

Построение архивов результатов обработки спутниковых данных для систем динамического формирования производных информационных продуктов

И.В. Балашов, М.А. Бурцев, В.Ю. Ефремов, Е.А. Лупян,
А.А. Прошин, В.А. Толпин

*Институт космических исследований РАН
117997, Москва, ул. Профсоюзная, 84/32
E-mail: info@smis.iki.rssi.ru*

В статье рассматриваются различные вопросы, связанные с построением архивов спутниковых данных, ориентированных на динамическое формирование информационных продуктов. Приводится общая постановка задачи построения таких архивов и выработанные нами пути ее решения. Более подробно освещены вопросы, связанные с масштабированием системы архивации и доступа к спутниковым данным, в частности, рассматриваются вопросы построения распределенных архивов спутниковых данных. В заключении приводятся примеры интерфейсов доступа к данным, реализованные в рамках системы ИСДМ «Рослесхоз».

В настоящее время к информационным системам спутникового мониторинга предъявляются высокие требования к организации доступа к спутниковым данным и продуктам их тематической обработки. Они должны обеспечивать доступ к спутниковым изображениям по произвольной области интереса в удобной для пользователя географической проекции. Важным требованием является также возможность динамического формирования необходимого пользователю изображения на основе данных, содержащихся в архиве. Это позволяет существенно увеличить набор информационных продуктов доступных пользователю. Таким образом, могут быть получены «на лету» различные композитные изображения, статистические характеристики изображений и т.п. И что особенно важно, динамическое формирование спутниковых изображений позволяет не только обеспечить пользователю доступ к архивам данных, но и предоставить ему различные инструменты для их анализа. Отметим, что вышеописанная функциональность интерфейсов доступа к спутниковым изображениям присутствовала ранее только в специализированных ГИС системах.

Естественно, что задача построения систем доступа к спутниковым данным неотделима от задачи организации архивов спутниковых данных. До недавнего времени архивы спутниковых данных, как правило, содержали изображения по фиксированному набору регионов наблюдения, что позволяло обеспечивать доступ к спутниковым изображениям по предопределенным географическим областям в фиксированной проекции. Это, в первую очередь, было необходимо для решения задач мониторинга отдельных регионов. Однако, такой подход не позволяет эффективно реализовать доступ к произвольной области изображения в заданной пользователем проекции, что затрудняет создание систем доступа к спутниковым данным глобального уровня, например, по территории России. Это приводит к необходимости пересмотра организации хранения спутниковых данных в архиве.

Для построения систем доступа к спутниковым данным, удовлетворяющих вышеперечисленным требованиям, был выработан новый подход к организации хранения данных в архиве [1]. Он основывается на том, что изображения в архиве хранятся в виде фиксированного разбиения на гранулы небольшого размера, общего для всех продуктов. При этом, для организации доступа к изображениям в произвольном масштабе, используются дополнительные разбиения на гранулы, формирующие пирамиду разрешений. Необходимое пользователю изображение в этом случае динамически формируется на сервере на основе соответствующих гранул нужного масштаба. Реализация новой схемы организации архивов спутниковых данных потребовала определенной адаптации системы обработки спутниковых данных. Теперь в рамках этой системы готовые спутниковые изображения разбиваются на гранулы базового масштаба и уже в таком виде поступают на архивацию. На этапе архивации на основе гранул базового масштаба получают также грану-

лы дополнительных масштабов, формирующих пирамиду разрешений.

Нами также был рассмотрен вариант с использованием индивидуального разбиения на гранулы для каждого спутникового изображения. Этот подход, на наш взгляд, целесообразно использовать в том случае, когда разбиение на гранулы реализуется в рамках самого графического формата, применяемого для хранения спутниковых изображений. В настоящее время в качестве такого формата реально может быть использован, пожалуй, только формат GeoTIFF. Разбиение изображения на гранулы, формирующие пирамиду разрешений, реализовано в рамках стандартной реализации этого формата. Использование формата GeoTIFF позволяет упростить реализацию архива, однако, при этом, накладывает определенные ограничения на область его применения. В частности, не может быть эффективно реализована модификация спутникового изображения большого размера, что легко достигается при использовании явного разбиения на гранулы. Необходимость в модификации изображения, в частности, возникает при поэтапном получении «глобальных» композитных изображений, когда требуемое изображение не может быть получено сразу целиком, а создается «по кусочкам». Кроме этого, в меньшей степени, могут быть оптимизированы различные задачи по динамическому формированию изображений. В случае использования явного разбиения на гранулы, мы можем ограничить набор используемых в обработке гранул, например, на основе информации о наложении участвующих в композите изображений. На наш взгляд, использование формата GeoTIFF наиболее целесообразно для хранения разрозненных спутниковых изображений, в частности, для хранения данных высокого разрешения, когда отдельные изображения существенно разнесены по зоне видимости и времени, и, поэтому, задача получения единого «покрытия» не очень актуальна.

