

УДК 528.8

Технология создания инструментов обработки и анализа данных свехбольших распределенных спутниковых архивов

А.В. Кашницкий, Е.А. Лупян, И.В. Балашов, А.М. Константинова*

*Институт космических исследований РАН
117997, г. Москва, ул. Профсоюзная, 84/32*

Поступила в редакцию 12.05.2016 г.

Значительное увеличение в последние годы объемов данных, поступающих от систем спутниковых наблюдений Земли, делает все более актуальной задачу разработки новых технологий, позволяющих организовать эффективную работу со свехбольшими, распределенными, постоянно пополняющимися архивами, обеспечивающими не только возможности эффективного поиска и выбора данных, но и их обработку. В настоящей статье рассмотрены возможности разработанных в Институте космических исследований РАН технологий, позволяющих создавать различные инструменты обработки и обеспечивающих использование для проведения анализа спутниковой информации распределенных вычислительных ресурсов центров обработки, архивации и представления данных дистанционного зондирования. Описаны преимущества и возможности предлагаемых подходов, приводятся примеры реализации инструментов для распределенной обработки данных, поступающих от различных спутниковых систем дистанционного зондирования. Приведенные примеры ориентированы на то, чтобы продемонстрировать возможности использования разработанных инструментов для решения задач анализа различных явлений, в том числе возникающих в атмосфере и на поверхности океана.

Ключевые слова: дистанционное зондирование, информационные системы, распределенная обработка данных, обработка спутниковых данных, свехбольшие архивы данных, технологии работы с данными, дистанционные методы наблюдений атмосферы и океана; remote sensing, information systems, distributed data processing, satellite data processing, very large data archives, data management technologies, ocean and atmosphere remote observation approaches.

Введение

Активное развитие технологий дистанционного зондирования привело к резкому увеличению в последние годы числа высококачественных систем спутникового наблюдения Земли. При этом объемы получаемой этими системами информации увеличиваются намного быстрее возможностей по ее усвоению, обработке и анализу с применением традиционных подходов к работе со спутниковыми данными. Это потребовало создания новых информационных технологий и подходов к организации работы с данными дистанционного зондирования. Традиционные подходы, используемые обычно в научных и прикладных проектах, предполагают сначала получение необходимых для решения конкретной задачи наборов данных (заказ съемки или выбор из архивов центров предоставления данных), создание их локальных копий у конкретного пользователя, а потом проведение их обработки и анализа. Как показано

в работах [1, 2], такой путь в условиях быстро растущей информативности спутниковых систем наблюдения становится все менее эффективным, а в перспективе и нереализуемым.

В последние годы стали создаваться и развиваться системы, предоставляющие возможности доступа к архивам спутниковых данных одновременно со средствами, обеспечивающими проведение обработки и анализа [2, 3]. Данные в таких системах могут находиться в территориально распределенных архивах, а их обработка и анализ производится под управлением удаленных пользователей на вычислительных средствах, расположенных в центрах архивации данных. С развитием информационных технологий в последние годы стало возможным создание веб-интерфейсов, обеспечивающих не только доступ к данным и их визуализацию, но и управление процессами их обработки. Во многих случаях такие системы по функциональности не уступают 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. Создание

* Александр Витальевич Кашницкий (kashnizky@gmail.com); Евгений Аркадьевич Лупян (evgeny@d902.iki.rssi.ru); Иван Васильевич Балашов (ivbalashov@d902.iki.rssi.ru); Анна Михайловна Константинова (konstantinova.anouk@gmail.com).

таких систем дает возможность принципиально поменять подходы к работе со спутниковой информацией, существенно упростить доступ к ней и повысить эффективность использования. Это позволяет избежать необходимости обмена большими объемами информации, создания дополнительных дорогостоящих систем хранения и обработки данных и в конечном итоге расширить круг специалистов, которые могут использовать распределенные ресурсы для эффективной работы со сверхбольшими объемами данных.

