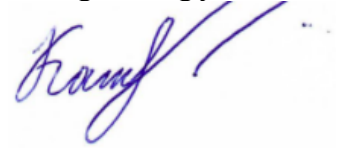


На правах рукописи



КАШНИЦКИЙ Александр Витальевич

**МЕТОДЫ, МОДЕЛИ И ПРОГРАММНЫЕ ИНФРАСТРУКТУРЫ
ГЛОБАЛЬНО РАСПРЕДЕЛЕННОЙ ОБРАБОТКИ
АРХИВНЫХ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ**

Специальность 05.13.11 – Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Москва – 2017

Работа выполнена в
Федеральном государственном бюджетном учреждении науки Институте
космических исследований Российской академии наук (ИКИ РАН)

Научный руководитель: **Лупян Евгений Аркадьевич**,
доктор технических наук, заведующий отделом «Технологии
спутникового мониторинга», заместитель директора,
Федеральное государственное бюджетное учреждение науки
Институт космических исследований Российской академии
наук (ИКИ РАН), г. Москва

Официальные оппоненты: **Алексанин Анатолий Иванович**,
доктор технических наук, заведующий лабораторией
спутникового мониторинга, Федеральное государственное
бюджетное учреждение науки Институт автоматике и
процессов управления Дальневосточного отделения
Российской академии наук, г. Владивосток

Кочергин Андрей Михайлович,
кандидат технических наук, старший научный сотрудник,
Федеральное государственное бюджетное образовательное
учреждение высшего образования «Рязанский
государственный радиотехнический университет», г. Рязань

Ведущая организация: **Федеральное государственное автономное образовательное
учреждение высшего образования «Сибирский
федеральный университет»**, г. Красноярск

Защита состоится **28 февраля 2018 года в 11.00 часов** на заседании
диссертационного совета Д 212.211.01 в Федеральном государственном
бюджетном образовательном учреждении высшего образования «Рязанский
государственный радиотехнический университет» по адресу:
390005, г. Рязань, ул. Гагарина, д. 59/1.

С диссертацией можно ознакомиться в библиотеке РГРТУ и на сайте
РГРТУ: <http://www.rsreu.ru>.

Автореферат разослан _____

Ученый секретарь
диссертационного совета Д 212.211.01
доктор технических наук, доцент



**Прузков
Александр
Викторович**

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования и степень ее разработанности. Постоянно возрастающая потребность в данных космической съемки и развитие технологий в последнее десятилетие привели к появлению большого количества высококачественных систем дистанционного зондирования Земли (ДЗЗ). Объем получаемой с них и доступной широкому кругу потребителей информации последние годы стремительно увеличивается. Для обеспечения эффективного использования имеющихся спутниковых данных требуется разработка новых подходов и методов работы с ними. Такие подходы должны обеспечить не только поиск и получение информации, но и организацию ее анализа с использованием распределенных вычислительных ресурсов. Работы в этом направлении ведутся в последнее десятилетие как в России, так и за рубежом различными исследователями (Александрин А.И., Асмус В.В., Кузнецов А.Е., Лупян Е.А., Лошкарев П.А., Маглинец Ю.А., Шокин Ю.И., Acker J. G., Gorelick N., Hansen M. C., Koubarakis M., Leptoukh G., Moore R. T. и др.). Данная диссертационная работа направлена на создание новых методов, модели и программных инфраструктур, позволяющих эффективно проектировать и поддерживать с использованием распределенных вычислительных ресурсов различные инструменты анализа и обработки информации, предоставляемой сверхбольшими архивами данных ДЗЗ.

Целью диссертации является разработка методов, модели данных и на их основе программных инфраструктур для построения систем распределенной обработки архивных данных ДЗЗ, которые должны обеспечить пользователям не только возможности эффективного доступа к данным, но и предоставление средств и вычислительных ресурсов для проведения их обработки и анализа. Для этого в работе решаются следующие **задачи**:

- Анализируются существующие тенденции развития систем ДЗЗ, подходы к организации обработки и предоставления исследователям получаемой с них информации, в том числе проводится обзор существующих информационных систем, позволяющих выполнять анализ и обработку спутниковых данных из больших распределенных архивов.
- На основе проведенного анализа формулируются основные проблемы при организации работы с большими распределенными непрерывно пополняющимися архивами информации ДЗЗ и намечаются пути их решения.
- Разрабатывается модель данных ДЗЗ для ведения больших распределенных систем хранения с возможностью эффективной работы с виртуальными (формируемыми по запросу) информационными продуктами.
- Создается программная инфраструктура ведения больших распределенных архивов данных ДЗЗ с возможностью предоставления удаленным исследователям сложных "виртуальных" информационных продуктов.
- Разрабатываются принципы построения больших распределенных архивов для предоставления информации, необходимой для организации анализа и обработки разнородных данных.
- Разрабатываются методы построения человеко-машинных интерфейсов систем распределенной интерактивной обработки спутниковых данных, которые обеспе-

чат возможность организации управляемых пользователями процедур обработки с выполнением их на вычислительных ресурсах центров хранения.

- На основе этих методов создается и внедряется программная инфраструктура, обеспечивающая работу систем распределенной обработки спутниковых данных.
- Показывается эффективность разработанных решений путем внедрения и использования их в различных научных и прикладных системах дистанционного мониторинга.

Научная новизна заключается в том, что в работе предложены новые методы работы со сверхбольшими распределенными архивами спутниковых данных с использованием развитых человеко-машинных интерфейсов, созданных на основе web-технологий. Основой для разработки данных методов стала предложенная в работе новая модель данных, обеспечивающая ведение распределенных систем хранения и ориентированная на динамическое формирование сложных виртуальных информационных продуктов в момент их запроса на основе хранящихся. Эта модель позволила существенно оптимизировать системы хранения, в том числе значительно сократить объемы архивов и унифицировать схему работы с разнородными данными. Разработанная модель, методы и созданная на их основе программная инфраструктура значительно повышают эффективность работы с информацией ДЗЗ за счет того, что пользователи получают возможность не только поиска и выбора необходимых им наборов данных, но и проведение их обработки и анализа с использованием распределенных вычислительных ресурсов центров, обеспечивающих хранение и предоставление спутниковой информации. Это, в частности, позволяет существенно сократить необходимость передачи больших объемов данных.

Основные положения, выносимые на защиту:

1. На основе системного анализа особенностей хранения и предоставления информации ДЗЗ и перспектив их развития разработана модель данных для ведения сверхбольших распределенных систем хранения с возможностью эффективной работы с "виртуальными" (формируемыми по запросу) информационными продуктами. Предложенная модель позволяет принципиально повысить эффективность работы со сверхбольшими архивами, в том числе на порядки уменьшить объем хранимых данных и время на модификацию и расширение состава предоставляемых на их основе информационных продуктов.

2. На основе разработанной модели создана программная инфраструктура ведения сверхбольших распределенных архивов данных ДЗЗ, которая внедрена в различные научные и прикладные информационные системы дистанционного мониторинга. В том числе создана система ведения распределенных архивов данных Центра коллективного пользования "ИКИ-Мониторинг", который по объемам информации ДЗЗ, доступной в режиме "online", входит в десятку самых крупных центров в мире.