При построении архивов спутниковых изображений, естественно, важную роль играет правильный выбор используемого графического формата. Так как спутниковые данные зачастую имеют большой размер, то для их хранения могут быть использованы графические форматы, предусматривающие сжатие данных. Форматы без потери качества, такие как GIF, PNG и GeoTIFF с использованием LZW компрессии, целесообразно использовать в тех случаях, когда недопустимо даже малейшее ухудшение качества изображений, либо изображения представляют собой матрицу значений, например соответствующих вычисленному индексу вегетации NDVI. В остальных случаях для хранения растровых изображений, пожалуй, более подходят форматы с потерей качества. Самым популярным из таких форматов является JPEG, поддерживаемый также в рамках формата GeoTIFF. Этот формат при незначительном ухудшении визуального качества изображений позволяет уменьшить в разы объемы хранимых изображений. Важно отметить, что уменьшение объема изображений играет важную роль также и в увеличении скорости работы интерфейсов доступа к данным по сети Интернет. К сожалению, при использовании форматов с потерей качества возникает неопределенность с границами изображения, что не позволяет, в частности, корректно накладывать изображения друг на друга. Эта проблема решается путем создания дополнительного информационного продукта, реализующего маску «наличия» изображения, например в формате GIF. Однако даже с учетом использования этого дополнительного продукта, выигрыш в суммарном объеме данных остается очень существенным.

Представленный выше подход к организации архивов спутниковых данных открывает новые перспективы для решения задач масштабирования систем хранения и представления данных, поскольку в рамках его использования органически может быть построена система распараллеливания, как хранения, так и динамического формирования изображений. В частности, при хранении подмножеств данных на разных серверах, для динамического формирования изображения «на лету», на каждом из серверов, где хранятся данные, может быть «на лету» сформирован соответствующий фрагмент результирующего изображения, после чего отдельные фрагменты могут быть использованы для сборки итогового изображения. Этот подход аналогичен технологии MapReduce, используемой Google для построения поисковых систем. В случае работы с архивами спутниковых данных такой подход может быть, в первую очередь, использован для получения различных композитных изображений на основе данных, получаемых в территориально разнесенных

центрах приема и обработки спутниковой информации. Кроме этого, даже в рамках одного информационного центра, может быть реализовано хранение спутниковых продуктов на нескольких серверах, что позволяет повысить скорость работы интерфейса доступа к данным.

Для построения высокопроизводительных систем доступа к спутниковым данным на основе представленного подхода могут быть использованы современные кластерные технологии. В этом случае, на наш взгляд, целесообразно использование «параллельных» файловых систем, наподобие Lustre [2]. В рамках таких файловых системы файлы или их фрагменты распределяются на дисковых подсистемах всех узлов кластера, а специальное программное обеспечение позволяет работать с ними, как если бы они находились в одном дисковом пространстве. В этом случае, естественным путем распараллеливания обработки, необходимой для получения результирующего изображения, является обработка каждой из задействованных гранул на отдельном компьютере. После этого изображения, полученные на основе отдельных гранул должны быть «склеены» на некоем незадействованном в обработке гранул компьютере. Ключевым моментом оптимизации является обработка исходных гранул по возможности именно на тех компьютерах, на дисковых подсистемах которых они расположены.