В настоящей статье мы кратко опишем новые возможности технологии, разработанной в Институте космических исследований Российской академии наук (ИКИ РАН) и ориентированной на создание систем распределенной работы с данными дистанционного зондирования Земли (ДЗЗ). Эта технология позволяет создавать и интегрировать в различные информационные системы различные инструменты обработки и анализа спутниковой информации.

Отметим, что аналогичные работы сейчас выполняются во многих организациях. Однако, в отличие от подхода, предлагаемого в данной статье, в части обработки данных они в основном направлены на offline-обработку, проводимую через предварительный заказ. Из авторов, проводящих исследования в России, можно упомянуть работы ИВТ СО РАН [4], ДВО РАН [5]. За рубежом подобные подходы также активно развиваются, например, в системах EOSDIS [6], NASA GIOVANNI [7] и наиболее развитой по описываемым подходам системе GOOGLE EARTH ENGINE [8].

Особенности построения технологии распределенной обработки

Следует учитывать, что пользователи различных прикладных информационных систем, являясь специалистами в конкретных предметных областях, заинтересованы в использовании инструментов для получения и анализа данных, не требующих дополнительного сложного программирования. Для построения таких инструментов, ориентированных на работу со сверхбольшими распределенными архивами данных, в ИКИ РАН была разработана специальная технология [3]. Она базируется на разработках ИКИ РАН по созданию систем сбора, архивации, обработки и предоставления спутниковых данных, описанных в статьях [9–11]. Технология позволяет создавать в рамках картографических веб-интерфейсов инструменты обработки данных на вычислительных ресурсах, расположенных в центрах, обеспечивающих доступ к данным. Подобные инструменты дают возможность, выбрав необходимые наборы информации, сформировать задания на их обработку, а потом непосредственно в веб-интерфейсе получить и проанализировать результаты. На рис. 1 представлена схема выполнения задания на обработку и преобразования данных на разных этапах. В соответствии с технологией реализуются четыре основных блока, обеспечивающих характерные этапы работы с данными при проведении их обработки, а именно: выбор данных, управление и настройку операций обработки, выполнение операций