3. На основе использования современных подходов работы со сверхбольшими архивами данных ДЗЗ разработаны методы построения человеко-

машинных интерфейсов систем распределенной интерактивной обработки спутниковых данных, в том числе:

- метод организации систем распределенной интерактивной обработки информации, предоставляемой сверхбольшими глобальными архивами данных ДЗЗ;
- метод организации массовой интерактивной распределенной обработки спутниковых данных ДЗЗ для картографирования различных объектов.

Предложенные методы позволили создать новый класс эффективных систем обработки и анализа спутниковой информации, которые не требуют использования дорогостоящих локальных приложений и вычислительных ресурсов, а также позволяют на несколько порядков сократить объемы передаваемых данных.

4. Для реализации предложенных методов разработана программная инфраструктура, обеспечивающая работу систем распределенной обработки спутниковых данных, которая внедрена в различные научные и прикладные информационные системы. В том числе, создана система распределенного анализа данных в информационной системе "ВЕГА-Science", которая по возможностям удаленного интерактивного анализа данных ДЗЗ не имеет аналогов в мире.

Все основные результаты получены автором лично.

Теоретическая и практическая значимость работы. Ценность работы заключается в том, что полученные результаты существенно повышают эффективность работы со сверхбольшими глобально распределенными непрерывно пополняющимися архивами спутниковых данных в различных научных и прикладных системах дистанционного мониторинга. Построенные на основе созданных методов и программной инфраструктуры системы по функциональности во многих случаях не уступают дорогостоящим настольным приложениям, ориентированным на обработку информации ДЗЗ. Это обеспечивает возможность работы со сверхбольшими распределенными массивами информации с любого компьютера без использования сложных локальных приложений, вычислительных ресурсов и средств хранения. Внедрение созданных в работе подходов позволило решать различные научные задачи с использованием возможностей центров предоставления спутниковой информации, в том числе вычислительных ресурсов этих центров. Разработанная модель данных и созданная на ее основе программная инфраструктура позволили оптимизировать действующие системы хранения в различных научных и прикладных системах дистанционного мониторинга, существенно уменьшив требования к необходимым ресурсам хранения, сократив объемы обмена данными и обеспечив возможность легкого изменения состава спутниковых информационных продуктов без проведения массовой переобработки архивов.

Полученные результаты позволяют принципиально изменить подходы к работе со спутниковой информацией, существенно упростив организацию ее использования и, как следствие, расширить область применения данных дистанционного зондирования. Опыт успешной эксплуатации созданных на основе разработанных методов и подходов интерактивных инструментов анализа и обработки подтверждает значимость и эффективность полученных результатов. Практическая ценность работы также подтверждается значительным числом их внедрений в различные научные и прикладные системы дистанционного мониторинга.

Апробация работы. Основные положения диссертации докладывались и обсуждались на более чем 20-ти международных и всероссийских конференциях, в том числе: на 10-й - 15-ой всероссийских открытых конференциях "Современные проблемы дистанционного зондирования Земли из космоса", Москва, ИКИ РАН, 2012, 2013, 2014, 2015, 2017; на третьей и четвертой международных научно-технических конференциях "Актуальные проблемы создания космических систем дистанционного зондирования Земли", Москва, АО "Корпорация "ВНИИЭМ", 2015, 2016; на шестой Всероссийской (с международным участием) конференции "Аэрокосмические методы и геоинформационные технологии в лесоведении и лесном хозяйстве", Москва, ЦЭПЛ РАН, 2016; на 9-ой - 14-ой конференциях молодых ученых "Фундаментальные и прикладные космические исследования", Москва, ИКИ РАН, 2011, 2012, 2013, 2014, 2016, 2017, всероссийской конференции, проводимой ИВТ СО РАН "Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов", Бердск, Новосибирская область, 2017, 7-ой Международной научно-технической конференции К.Э. Циолковский – 160 лет со дня рождения. Космонавтика. Радиоэлектроника. Геоинформатика, Рязань, 2017.

В 2016 году за работу, положенную в основу диссертации, автор стал победителем конкурса молодых ученых Всероссийской конференции "Современные проблемы дистанционного зондирования Земли из космоса".

Публикации. Результаты диссертации опубликованы в 28 статьях (в том числе 15 статьях в изданиях по списку ВАК России) и 62 тезисах докладов на международных и всероссийских конференциях. Получено 2 свидетельства о государственной регистрации программ.

Внедрение результатов работы. Диссертационная работа выполнена в Институте космических исследований Российской академии наук (ИКИ РАН). На проведение исследований, входящих в данную работу, автором был выигран грант РФФИ 16-37-00427. Результаты работы использованы при выполнении проекта РАН № 01.20.0.2.00164 (тема "Мониторинг"), проектов Минобрнауки № 14.616.21.0063, 14.607.21.0122, грантов РФФИ № 16-37-00427, 15-07-05564, 16-29-09615, проекта РНФ № 16-17-00042, а также в рамках проекта SIGMA при поддержке Рамочной программы Европейской комиссии FP7.

Результаты работы используются во многих организациях, в том числе: ИКИ РАН, ФБУ "Авиалесоохрана", ФГБУ "НИЦ "Планета", Институт вулканологии и сейсмологии ДВО РАН. Результаты работы применены при создании и развитии ЦКП "ИКИ-Мониторинг", пользователями которого в настоящее время является более 50 организаций.

Достоверность результатов работы заключается в оценке их эффективности и применимости. Достижение цели работы было подтверждено в результате внедрения и применения созданных подходов при решении с использованием спутниковых данных практических задач самого разного назначения. Количественными оценками эффективности является значительное сокращение необходимых ресурсов, которое нужно затратить на решение различных практических задач и требуемых при создании больших архивов информации ДЗЗ и

информационных систем дистанционного мониторинга (в том числе ресурсов хранения, вычислительных, экономических и человеческих ресурсов).

Методология и методы исследования. В работе использованы методы математического моделирования, методы прикладного программирования, методы системного анализа, методы дистанционного зондирования Земли, методы классификации изображений, методы статистического анализа.

Соответствие паспорту специальности 05.13.11. Диссертация соответствует паспорту специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» в части:

– пункта 7 «Человеко-машинные интерфейсы; модели, методы, алгоритмы и программные средства машинной графики, визуализации, обработки изображений, систем виртуальной реальности, мультимедийного общения»;

– пункта 8 «Модели и методы создания программ и программных систем для параллельной и распределенной обработки данных, языки и инструментальные средства параллельного программирования»;

– пункта 9 «Модели, методы, алгоритмы и программная инфраструктура для организации глобально распределенной обработки данных».

Структура и объем работы. Работа состоит из введения, четырех глав, заключения, 5 приложений и списка литературы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе представлены результаты анализа развития спутниковых систем ДЗЗ и подходов к организации работы с информацией, поступающей от них. Выполненный анализ показывает, что наблюдающийся в последние годы резкий рост возможностей спутниковых систем ДЗЗ и объемов поступающей с них информации требует создания новых методов и технологий организации работы со сверхбольшими глобально распределенными постоянно обновляющимися архивами данных ДЗЗ для обеспечения их эффективного использования при решении научных и прикладных задач. Проведенный анализ позволил сформулировать основные требования к перспективным методам и технологиям работы с данными ДЗЗ.