Частным, но очень важным случаем масштабирования архивов спутниковых данных является построение распределенных архивов спутниковых данных, включающих в себя данные, принимаемые и обрабатываемые в территориально разнесенных центрах приема спутниковых данных. Системы доступа к таким архивам должны обеспечивать возможность получения спутниковых изображений в независимости от того, в каком центре приема и обработки они физически находятся. Существенно, также, что запрошенное пользователем изображение может быть получено на основе данных, физически располагаемых на различных серверах. Для эффективной работы такой системы доступа должен быть реализован единый каталог данных, в котором бы указывалось, на каком из территориально разнесенных серверов находятся те или иные данные. Для этой цели нами был выработан нижеописанный подход.

Для удобства пользователей полнофункциональный интерфейс доступа к распределенному архиву устанавливается в каждом из центров приема и обработки спутниковых данных. Это приводит к необходимости ведения в каждом центре единого каталога данных получаемых во всех центрах приема. В каждом из центров в архив поступают продукты обработки спутниковых данных, принимаемые на локальной станции приема. После занесения данных в архив, информация о данных (без самих изображений) передается на сервера остальных центров приема и обработки спутниковых данных. Таким образом, БД в центре приема, соответствующая архиву спутниковых данных, содержит описания данных, находящихся во всех центрах приема и обработки спутниковых данных. В рамках работы интерфейса доступа запрос на получение спутникового изображения переадресуется на сервер, на котором физически хранится это изображение. Таким образом, пользователь получает доступ к любым данным в независимости от того, где они физически находятся.

В случае, когда пользователя интересует композитное изображение, каждый из фрагментов запрашивается с соответствующего сервера, после чего полученные фрагменты «склеиваются» в результирующее изображение. При этом для повышения скорости доступа с серверов могут быть запрошены не полные изображения по интересующей пользователя области, а только те их фрагменты, которые войдут в результирующее изображение. В частности, нами реализован вариант, когда на этапе подготовки композитного изображения с серверов запрашиваются изображения с указанием списка гранул, которые должны использоваться при его получении. На рисунке 1 приведена функциональная схема процедуры выполнения запроса на получение композитного изображения по данным распределенного архива спутниковых данных. На схеме обычными стрелками указаны запросы на получение данных, а полыми стрелками показаны сами потоки данных. Таким образом, по запросу пользователя локальный сервис предоставления изображений запрашивает информацию из локального экземпляра единого каталога данных. После получения информации, запросы на получение необходимых фрагментов направляются на соответствующие

информационные узлы. Сформированные на этих узлах изображения отсылаются на первый узел, где формируется интересующее пользователя результирующее изображение.