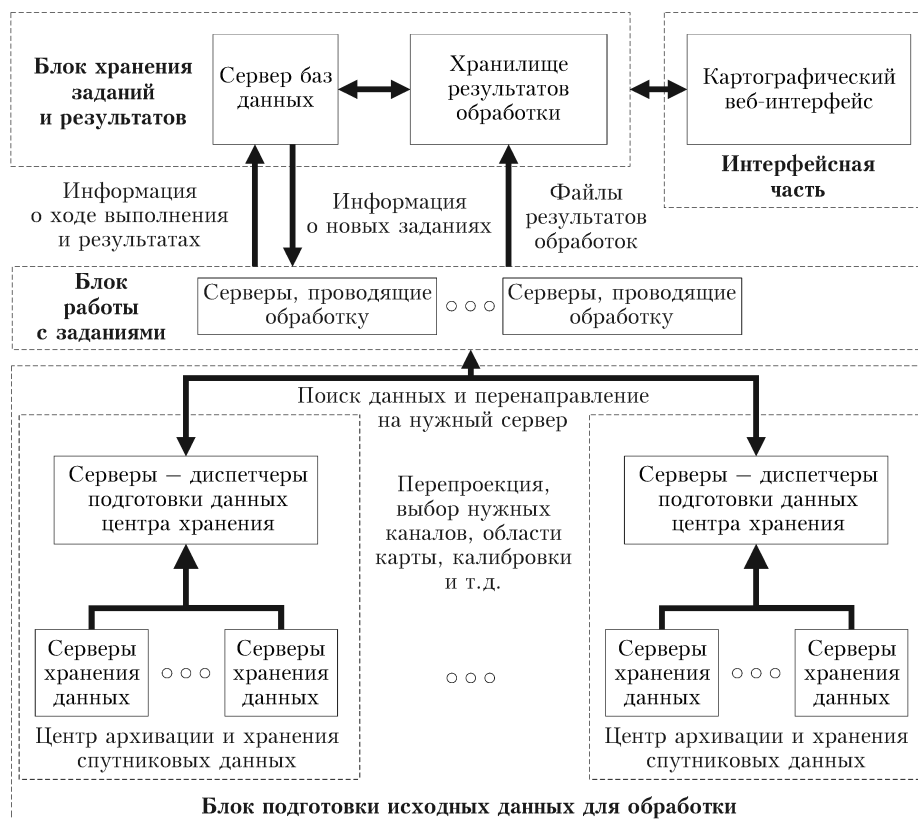


Рис. 1. Схема выполнения задания на обработку и преобразования данных на разных этапах

и проведение анализа результатов. Перечислим эти основные блоки, обеспечивающие работу технологии, и их основные функции.

- *Интерфейсная часть.* Обеспечивает возможность удаленного управления процедурами обработки (выбор данных, задание параметров, контроль исполнения и т.д.) и проведение анализа результатов. Для выбора наборов данных для обработки в интерфейсах используются стандартные средства поиска и выбора информации из архивов спутниковых данных. При разработке технологии создан ряд достаточно универсальных блоков интерфейсов, которые могут использоваться для настройки параметров различных процедур обработки данных.

- *Блок хранения заданий и результатов.* Состоит из базы данных заданий, файлового архива результатов проведенных обработок, библиотеки доступа к ним и модулей представления данных.

- *Блок работы с заданиями.* Предназначен для контроля очереди заданий, запуска процедур обработки и управления ими. Любая такая процедура включает запрос в центры архивации для выбора и подготовки исходных данных.

- *Блок подготовки исходных данных для обработки.* Обеспечивает выбор наборов данных из архивов и передачу их в процедуры обработки. Подготовка включает поиск необходимой информации, перепроецирование, вырезание нужных каналов исходных спутниковых данных, обрезку только по требуемой области, калибровку и т.д.

Одним из важнейших моментов технологии является быстрое получение данных из архивов для подготовки их к дальнейшей обработке. Серверы, проводящие обработку, для каждого задания могут параллельно запрашивать данные из разных центров архивации. Для подготовки данных используются мощности самих центров архивации, которые предоставляют для обработки стандартизированные наборы информации. Это позволяет унифицировано проводить совместную обработку данных, получаемых от различных систем. Кроме того, при таком подходе удалось минимизировать объем передаваемой информации из центров архивации на серверы, проводящие обработку, а также максимально задействовать возможности самих центров.

Примеры построенных на основе созданной технологии инструментов анализа данных

На основе разработанной технологии распределенной обработки данных были созданы различные инструменты, позволяющие с использованием веб-интерфейсов проводить обработку и анализ данных, в том числе классификацию различных наборов спутниковой информации, расчет различных индексов, различные операции с изображениями (инструмент «алгебра изображений») и т.д.

В инструментах классификации реализовано несколько методов с обучением и без. Эти инструменты создавались с использованием модулей GRASS

GIS [12]. Были также реализованы различные возможности для построения обучающих выборок, в том числе для обеспечения проведения классификации данных высокого пространственного разрешения с обучением по картам, построенным на основе данных менее детальной спутниковой информации. В классификации доступны разнообразные опции, например фильтрация результата, применение масок облачности, водных объектов, области интереса, топографическая коррекция и многие другие.

Инструмент «алгебра изображений» позволяет проводить арифметические, логические операции и различные математические преобразования над данными. Для этого непосредственно в картографическом веб-интерфейсе есть возможность задать формулу преобразования данных в выбранных каналах любых доступных спутниковых данных и по ней рассчитать результирующее растровое изображение.