Проведенный в первом параграфе анализ показывает, что одной из важнейших тенденций в развитии современных систем ДЗЗ является непрерывный рост качества и количества получаемой информации. Во втором параграфе показано, что в связи с этим становятся практически неприменимы "традиционные" подходы к работе с данными ДЗЗ, предполагающие выбор необходимых наборов данных из распределенных центров хранения, создание локальных архивов и средств обработки под конкретные задачи и проведение с их помощью обработки и анализа полученных данных. Такой подход уже не может обеспечить эффективное использование получаемых современными системами ДЗЗ данных, поскольку даже при решении локальных задач объемы необходимой информации могут превышать десятки и сотни терабайт. Основным сдерживающим фактором при этом является необходимость создания в рамках отдельных проектов достаточно дорогостоящих специализированных комплексов

хранения и обработки данных. Альтернативой "традиционного" подхода является создание систем, которые кроме доступа к архивам информации ДЗЗ, предоставляют возможность их анализа и обработки с использованием распределенных вычислительных ресурсов.

В третьем параграфе проведен обзор современных систем работы с информацией ДЗЗ, которые обеспечивают возможность обработки и анализа данных с использованием распределенных вычислительных ресурсов. Отмечено, что хотя работы в данном направлении ведутся, однако, как в России, так и за рубежом находятся на первых стадиях развития.

В четвертом параграфе сформулированы основные проблемы, возникающие при работе с большими объемами информации ДЗЗ. К ним, в частности, относится организация хранения и предоставления удаленным пользователям и системам обработки больших объемов разнородной информации ДЗЗ и значительного числа различных информационных продуктов, получаемых на ее основе. Обсуждается понятие виртуальных информационных продуктов, формируемых по запросу пользователей в режиме "online". Использование таких продуктов позволяет существенно сократить объемы хранящихся данных. Показано, что для эффективного использования в процедурах обработки и анализа информации ДЗЗ (в том числе при получении виртуальных продуктов) требуется создание новых подходов к организации и ведению сверхбольших распределенных архивов. Формулируются основные требования к таким архивам. Также сформулированы основные задачи, которые необходимо решить в работе.

Во второй главе представлены разработанная модель данных для организации сверхбольших распределенных систем хранения данных ДЗЗ и созданная на ее основе программная инфраструктура, ориентированные на предоставление виртуальных информационных продуктов. Созданная модель и программная инфраструктура повышают эффективность ведения больших архивов, уменьшают объемы информации, хранение которой необходимо обеспечить, позволяют легко модифицировать состав предоставляемых информационных продуктов и дают возможность создания процедур обработки и анализа разнородных данных из архивов.

Модель данных включает следующие основные составляющие: набор типов и логических структур данных, набор правил манипулирования данными и набор правил поддержки целостности. Наименьшим логическим элементом хранимого растрового цифрового изображения являются пиксели. Наборы пикселей объединены в понятие канала: $K = \{px_{DN}(m, n); m = 1, M; n = 1, N\}$, где px_{DN} - цифровое значение яркости пикселя с координатами m, n в канале. Каналы объединяются в отдельный растровый файл: $I = \{K_1, K_2, \dots, K_L\}$, где L - количество каналов в файле. Для ускорения выборки и просмотра также организуется хранение обзорных (прореженных) уровней данных, которые могут находиться в отдельных файлах. Отметим, что на основе одного и того же набора базовых данных обычно формируется значительное число производных информационных продуктов. Будем далее называть термином *реальный продукт* заранее сформированный информационный продукт, физически хранящийся в архиве спутниковых данных.

Состав реального продукта зависит от его особенностей и планируемых способов использования. Возможно создание реальных продуктов на основе различных данных, различными версиями программного обеспечения и так далее. В результате в данное понятие хранящегося в архиве реального продукта входит:

$$R = \{I_1, I_2, \dots, I_S, \dots, I_S, Q_1, Q_2, \dots, Q_S, \dots, Q_S; s = 1, S\},$$

где R – реальный информационный продукт, $I_S = \{px_{DN}(m, n, k); m = 1, M; n = 1, N; k = 1, K\}$ – версия номер S построения реального продукта, а $Q_S = \{px_{DNq}(m_q, n_q, k); m_q = 1, M_q; n_q = 1, N_q; k = 1, K\}$ – обзорные прореженные уровни для быстрого доступа на мелком масштабе к варианту продукта номер S , px_{DN} – цифровое значение яркости пикселя с координатами m, n в канале k продукта, px_{DNq} – цифровое значение яркости пикселя с координатами m_q, n_q в канале k обзорного прореженного уровня продукта. Также для каждого значения пикселя любого канала любого варианта построения любого продукта должно существовать правило его преобразования в физическую величину $px_{physicalvalue} = f(px_{DN})$ и обратное преобразование $px_{DN} = f^{-1}(px_{physicalvalue})$, где px_{DN} – цифровое значение в пикселе исходного файла продукта, $px_{physicalvalue}$ – соответствующее ему значение физической величины.

Формирование виртуального продукта на основе хранящихся реальных продуктов в общем случае можно представить следующим образом:

$$\left. \begin{array}{l} R_1 \xrightarrow{F_{B1}, F_{B2}, \dots, F_{BN}} R_{1prep}, \\ R_2 \xrightarrow{F_{B1}, F_{B2}, \dots, F_{BN}} R_{2prep}, \\ \vdots \\ R_K \xrightarrow{F_{B1}, F_{B2}, \dots, F_{BN}} R_{Kprep}, \end{array} \right\} \xrightarrow{F_1, F_2, \dots, F_M} V,$$

где R_1, R_2, R_K – реальные продукты, используемые для создания предоставляемого виртуального продукта. F_{B1}, F_{B2}, F_{BN} – базовые операции преобразования, в результате которых получают подготовленные к совместному преобразованию реальные продукты $R_{1prep}, R_{2prep}, R_{Kprep}$. F_1, F_2, F_N – операции преобразования подготовленных реальных продуктов, в результате которых получается итоговый виртуальный. При формировании виртуального продукта к базовым операциям преобразования можно отнести операции по трансформированию снимков в определенную систему координат, приведение размеров и типов данных, вырезание конкретной географической области и так далее. К операциям финального преобразования хранящихся реальных продуктов в виртуальные могут относиться как достаточно простые преобразования, например, синтез нескольких каналов в цветное изображение, так и сложные многоуровневые операции, например, атмосферная или топографическая коррекция. В главе проводится анализ основных типов операций преобразования, необходимых для формирования продуктов, и определяется их сложность. Предложен метод оценки вычислительных ресурсов, необходимых для реализации конкретных архивов данных, обеспечива-

ющих формирование виртуальных продуктов определенного класса. Также описаны включенные в модель правила поддержки целостности.

Следует отметить, что представленная модель данных позволяет не только создавать архивы, рассчитанные на организацию работы с разнородными реальными и виртуальными продуктами, но и обеспечивать автоматическую подготовку стандартизированных наборов данных для проведения сложных комплексных операций обработки и анализа с использованием распределенных вычислительных ресурсов. Методам организации такой обработки посвящена глава 3 настоящей работы.

В главе также описана архитектура построения больших распределенных архивов данных ДЗЗ, представленная на рис.1.