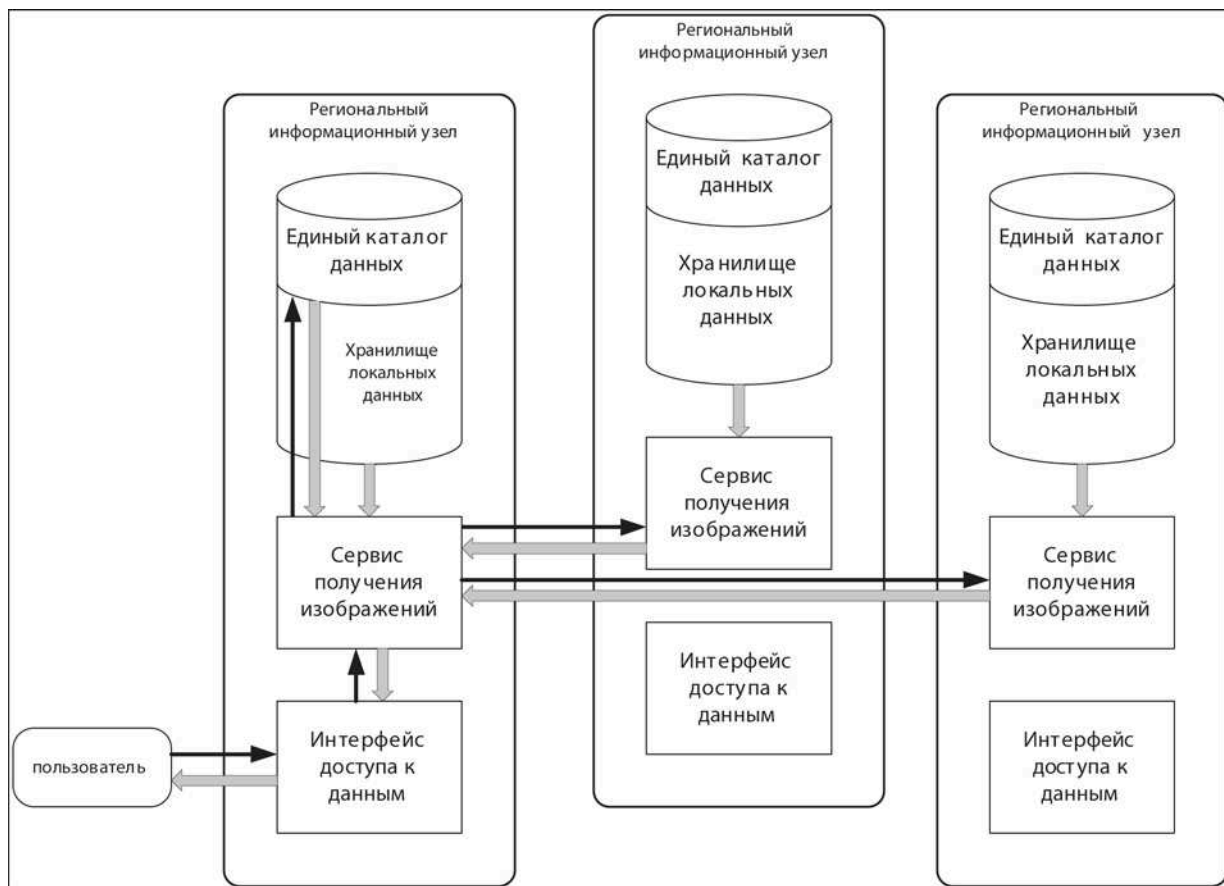


Рис. 1. Схема работы интерфейса получения композитного изображения по данным распределенного архива

Представленный в настоящей статье подход к построению архивов спутниковых данных для систем динамического формирования производных продуктов в настоящее время успешно нами используется при построении интерфейсов к данным различных информационных систем. Ниже, в качестве примеров, приведены две системы доступа, реализованные в рамках информационной системы мониторинга лесных пожаров Федерального агентства лесного хозяйства (ИСДМ «Рослесхоз») [3-4]: это система доступа к распределенному архиву результатов обработки спутниковых данных приборов AVHRR, MODIS, SPOT VGT и система доступа к распределенному архиву спутниковых данных высокого разрешения. Отметим, что приведенные ниже системы доступа к архивам построены в соответствии с технологией построения автоматизированных систем хранения спутниковых данных, разработанной в ИКИ РАН [5].

Система доступа к распределенному архиву результатов обработки данных приборов AVHRR, MODIS, SPOT VGT обеспечивает оперативный доступ к различным информационным продуктам, используемым для мониторинга лесных пожаров. В настоящее время в рамках системы ИСДМ «Рослесхоз» такие данные принимаются в семи территориально разнесенных центрах приема и обработки спутниковых данных, а также поступают из архивов компании VITO. В предыдущей реализации системы доступа архив каждого центра приема содержал информацию только о локальных данных. В результате для доступа к данным разных станций приема пользователю надо было обращаться к WEB серверам соответствующих станций, что, естественно, было не совсем удобно. Для того чтобы пользователь мог получить сразу доступ к данным всех центров приема, в соответствии с подходом, описанным в настоящей статье, была реализована единая система доступа к распределенному архиву данных

среднего разрешения. Хранение данных в архиве реализовано в виде гранул, по описанной выше технологии. Пример работы интерфейса этой системы приведен на рисунке 2. В левом фрейме выведен список сеансов со всех станций за указанную дату, попавших в интересующую пользователя географическую область. В правом фрейме отображаются различные графические информационные продукты. При этом отметим, что пользователь имеет возможность выбрать тип продукта, произвольный географический регион и набор слоев, необходимых ему для работы. Пользователь также может осуществлять масштабирование продуктов, детально анализируя те или иные районы.