Еще одним примером реализованных в настоящее время инструментов является расчет различных спектральных индексов. Заметим, что на сегодняшний день существует большое количество стандартных спектральных индексов, ориентированных на анализ различных процессов и объектов, в том числе на анализ различных атмосферных процессов. Применимость таких индексов в задачах мониторинга атмосферы показана, например, в работах [13, 14]. Работа со спектральными индексами осуществляется следующим образом: выбор спутниковых данных, выбор необходимого индекса, расчет индекса для конкретных данных и анализ полученного результата. Инструмент для расчета индексов позволяет применить выбранный пользователем индекс «на лету» к любым предоставляемым системой спутниковым данным, имеющим необходимые спектральные каналы для расчета конкретного индекса. Результатом применения индекса является изображение, каждый пиксель которого вычислен по формуле выбранного индекса из каналов с соответствующей индексу длиной волны. При этом весь процесс расчета и представления результата занимает малое время, исчисляемое в секундах, и у пользователей нет необходимости скачивать полный для проведения обработки набор выбранных спутниковых данных. Для проведения обработки пользователю также нет необходимости устанавливать на своем локальном компьютере специальное программное обеспечение. Всю обработку он может осуществлять с использованием только стандартного веб-браузера. Кроме того, созданный инструмент позволяет вести базу данных спектральных индексов: хранить, добавлять, редактировать и искать индексы с фильтрацией по области применения и типу данных, на работу с которыми они ориентированы.

Благодаря универсальности разработанной технологии, созданные на ее основе инструменты распределенного анализа данных сегодня удалось внедрить в различные специализированные информационные системы, например: спутниковый сервис ВЕГА (<http://pro-vega.ru/>), предназначенный для решения задач дистанционного мониторинга растительного покрова [15]; информационную систему

дистанционного мониторинга Федерального агентства лесного хозяйства (ИСДМ-Рослесхоз) (<https://nffc.aviales.ru/>) [16, 17]; объединенную систему работы с данными центров приема и обработки спутниковых данных ФГБУ НИЦ «Планета» Росгидромета (<http://moscow.planeta.smislab.ru/>) [18]; спутниковый сервис See the Sea (<http://ocean.smislab.ru>) для решения задач, связанных с исследованием процессов на поверхности океана [19]; информационную систему VolSatView «Дистанционный мониторинг активности вулканов Камчатки и Курил» (<http://volcanoes.smislab.ru>) [20].

Использование разработанных инструментов для обеспечения возможности работы с данными в информационной системе ВЕГА-Science для решения задач исследования и мониторинга процессов в атмосфере и океане

Разработанная технология, на наш взгляд, особенно важна при создании новых подходов и инструментов анализа спутниковой информации для

решения различных научных задач. Следует учитывать, что в последние годы все более востребованными становятся возможности анализа разнородной, в том числе разновременной, информации включая ряды наблюдений. Поэтому в научных исследованиях становятся необходимы системы, обеспечивающие не только удобный доступ к данным, но и предоставляющие различные инструменты и технические мощности для их обработки и анализа.

Одной из таких систем, развивающихся в последние годы в России, является Центр коллективного пользования «ИКИ-Мониторинг», подробно описанный в [21]. Центр предоставляет сегодня исследователям возможность работы со сверхбольшими архивами спутниковых данных и результатами их обработки. В архивах находится информация, полученная более чем 20 различными приборами ДЗЗ, имеются данные с 1984 г. по настоящее время, которые полностью покрывают территорию Северной Евразии, а также некоторые другие регионы земного шара. В непосредственном доступе (online) исследователям доступно более петабайта данных. В архивы ежедневно поступают новые оперативно получаемые данные. Для распределенной работы с информацией, предоставляемой Центром, создан спутниковый сервис ВЕГА-Science (<http://sci-vega.ru/>), который, в частности, предоставляет пользователям различные