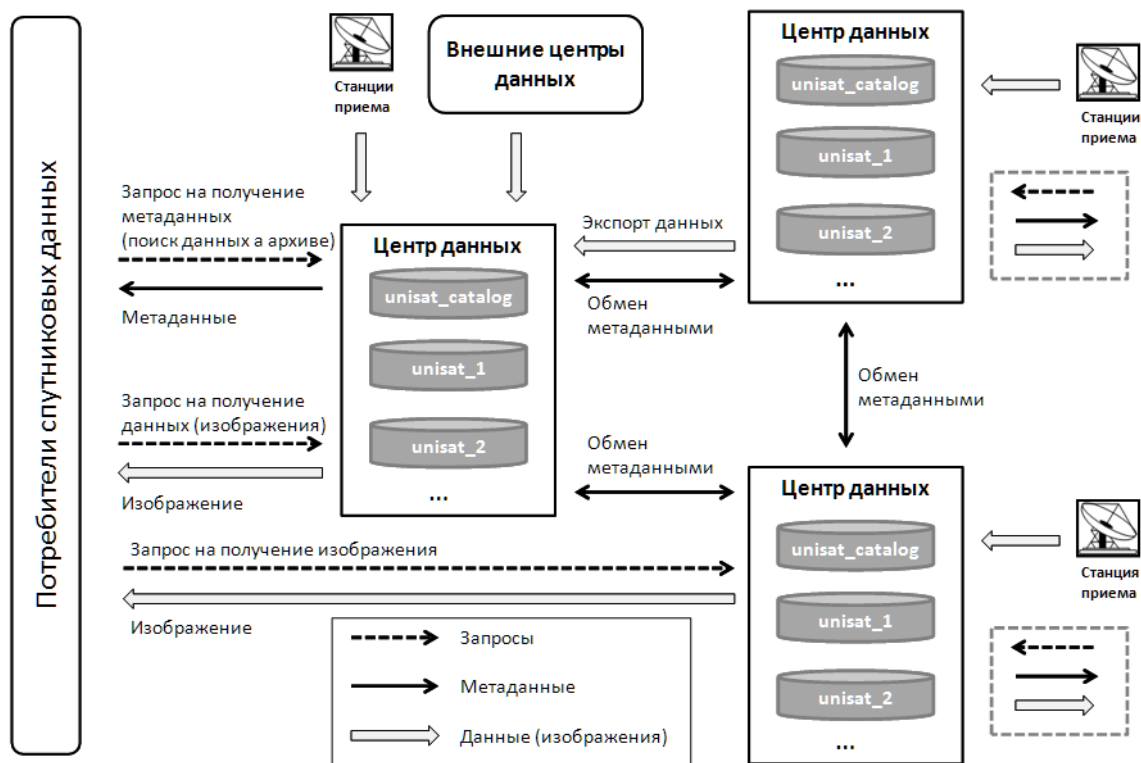


Рис.1. Схема построения распределенных архивов спутниковых данных

В соответствии с приведенной схемой информация ДЗЗ может поступать в систему после обработки от станций приема или из внешних центров. Данные заносятся в файловое хранилище, расположенное локально в пределах самого центра, куда поступили данные, а информация об этом – в базы метаданных. При этом организована постоянная синхронизация баз метаданных, расположенных во всех центрах. В результате, каждый центр имеет полную информацию о наличии данных во всех узлах распределенного архива. При этом поиск информации о наличии данных в распределенном архиве может осуществляться в любом из узлов, а их получение и предоставление обеспечивает узел, в котором они реально хранятся.

На основе предложенной модели данных и описанной архитектуры разработана программная инфраструктура (ПИ) построения сверхбольших распреде-

ленных архивов данных ДЗЗ. На рис. 2 представлены ее основные блоки и схема их взаимодействия.

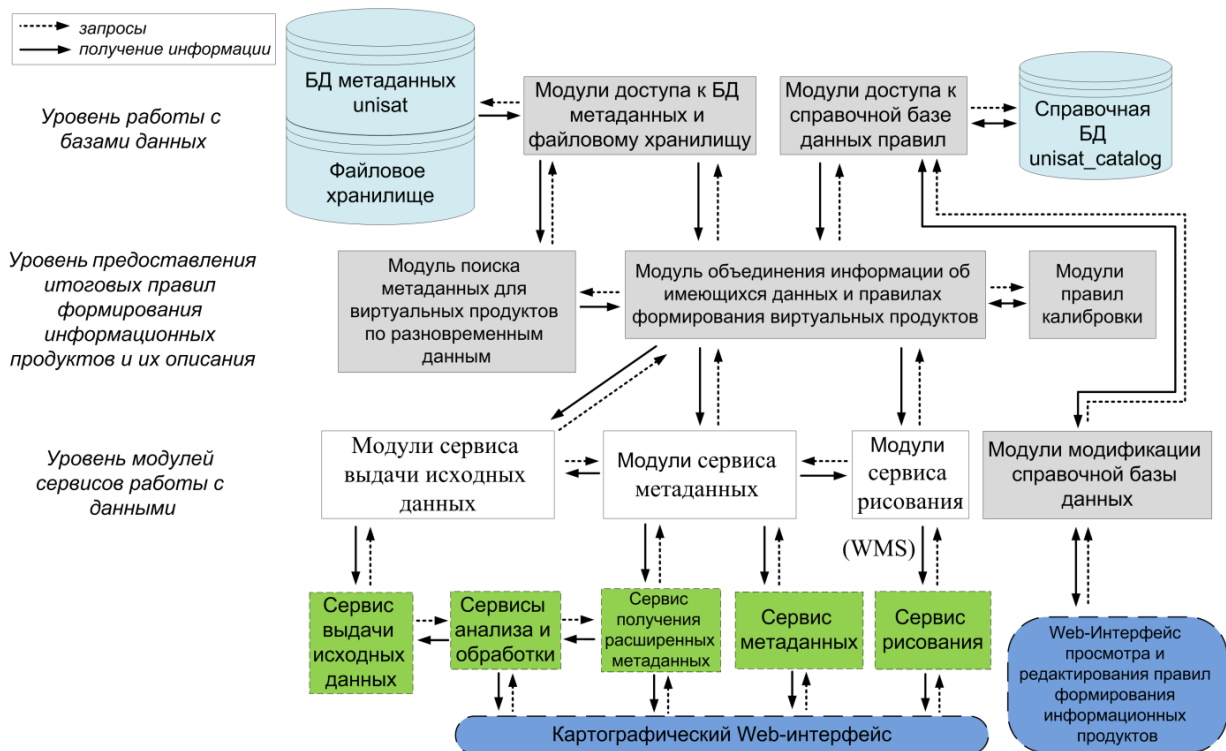


Рис.2. Основные элементы ПИ для построения распределенных архивов данных

Входящие в ПИ модули можно условно разбить на следующие уровни:

Уровень работы с базами данных (БД). Архив спутниковых продуктов поддерживается двумя основными реляционными БД: отвечающей за хранение самих данных (БД метаданных и связанное с ней файловое хранилище), и справочной БД. Структура справочной БД основана на созданной модели и повторяет все возможные особенности хранимых данных и информацию о том, из каких реальных продуктов могут формироваться виртуальные и какие для этого необходимы преобразования.