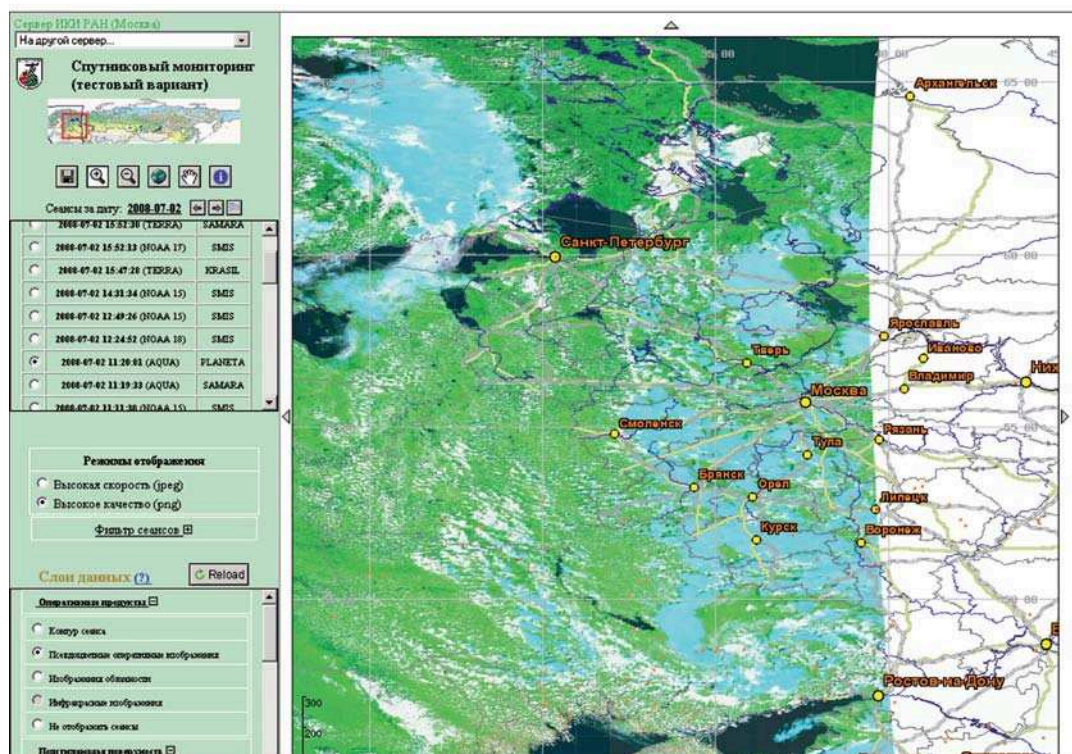


Рис. 2. Интерфейс системы доступа к распределенному архиву данных среднего разрешения (ИСДМ «Рослесхоз»)

Система доступа к данным высокого разрешения обеспечивает, в первую очередь, доступ к данным спутников SPOT, принимаемых на постоянной основе в трех центрах приема и обработки спутниковых данных, работающих в интересах ИСДМ «Рослесхоз». Система также обеспечивает доступ к архивам данных Ресурс и Landsat, полученных в различное время ФГУ «Авиалесоохрана». Данные высокого разрешения в основном используются для оценки последствий пожаров. В настоящее время рассматриваемый архив данных высокого разрешения содержит псевдоцветные изображения, полученные по сеансам спутников SPOT. При этом пользователи имеют возможность работы в полном пространственном разрешении для сцен, облачность которых не превышает 50%. Хранение изображений реализовано в виде файлов в формате GeoTIFF. Ниже на рисунке 3 приведен пример работы с интерфейсом доступа к данным высокого разрешения. В левом фрейме внизу выведен список сцен со всех центров приема за указанный диапазон дат по заданной пользователем географической области. В правом фрейме на выбранной сцене данных спутника SPOT хорошо виден пожар и его последствия.

Описанный в настоящей статье подход к построению архивов спутниковых данных, ориентированных на динамическое формирование продуктов, показал свою результативность. В то же время, мы постоянно работаем над его развитием, поэтому его нельзя рассматривать как что-то окончательно сложившееся. Одним из основных направлений его развития является масштабирование системы архивации и доступа к данным с целью увеличения скорости работы интерфейсов

доступа к данным и наращивания их функциональности. В частности, изучается вопрос использования для этих целей различных кластерных технологий.