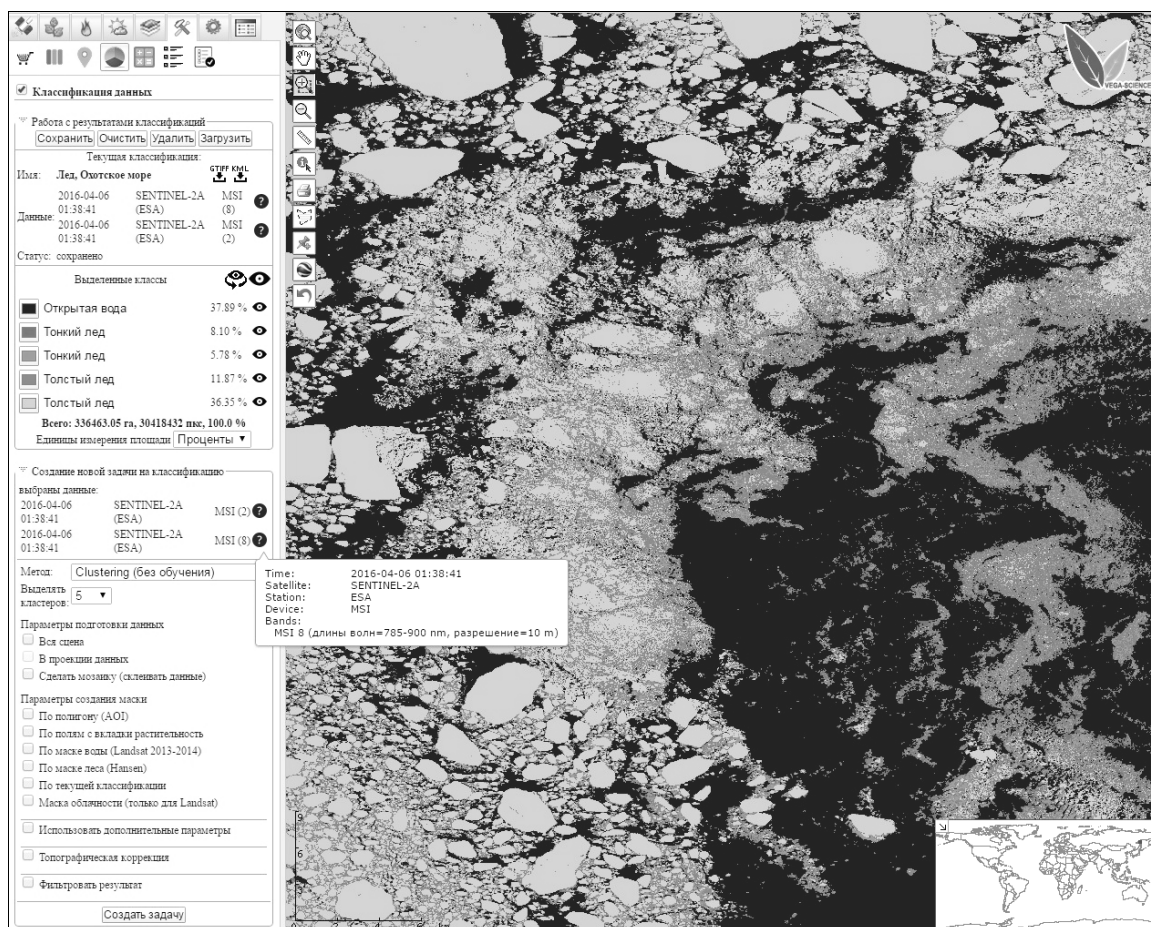


Рис. 2. Пример классификации ледовых полей с использованием данных спутника Sentinel-2A в интерфейсе системы ВЕГА-Science

инструменты для обработки спутниковых данных, созданные на основе описываемой в данной работе технологии. Мы приведем несколько примеров возможностей таких инструментов.

На рис. 2 приведен пример использования инструмента классификации для построения карт ледовых полей. На нем представлен результат классификации изображения, полученного 19 февраля 2016 г. со спутника Sentinel-2A по акватории Охотского моря. Данный инструмент позволяет не только провести классификацию различными методами, но и получить различную информацию о сформированных классах.