Уровень предоставления итоговых правил. Модули этого уровня объединяют информацию о хранящихся данных, полученную из БД метаданных, и правила формирования на их основе виртуальных продуктов, полученные из справочной БД. В результате на уровень модулей конкретных сервисов передается полная информация о хранящихся данных, объединенная с набором правил построения на их основе информационных продуктов, включая пути к файлам в архиве и необходимые операции преобразования.

В **уровень модулей сервисов работы с данными** входят: модули, реализующие основные сервисы, необходимые для организации работы картографических web-интерфейсов с распределенными архивами данных; модули формирования наборов данных для процедур "online" обработки данных (см. главу 3); модули модификации справочной БД, которые поддерживают отдельный web-интерфейс для изменения и добавления продуктов и правил их преобразования.

Предложенная структура ПИ является достаточно гибкой и позволяет создавать блоки ведения архивов и работы с виртуальными продуктами в различ-

ных специализированных информационных системах. Модули ПИ реализованы для UNIX-подобной ОС, написаны на языках Perl и C++, в качестве СУБД используется MariaDB или MySQL.

В главе также проводится анализ основных преимуществ и недостатков представленных в главе походов, модели данных и программной инфраструктуры, обеспечивающих возможность однотипной работы со сверхбольшими разнородными архивами данных с максимальным использованием виртуальных продуктов. В качестве основных минусов отмечается, что при необходимости обеспечивать работу большого числа пользователей могут потребоваться значительные вычислительные ресурсы в центрах хранения и предоставления информации. Однако использование таких ресурсов происходит коллективно и может быть оптимизировано. Основные преимущества предлагаемых подходов по сравнению с традиционными приведены в *таблице 1*.

Таблица 1. Сравнение реализаций базовых функций в новой системе с традиционными системами ведения архивов спутниковых данных

<i>Новая система ведения архивов</i>	<i>Традиционные системы ведения архивов</i>
Хранение данных	
Экономия средств хранения из-за отсутствия необходимости хранить продукты, перевод их на виртуальную схему	При широком диапазоне использования существенное возрастание объемов хранения из-за множества продуктов
Создание новых продуктов	
Создание нового продукта заключается лишь в описании правил его получения	Существенные затраты на подготовку и предварительную обработку
Гибкая схема добавления новых и модификации существующих продуктов – можно добавить или поменять в любой момент	Негибкая схема – любые изменения в схемах формирования продуктов ведут к необходимости обработки значительных объемов данных
Предоставление данных	
Возможность динамического формирования практически неограниченного числа различных информационных продуктов	Наличие существенных ограничений на число продуктов, связанное с объемами архивов
Особенности использования	
Быстрое и не требующее работы от пользователя внедрение новых данных и продуктов	Уникальная схема работы с каждым новым типом информации

В третьей главе разрабатываются методы и программная инфраструктура построения систем распределенной интерактивной обработки информации ДЗЗ, предоставляемой сверхбольшими глобальными архивами. Разработанные методы и программная инфраструктура обеспечивают:

- минимизацию объемов передаваемой информации за счет максимальной локализации процедур, выполняющих различные этапы обработки в местах их хранения;
- возможность использования вычислительных мощностей центров архивации для проведения обработки данных по запросам удаленных пользователей;
- организацию удаленного управления процедурами обработки и настройки их параметров с использованием web-интерфейсов;

- унификацию работы с разнородными данными в процедурах обработки и анализа за счет стандартизации форматов и наборов данных, передаваемых в них;
- унификацию создания и внедрения различных типов процедур обработки на основе стандартизации входной и выходной информации, а также схем подготовки данных и предоставления результатов.

На основе предложенных методов разработана схема организации распределенной обработки, обеспечивающая возможность выбора данных, подготовку их в виде, необходимом для проведения конкретной процедуры, передачу в процедуру обработки, выполнение обработки и организацию работы с ее результатами. При этом учитывается, что операции подготовки данных должны осуществляться с использованием вычислительных ресурсов центров, в которых осуществляется их хранение.

Для реализации предложенных методов разработана программная инфраструктура, основные блоки которой представлены на рис. 3.

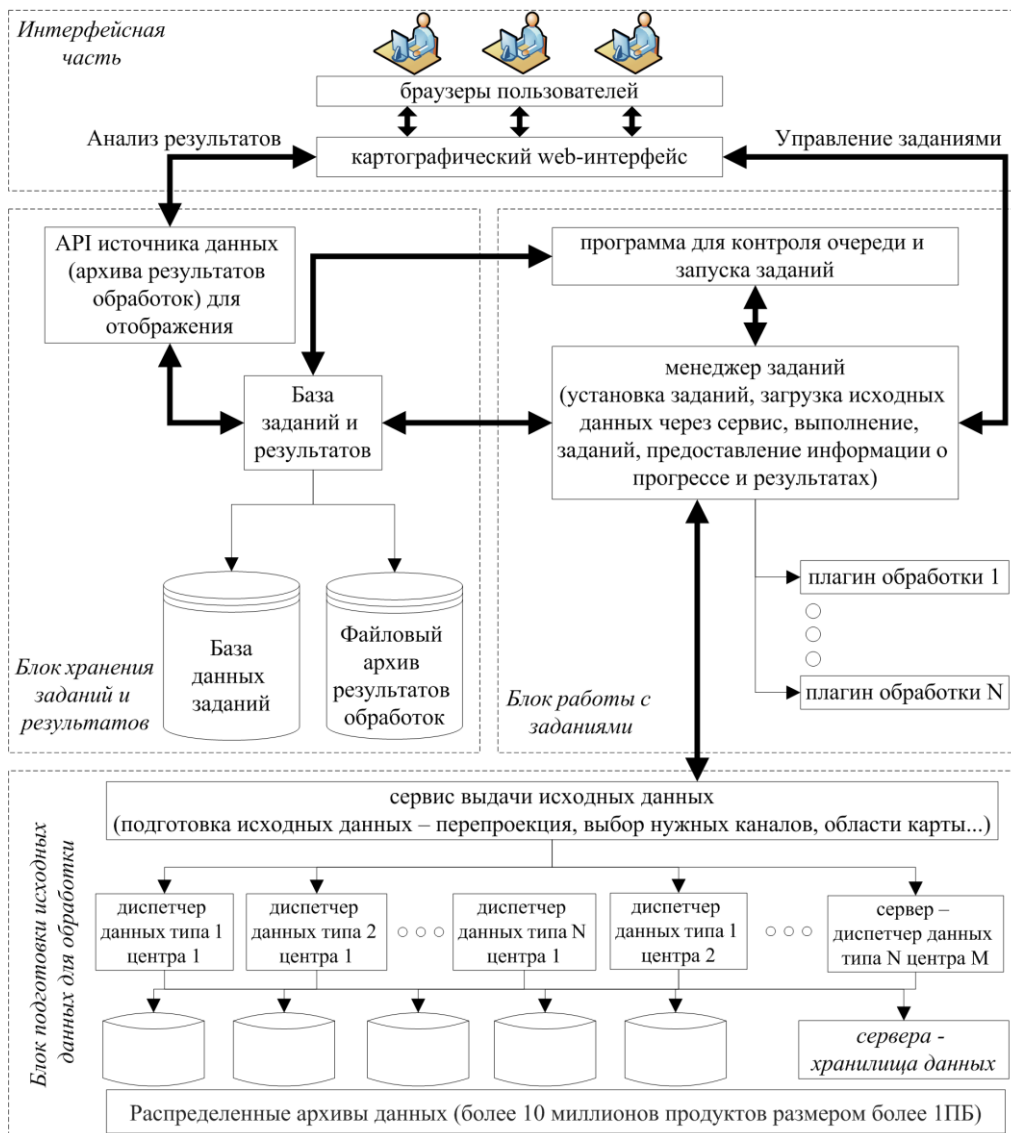


Рис. 3. Принципиальная схема построения программной инфраструктуры, обеспечивающей удаленную интерактивную обработку информации ДЗЗ из больших глобально распределенных архивов

В главе подробно описываются функции и особенности функционирования следующих основных ее блоков:

Интерфейсная часть. Обеспечивает возможность удаленного управления процедурами обработки (поиск и выбор данных, настройка параметров, запуск и контроль выполнения, а также анализ полученных результатов).