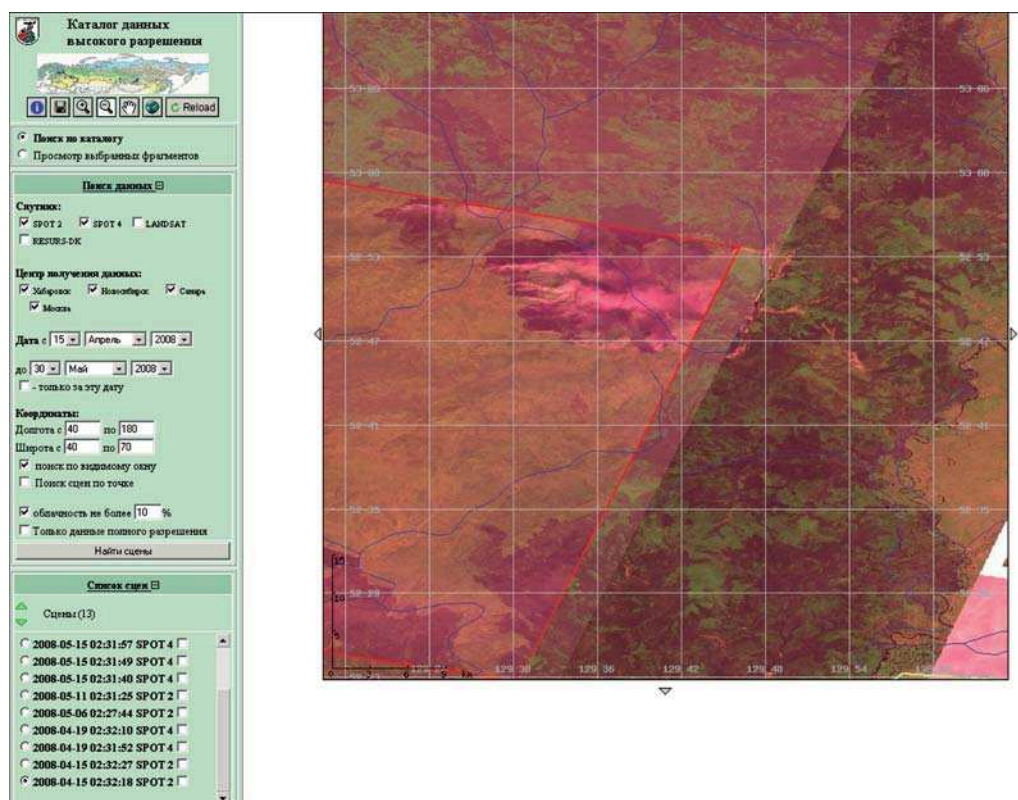


Рис. 3. Интерфейс доступа к распределенному архиву данных высокого разрешения (ИСДМ «Рослесхоз»)

Литература

1. Ефремов В.Ю., Крашенинникова Ю.С., Луян 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: Физические основы, методы и технологии мониторинга окружающей среды, потенциально опасных объектов и явлений. Сборник научных статей. М.: ООО «Азбука-2000», 2007. Выпуск 4. Т. 1. С. 125-134.

2. <http://www.lustre.org>

Беляев А.И., Коровин Г.Н., Луян Е.А. Использование спутниковых данных в системе дистанционного мониторинга лесных пожаров МПР РФ // Современные проблемы дистанционного зондирования Земли из космоса: Физические основы, методы и технологии мониторинга окружающей среды, потенциально опасных объектов и явлений. Сборник научных статей. М.: GRANP polygraph, 2005. Т. 1. С. 20-29.

3. Ершов Д.В., Коровин Г.Н., Луян Е.А., Мазуров А.А., Таццилин С.А. Российская система спутникового мониторинга лесных пожаров // Современные проблемы дистанционного зондирования Земли из космоса: Физические основы, методы и технологии мониторинга окружающей среды, потенциально опасных объектов и явлений. Сборник научных статей. М.: ООО "Полиграф сервис", 2004. С. 47-57.

4. Ефремов В.Ю., Луян Е.А., Мазуров А.А., Прошин А.А., Флитман Е.В. Технология построения автоматизированных систем хранения спутниковых данных // Современные проблемы дистанционного зондирования Земли из космоса: Физические основы, методы и технологии мониторинга окружающей среды, потенциально опасных объектов и явлений. Сборник научных статей. М.: ООО "Полиграф сервис", 2004. С.437-443.