Особо отметим, что предоставляемые Центром «ИКИ-Мониторинг» возможности позволили как достаточно быстро осуществить выбор необходимой информации, так и произвести ее обработку для подготовки данных примеров и не потребовали использования никакого специального программного обеспечения, кроме веб-браузера.

Заключение

Таким образом, для развития различных систем, обеспечивающих работу с данными ДЗЗ, удалось создать и использовать технологию, позволяющую реализовывать достаточно сложные процедуры распределенной обработки и анализа данных дистанционного зондирования, оперативно получаемых из сверхбольших распределенных архивов. Разработанная технология позволяет реализовывать принципиально новые подходы к работе со сверхбольшими распределенными архивами данных спутниковых наблюдений Земли. В частности, она позволяет избежать необходимости формирования у пользователей крупных архивов спутниковых данных и создания достаточно дорогих программно-аппаратных комплексов, необходимых для обработки больших объемов информации. Это позволяет существенно упростить процессы получения и анализа спутниковой информации в рамках различных научных и прикладных проектов.

Работа выполнена при поддержке ФАНО (тема «Мониторинг», госрегистрация № 01.20.0.2.00164) и РФФИ (грант № 16-37-00427 мол_а).

1. *Лузян 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.
2. *Лузян Е.А., Балашов И.В., Бурцев М.А., Ефремов В.Ю., Кашицкий А.В., Кобец Д.А., Крашенинникова Ю.С., Мазуров А.А., Назиров Р.Р., Прошин А.А., Сычугов И.Г., Толтин 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.
3. *Кашицкий А.В., Балашов И.В., Лузян Е.А., Толтин В.А., Уваров И.А.* Создание инструментов для удаленной обработки спутниковых данных в современ-

ных информационных системах // *Современные проблемы дистанционного зондирования Земли из космоса*. 2015. Т. 12, № 1. С. 156–170.

4. *Шокин Ю.И., Добрецов Н.Н., Мамаш Е.А., Кухтенко 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.
5. *Недолужко И.В., Бабяк П.В., Тарасов Г.В., Ерёмченко В.С.* Инфраструктура приема, распределенной 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.
6. *Ramapriyan H.K., Behnke J., Sofinowski E., Lowe D., Esfandiari M.A.* 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.
7. *Acker J.G., Leptoukh G.* Online analysis enhances use of NASA earth science data // *Eos. Transactions American Geophysical Union*. 2007. V. 88, N 2. P. 14–17.
8. *Gorelick N.* Google Earth Engine // *EGU General Assembly Conference Abstracts*. 2013. V. 15. P. 11997.
9. *Балашов И.В., Бурцев М.А., Ефремов В.Ю., Лузян Е.А., Прошин А.А., Толтин В.А.* Построение архивов результатов обработки спутниковых данных для систем динамического формирования производных информационных продуктов // *Современные проблемы дистанционного зондирования Земли из космоса*. 2008. Т. 5, № 1. С. 26–32.
10. *Ефремов В.Ю., Крашенинникова Ю.С., Лузян Е.А., Мазуров А.А., Прошин А.А., Флитман Е.В.* Оптимизированная система хранения и представления географически привязанных спутниковых данных // *Современные проблемы дистанционного зондирования Земли из космоса*. 2007. Т. 4, № 1. С. 125–132.
11. *Толтин В.А., Балашов И.В., Ефремов В.Ю., Лузян Е.А., Прошин А.А., Уваров И.А., Флитман Е.В.* Создание интерфейсов для работы с данными современных систем дистанционного мониторинга (система GEOSMIS) // *Современные проблемы дистанционного зондирования Земли из космоса*. 2011. Т. 8, № 3. С. 93–108.
12. *Neteler M., Bowman M.H., Landa M., Metz M.* GRASS GIS: A multi-purpose open source GIS // *Environmental Modelling & Software*. 2012. V. 31. P. 124–130.
13. *He J., Zha Y., Zhang J., Gao J.* Aerosol indices derived from MODIS data for indicating aerosol-induced air pollution // *Remote Sens.* 2014. V. 6, N 2. P. 1587–1604.
14. *Marshak A., Knyazikhin Y., Davis A.B., Wiscombe W., Pilewskie P.* Cloud – vegetation interaction: Use of Normalized Difference Cloud Index for estimation of cloud optical thickness // *Geophys. Res. Lett.* 2000. V. 27, N 12. P. 1695–1698.
15. *Лузян Е.А., Савин И.Ю., Барталев С.А., Толтин В.А., Балашов И.В., Плотников Д.Е.* Спутниковый сервис мониторинга состояния растительности («ВЕГА») // *Современные проблемы дистанционного зондирования Земли из космоса*. 2011. Т. 8, № 1. С. 190–198.
16. *Барталев С.А., Ершов Д.В., Коровин Г.Н., Котельников Р.В., Лузян 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.