Блок хранения заданий и результатов. Обеспечивает хранение заданий на обработку и их результатов. Состоит из базы данных заданий, файлового архива результатов проведенных обработок, библиотеки доступа к ним и модулей предоставления данных.

Блок работы с заданиями. Поддерживает функции по диспетчеризации и управлению заданиями, а также по получению данных из мест их хранения через блок подготовки.

Блок подготовки данных для обработки. Обеспечивает выбор наборов данных из архивов и передачу их в процедуры обработки в унифицированном виде. В главе также описывается метод создания интерфейсов для управления интерактивными процедурами распределенной обработки данных ДЗЗ с использованием разработанной программной инфраструктуры. Приводятся примеры созданных на ее основе различных интерактивных инструментов для проведения распределенной обработки. На рис. 4 приведен пример такого инструмента для проведения классификации спутниковых данных, позволяющий, в том числе, автоматически создавать обучающие выборки на основе различных карт.

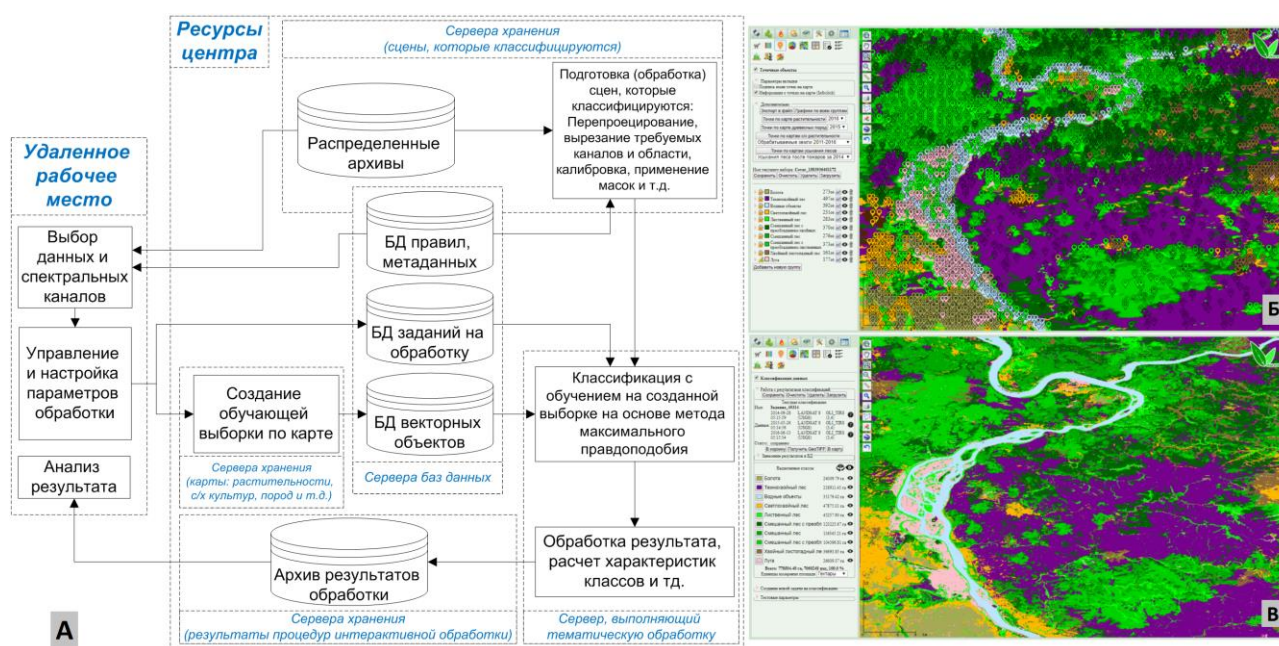


Рис. 4. Схема (А) и пример работы инструмента классификации данных LANDSAT (В) с обучением на автоматически создаваемой обучающей выборке. Для построения выборки используется карта растительности, построенная на основе данных MODIS (Б)

В главе также представлен метод организации процедур массовой интерактивной распределенной обработки спутниковых данных ДЗЗ для картографирования различных объектов. Организация таких процедур необходима в случаях, когда невозможно полностью автоматизировать конкретные процессы об-

работки. Процесс обработки в соответствии с методом разбивается на элементарные этапы, которые могут быть полностью автоматизированы. Для каждого такого этапа необходимо создание блоков автоматической обработки, а также специализированных интерфейсов управления для тех этапов, по которым необходимо принятие решений человеком. При этом, если возможно, выполняется максимальное упрощение процедуры принятия решения, в том числе за счет предоставления всей необходимой для этого информации в удобном для анализа виде. Предложенный метод позволяет значительно упростить и ускорить интерактивное картографирование объектов и делает возможным организацию интерактивной согласованной обработки данных по большим территориям распределенными коллективами.

Рассматривается пример организации массового картографирования лесных гарей на основе разработанных методов и программной инфраструктуры. Схема и пример проведения картографирования приведены на рис. 5. Разработанный способ, в том числе, позволяет осуществлять массовое картографирование гарей на всей территории России (более 10 000 пожаров за сезон).

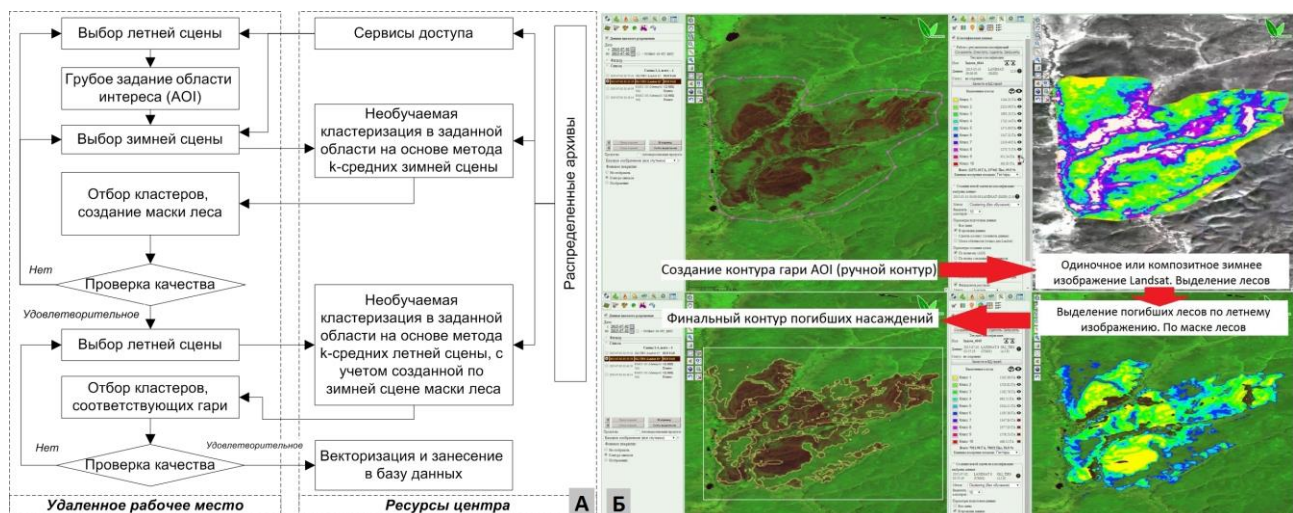


Рис.5. Схема организации картографирования погибших лесов (А) и основные этапы работы по такой схеме в картографическом web-интерфейсе (Б)

