., Алексанин А.И., Назиров Р.Р., Недолужко И.В., Панова О.Ю. Современные подходы и технологии организации работы с данными дистанционного зондирования Земли для решения научных задач // Современные проблемы дистанционного зондирования Земли из космоса, 2012. Т.9. № 5. С.21-44.
- [12] Шокин Ю.И., Антонов В.Н., Добрецов Н.Н., Кихтенко В.А., Лагутин А.А., Смирнов В.В., Чубаров Д.Л., Чубаров Л.Б. Распределенная система приема и обработки спутниковых данных Сибири и Дальнего Востока. Текущее состояние и перспективы развития // Современные проблемы дистанционного зондирования Земли из космоса, 2012. Т. 9. №. 5. С. 45-54.

- [13] Шокин Ю.И., Добрецов Н.Н., Кихтенко В.А., Смирнов В.В., Чубаров Д.Л., Чубаров Л.Б. О распределенной инфраструктуре системы оперативного спутникового мониторинга ЦКП ДДЗ СО РАН // Вычисл. технологии. 2013. Т. 18. С. 86–94.
- [14] Шокин Ю.И., Добрецов Н.Н., Мамаш Е.А., Кихтенко В.А., Воронина П.В., Смирнов В.В., Чубаров Д.Л. Информационная система приема, обработки и доступа к спутниковым данным и ее применение для решения задач мониторинга окружающей среды // Вычислительные технологии. 2015. Т. 20. № 5. С. 157-174.
- [15] Acker J. G., Leptoukh G. Online analysis enhances use of NASA earth science data // Eos, Transactions American Geophysical Union. – 2007. – Т. 88. – №. 2. – С. 14-17.
- [16] Недолужко И. В. и др. Инфраструктура приёма, распределённой обработки и поставки спутниковых данных в Центре коллективного пользования Регионального спутникового мониторинга ДВО РАН // Современные проблемы дистанционного зондирования Земли из космоса. – 2012. – Т. 9. – №. 3. – С. 324-331.
- [17] Бабяк П. В., Недолужко И. В. Подход к хранению и обработке данных в Центре коллективного пользования регионального спутникового мониторинга окружающей среды ДВО РАН // Материалы научной конференции "Интернет и современное общество". – 2012. – С. 16-22.
- [18] Левин В. А. и др. Состояние дел и перспективы развития ЦКП регионального спутникового мониторинга окружающей среды ДВО РАН в области современных информационных и телекоммуникационных технологий // Открытое образование. – 2008. – №. 4.
- [19] Moore R. T., Hansen M. C. Google Earth Engine: a new cloud-computing platform for global-scale earth observation data and analysis // AGU Fall Meeting Abstracts, 2011. Vol. 1. P.2.
- [20] Gorelick N. Google Earth Engine // EGU General Assembly Conference Abstracts. – 2013. – Т. 15. – С. 11997.
- [21] Koubarakis M. et al. TELEIOS: a database-powered virtual earth observatory // Proceedings of the VLDB Endowment. – 2012. – Т. 5. – №. 12. – С. 2010-2013.
- [22] Кашницкий А.В., Лупян Е.А., Барталев С.А., Барталев С.С., Балашов И.В., Ефремов В.Ю., Стыценко Ф.В. Оптимизация интерактивных процедур картографирования гарей в информационных системах дистанционного мониторинга природных пожаров // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 4. С. 7-16.
- [23] Рыбалко Е.А., Баранова Н.В., Лупян Е.А., Толпин В.А., Кашницкий А.В., Уваров И.А., Крашенинникова Ю.С., Иванченко В.И. Организация работы с данными наземных и дистанционных наблюдений для решения задач дистанционного мониторинга виноградников // Современные проблемы дистанционного зондирования Земли из космоса. 2016. Т. 13. № 1. С. 79-92. DOI: 10.21046/2070-7401-2016-13-1-79-92.
- [24] Саворский В.П., Кашницкий А.В., Константинова А.М., Балашов И.В., Крашенинникова Ю.С., Толпин В.А., Маклаков С.М., Савченко Е.В. Возможности анализа гиперспектральных индексов в информационных системах дистанционного мониторинга семейства «Созвездие-Вега» // Современные проблемы дистанционного зондирования Земли из космоса. 2016. Т. 13. № 3. С. 28-45. DOI: 10.21046/2070-7401-2016-13-3-28-45.
- [25] Гордеев Е.И., Гирина О.А., Лупян Е.А., Сорокин А.А., Мельников Д.В., Маневич А.Г., Романова И.М., Крамарева Л.С., Ефремов В.Ю., Кобец Д.А., Кашницкий А.В., Верхотуров А.Л., Бурцев М.А. ИС «VolSatView»: комплексный анализ данных об эксплозивных извержениях вулканов Камчатки // Вестник Дальневосточного отделения Российской академии наук. 2016. № 5. С. 397-410.