17. *Лулян Е.А., Барталев С.А., Ершов Д.В., Котельников Р.В., Балашов И.В., Бурцев М.А., Егоров В.А., Ефремов В.Ю., Жарко В.О., Ковганко К.А., Колбудаев П.А., Крашенинникова Ю.С., Прошин А.А., Мазуров А.А., Уваров И.А., Стыцenco Ф.В., Сычугов И.Г., Флитман Е.В., Хвостиков С.А., Шуляк П.П.* Организация работы со спутниковыми данными в информационной системе дистанционного мониторинга лесных пожаров Федерального агентства лесного хозяйства (ИСДМ-Рослесхоз) // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12, № 5. С. 222–250.
18. *Лулян Е.А., Милехин О.Е., Антонов В.Н., Крамарева Л.С., Бурцев М.А., Балашов И.В., Толтин В.А., Соловьев В.И.* Система работы с объединенными информационными ресурсами, получаемыми на основе спутниковых данных в центрах НИЦ «ПЛАНЕТА» // Метеорол. и гидрол. 2014. № 12. С. 89–97.
19. *Лулян Е.А., Матвеев А.А., Уваров И.А., Бочарова Т.Ю., Лаврова О.Ю., Митягина М.И.* Спутниковый сервис See the Sea – инструмент для изучения различных явлений на поверхности океана // Современные проблемы дистанционного зондирования Земли из космоса. 2012. Т. 9, № 2. С. 251–262.
20. *Ефремов В.Ю., Гурина О.А., Крамарева Л.С., Лулян Е.А., Маневич А.Г., Мельников Д.В., Матвеев А.М., Прошин А.А., Сорокин А.А., Флитман Е.В.* Создание информационного сервиса «Дистанционный мониторинг активности вулканов Камчатки и Курил» // Современные проблемы дистанционного зондирования Земли из космоса. 2012. Т. 9, № 5. С. 155–170.
21. *Лулян Е.А., Бурцев М.А., Балашов И.В., Барталев С.А., Ефремов В.Ю., Кашицкий А.В., Мазуров А.А., Матвеев А.М., Суднева О.А., Сычугов И.Г., Толтин 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.

A.V. Kashnitskii, E.A. Lupyana, I.V. Balashov, A.M. Konstantinova. **Technology for producing tools for processing and analysis of data from very large ever-expanding archives.**

Rapid growth of Earth satellite observation data over the recent years actualizes the problem of development of new technologies for effective data search, selection and processing within very large distributed ever-expanding archives. The paper describes the features of such technologies developed by Space Research Institute of RAS (IKI RAN). These technologies provide implementation of various data processing tools for satellite data analysis operating with distributed computing resources of remote sensing data processing and archiving centers. The paper shows their advantages and capabilities and gives the examples of developed tools for distributed processing of data from various satellite remote sensing systems. The examples given are provided to demonstrate the possibilities of described tools for various atmospheric and ocean surface phenomena analysis.