В главе 3 также обсуждается метод оценки необходимых человеческих и вычислительных ресурсов для организации процедур массовой интерактивной обработки информации ДЗЗ. На его основе проводится оценка эффективности разработанных методов.

В главе также проводится анализ основных преимуществ и недостатков способов работы с информацией ДЗЗ на основе разработанных методов и программной инфраструктуры. В качестве минусов отмечается, что пока возможности создаваемых web-интерфейсов еще уступают по ряду параметров специализированным настольным ГИС приложениям. Однако опыт создания и внедрения различных интерактивных инструментов распределенного анализа данных показывает, что стремительное развитие технологий позволит сократить это отставание в ближайшие годы. Основные преимущества предлагаемых подходов приведены в таблице 2.

Таблица 2. Преимущества разработанных способов работы с информацией ДЗЗ по сравнению с традиционными

<i>Предложенные</i>	<i>Традиционные</i>
Анализ и обработку можно проводить с любого компьютера, имеющего выход в интернет и стандартный web-браузер	Работу можно проводить только на компьютерах с установленным специализированным (во многих случаях дорогостоящим) ПО
Для работы доступны сверхбольшие распределенные архивы данных	Работа ведется только с локальными архивами данных
Существенное сокращение объемов передаваемой информации	Необходимость передачи больших объемов информации и формирования локальных архивов
Отсутствие необходимости в значительных локальных вычислительных мощностях	Необходимость наличия значительных локальных вычислительных мощностей
Возможность достаточно простой организации и проведения массовой интерактивной обработки данных распределенными коллективами	Сложность организации коллективной распределенной работы с данными
При появлении новых данных и/или возможностей обработки пользователи «мгновенно» получают возможности работы с ними без дополнительных усилий	Обновление системы и организация работы с новыми типами данных обычно требуют от пользователя значительных усилий

Четвертая глава посвящена практическому применению разработанных методов, модели и программной инфраструктуры при построении блоков обработки и анализа данных в информационных системах дистанционного мониторинга различного назначения. В ней показана эффективность созданных подходов и их универсальность при организации работы с различными типами данных при решении разных научных и прикладных задач.

В первом параграфе описываются блоки анализа и обработки информации ДЗЗ в информационной системе дистанционного мониторинга вулканов Камчатки и Курил (<http://volcanoes.smislab.ru>). Использование в системе разработанных методов и программной инфраструктуры позволило организовать эффективную работу с данными различных спутниковых систем, как для решения оперативных задач мониторинга вулканической активности, так и для решения различных научных задач, требующих анализа долговременных наблюдений. Рассматриваются возможности и особенности интерактивных инструментов, реализованных в системе. В том числе приведены примеры работы с быстро обновляющейся информацией (каждые 10 минут), получаемой со спутника HIMAWARI-8, и созданными для ее анализа "виртуальными" продуктами. Рассказано о разработанном способе получения контуров пепловых шлейфов от эксплозивных извержений с использованием данных различных спутниковых систем.

Во втором параграфе представлены возможности блоков анализа и обработки информации ДЗЗ в информационной системе дистанционного мониторинга лесов Приморского края (<http://primorsky.geosmis.ru>). Система ориентирована на комплексный дистанционный мониторинг лесных ресурсов, в том числе вырубок, пожаров и их последствий, охотничьих ресурсов. В параграфе приводится пример созданной процедуры массового картографирования рубок для решения задачи контроля лесопользования.

В третьем параграфе обсуждаются возможности блоков анализа и обработки информации ДЗЗ в спутниковом сервисе коллективного пользования "ВЕГА-Science" (<http://sci-vega.ru>), предназначенном для решения различных научных задач и используемом более чем 50 научными организациями. Сервис "ВЕГА-Science" сегодня является уникальной научной установкой, входящей в состав Центра коллективного пользования (ЦКП) "ИКИ-Мониторинг". Внедрение в него разработанных подходов позволило значительно повысить эффективность исследований с использованием информации ДЗЗ. Показано, что использование в системе разработанных подходов к работе с виртуальными продуктами позволило более чем на порядок сократить объемы хранимых данных, требующихся для эксплуатации ЦКП. Приведены примеры возможностей созданных в системе интерактивных инструментов для работы с разными типами спутниковых данных, в том числе радиолокационных. Рассказано о реализованных на базе возможностей ЦКП интерактивных инструментах анализа состояния сельскохозяйственных земель и культур, разработанных в том числе для решения задач глобального сельскохозяйственного мониторинга в рамках инициативы GEOGLAM.

В четвертом параграфе отмечается, что в настоящее время на основе результатов, полученных в диссертации, созданы блоки обработки и анализа в более чем десяти действующих прикладных и научных информационных системах дистанционного мониторинга. В том числе в таких крупных, как ИСДМ-Рослесхоз (<https://nffc.aviales.ru>), VEGA-GEOGLAM (<http://vega.geoglam.ru>), Объединенной системе работы с данными центров НИЦ "Планета" Росгидромета (<http://moscow.planeta.smislab.ru>) и других.

В заключении представлены результаты, выносимые на защиту.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Проведен анализ развития спутниковых систем ДЗЗ и подходов к организации работы с информацией, поступающей от них, который позволил сформулировать основные требования к перспективным методам и технологиям работы с данными ДЗЗ.

2. На основе системного анализа особенностей хранения и предоставления информации ДЗЗ и перспектив их развития разработана модель данных для ведения сверхбольших распределенных систем хранения с возможностью эффективной работы с "виртуальными" (формируемыми по запросу) информационными продуктами.

3. На основе разработанной модели создана программная инфраструктура ведения сверхбольших распределенных архивов данных ДЗЗ, которая внедрена в различные научные и прикладные информационные системы дистанционного мониторинга.

4. На основе использования современных подходов работы со сверхбольшими архивами данных ДЗЗ разработаны новые методы построения человеко-машинных интерфейсов систем распределенной интерактивной обработки спутниковых данных.

5. Для реализации предложенных методов разработана программная инфраструктура, обеспечивающая работу систем распределенной обработки спутниковых данных, которая внедрена в различные научные и прикладные информационные системы.

6. На примерах использования разработанных подходов, методов и программных инфраструктур при разработке различных систем дистанционного мониторинга показана их актуальность, востребованность и эффективность.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в изданиях из списка ВАК:

1. **Кашицкий А.В.** Создание инструментов для удаленной обработки спутниковых данных в современных информационных системах / *Кашицкий А.В., Балашов И.В., Лупян Е.А., Толпин В.А., Уваров И.А.* // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 1. С. 156-170. (**Scopus**)

2. **Кашицкий А.В.** Технология создания инструментов обработки и анализа данных больших распределенных спутниковых архивов / *Кашицкий А.В., Лупян Е.А., Балашов И.В., Константинова А.М.* // Оптика атмосферы и океана. 2016. Т. 29. № 9. С. 772-777. DOI: 10.15372/AOO20160908. (**Scopus, WoS**)

3. **Кашицкий А.В.** Оптимизация интерактивных процедур картографирования гарей в информационных системах дистанционного мониторинга природных пожаров / *Кашицкий А.В., Лупян Е.А., Барталев С.А., Барталев С.С., Балашов И.В., Ефремов В.Ю., Стыценко Ф.В.* // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 4. С. 7-16. (**Scopus**)

4. *Прошин А.А.* Создание унифицированной системы ведения архивов спутниковых данных, предназначенной для построения современных систем дистанционного мониторинга / *Прошин А.А., Лупян Е.А., Балашов И.В., Кашицкий А.В., Бурцев М.А.* // Современные проблемы дистанционного зондирования Земли из космоса. 2016. Т. 13. № 3. С. 9-27. DOI: 10.21046/2070-7401-2016-13-3-9-27. (**Scopus**)

5. *Толпин В.А.* Формирование информационной базы спутниковых и наземных данных для отработки методик дистанционного мониторинга виноградарства в Республике Крым / *Толпин В.А., Рыбалко Е.А., Баранова Н.В., Кашицкий А.В., Лупян Е.А., Уваров И.А.* // Современные проблемы дистанционного зондирования Земли из космоса. 2017. Т. 14. № 1. С. 101-110. DOI: 10.21046/2070-7401-2017-14-1-101-110. (**Scopus**)

6. *Рыбалко Е.А.* Организация работы с данными наземных и дистанционных наблюдений для решения задач дистанционного мониторинга виноградников / *Рыбалко Е.А., Баранова Н.В., Лупян Е.А., Толпин В.А., Кашицкий А.В., Уваров И.А., Крашенинникова Ю.С., Иванченко В.И.* // Современные проблемы дистанционного зондирования Земли из космоса. 2016. Т. 13. № 1. С. 79-92. DOI: 10.21046/2070-7401-2016-13-1-79-92. (**Scopus**)

7. *Саворский В.П.* Возможности анализа гиперспектральных индексов в информационных системах дистанционного мониторинга семейства «Созвездие-Вега» / *Саворский В.П., Кашицкий А.В., Константинова А.М., Балашов И.В., Крашенинникова Ю.С., Толпин В.А., Маклаков С.М., Савченко Е.В.* // Современные проблемы дистанционного зондирования Земли из космоса. 2016. Т. 13. № 3. С. 28-45. DOI: 10.21046/2070-7401-2016-13-3-28-45. (Scopus)

8. *Лупян Е.А.* Информационная система комплексного дистанционного мониторинга лесов «Вега-Приморье» / *Лупян Е.А., Барталев С.А., Балашов И.В., Барталев С.С., Бурцев М.А., Егоров В.А., Ефремов В.Ю., Жарко В.О., Кашицкий А.В., Колбудаев П.А., Крамарева Л.С., Мазуров А.А., Оксюкевич А.Ю., Плотников Д.Е., Прошин А.А., Сенько К.С., Уваров И.А., Хвостиков С.А., Ховратович Т.С.* // Современные проблемы дистанционного зондирования Земли из космоса. 2016. Т. 13. № 5. С. 11-28. DOI: 10.21046/2070-7401-2016-13-5-11-28. (Scopus)

9. *Гордеев Е.И.* Информационная система VOLSATVIEW для решения задач мониторинга вулканической активности Камчатки и Курил / *Гордеев Е.И., Гирина О.А., Лупян Е.А., Сорокин А.А., Крамарева Л.С., Ефремов В.Ю., Кашицкий А.В., Уваров И.А., Бурцев М.А., Романова И.М., Мельников Д.В., Маневич А.Г., Королев С.П., Верхотуров А.Л.* // Вулканология и сейсмология. 2016. № 6. С. 1-16. DOI: 10.7868/S0203030616060043. (Scopus, WoS)

10. *Гордеев Е.И.* ИС «VolSatView»: комплексный анализ данных об эксплозивных извержениях вулканов Камчатки / *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., Кашицкий А.В., Верхотуров А.Л., Бурцев М.А.* // Вестник Дальневосточного отделения Российской академии наук. 2016. № 5. С. 397-410.

11. *Гордеев Е.И.* Изучение продуктов извержений вулканов Камчатки с помощью гиперспектральных спутниковых данных в информационной системе VolSatView / *Гордеев Е.И., Гирина О.А., Лупян Е.А., Кашицкий А.В., Уваров И.А., Ефремов В.Ю., Мельников Д.В., Маневич А.Г., Сорокин А.А., Верхотуров А.Л., Романова И.М., Крамарева Л.С., Королев С.П.* // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 1. С. 113-128. (Scopus)

12. *Лупян Е.А.* Создание технологий построения информационных систем дистанционного мониторинга / *Лупян 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., Толпин В.А., Уваров И.А., Флитман Е.В.* // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 5. С. 53-75. (Scopus)

13. *Лупян Е.А.* Центр коллективного пользования системами архивации, обработки и анализа спутниковых данных ИКИ РАН для решения задач изучения и мониторинга окружающей среды / *Лупян Е.А., Прошин А.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., Толпин В.А., Уваров И.А.* // Современные про-

блемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 5. С. 263-284. (Scopus)

14. *Бурцев М.А.* Система работы с распределенными архивами результатов обработки спутниковых данных центров приема НИЦ "Планета" / *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.* // Современные проблемы дистанционного зондирования Земли из космоса. 2012. Т.9. № 5. С. 55-76.

15. *Гирина О.А.* Извержения Северной группы вулканов Камчатки 14–18 июня 2017 года / *Гирина О.А., Лупян Е.А., Мельников Д.В., Маневич А.Г., Кашницкий А.В., Бриль А.А., Сорокин А.А.* // Современные проблемы дистанционного зондирования Земли из космоса. 2017. Т. 14. № 3. С. 317-323. DOI: 10.21046/2070-7401-2017-14-3-317-323. (Scopus)

Кашницкий Александр Витальевич
МЕТОДЫ, МОДЕЛИ И ПРОГРАММНЫЕ ИНФРАСТРУКТУРЫ
ГЛОБАЛЬНО РАСПРЕДЕЛЕННОЙ ОБРАБОТКИ
АРХИВНЫХ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ

Автореферат диссертации на соискание ученой степени
кандидата технических наук

Подписано в печать _____

Формат А5

Бумага офсетная. Печать цифровая.

Тираж 150 Экз. Заказ No _____

Типография ООО "МДМпринт"

(Печатный салон МДМ)

119146, г. Москва, Комсомольский пр-кт, д.28

Тел. 8-495-256-10-00