

Построение систем, обеспечивающих динамическое формирование комплексных информационных продуктов на основе данных дистанционного зондирования

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

Институт космических исследований РАН

117997 Москва, ул. Профсоюзная, 84/32

E-mail: ivbalashov@smis.iki.rssi.ru

В последнее время, в связи с расширением областей использования спутниковой информации, возникла потребность в достаточно большом числе различных продуктов на основе данных дистанционного зондирования Земли. Это приводит к тому, что фактически взрывным образом растут объемы архивов спутниковых данных и результатов их обработки. В то же время, во многих случаях, целые классы информационных продуктов могут достаточно быстро готовиться на основе некоторых базовых заготовок, полученных в результате обработки спутниковых данных. Динамическое формирование производных информационных продуктов позволяет уйти от необходимости организации их хранения и существенно сократить объемы архивов. Для построения систем, обеспечивающих динамическое формирование комплексных информационных продуктов, необходимо использовать специальную технологию организации хранения и предоставления информации, обсуждению которой и посвящена настоящая статья.

Ключевые слова: дистанционное зондирование Земли, обработка спутниковых данных, комплексные информационные продукты, системы хранения и предоставления спутниковой информации.

Введение

В последние годы, в связи с расширением областей использования спутниковых данных, возникла потребность в достаточно большом числе различных продуктов получаемых на их основе. Следует отметить, что во многих случаях, в частности, для систем мониторинга, необходимо регулярное, а иногда и ежедневное получение таких продуктов. Это приводит к тому, что фактически взрывным образом растут объемы архивов спутниковых данных и результатов их обработки.

В то же время, во многих случаях, целые классы информационных продуктов могут достаточно быстро готовиться на основе некоторых базовых заготовок, полученных в результате обработки спутниковых данных. Простейшим примером здесь являются карты композитных изображений, построенные для различных регионов наблюдения. Для разных регионов эти карты могут достаточно быстро собираться из непересекающихся в пространстве фрагментов, и такая сборка может производиться на лету по запросу пользователей. Динамическое формирование производных информационных продуктов позволяет уйти от необходимости организации их хранения и существенно сократить объемы архивов. Кроме того, системы, обеспечивающие динамическое формирование производных продуктов (СДП), могут позволить организовать достаточно эффективную работу с распределенными вычислительными ресурсами, обеспечивающими хранение, обработку и представление данных. Это связано с тем, что в СДП не требуется постоянной синхронизации архивов, находящихся в разных центрах. Для таких систем требуется организация получения лишь фрагментов информации, необходимых для формирования конкретного запрошенного продукта.

Следует, однако, отметить, что для того, чтобы СДП можно было реально использовать, они должны обеспечивать достаточно быстрое формирование и представление информации по запросу пользователя. Поэтому, при построении таких систем, необходимо использовать специальные технологии и подходы к организации ведения архивов данных, формирования и предоставления информационных продуктов. Обсуждению элементов таких технологий, подходов и примеров их реализации и посвящена настоящая работа.

Архитектура системы

Безусловно, одним из ключевых вопросов при построении СДП является вопрос организации хранения и доступа к информации базовых продуктов, на основе которых производится динамическое формирование производных продуктов. На наш взгляд, наиболее эффективно в этом случае использовать подход, основанный на организации хранения базовых продуктов в виде гранул, содержащих информацию о географически непересекающихся областях. Это позволяет, с одной стороны, избежать ненужного дублирования информации и сократить объемы хранимой информации, а с другой обеспечить возможность однотипного построения продуктов по любой территории, на которую ориентирована система.

Идея такого хранения результатов обработки спутниковых данных не нова. Такой подход, в частности, используется в архивах центра LP DAAC (<http://lpdaac.usgs.gov>), входящего в состав проекта NASA EOSDIS, для хранения данных прибора MODIS, установленного на спутниках TERRA и AQUA [1], <http://lpdaac.usgs.gov>. Однако, достаточно долго он использовался только для минимизации объемов хранимой информации. Организация гранул и схема доступа к хранящейся в них информации была достаточно затратной. В случае же, когда стоит задача оперативного использования информации в СДП, должен быть реализован эффективный механизм доступа к информации, хранящейся в гранулах. При этом, должен быть выбран оптимальный размер гранул, информация в них должна иметь удобную структуру для быстрого чтения, в частности, должны быть организованы слои информации с различным уровнем пространственного разрешения. Подобный подход к ведению архивов использован, в частности, в Google Earth, (<http://earth.google.com>). Вопросы, связанные с организацией архивов спутниковых данных для СДП достаточно подобно рассмотрены также в [1,2].

Для реализации СДП, безусловно, недостаточно только организовать систему оптимального хранения базовых продуктов обработки спутниковых данных. Необходимо еще организовать автоматическое пополнения этих архивов, собственно динамическое формирование производных продуктов и представление их пользователям. Принципиальная схема блоков входящих в СДП, приведена на рисунке 1.

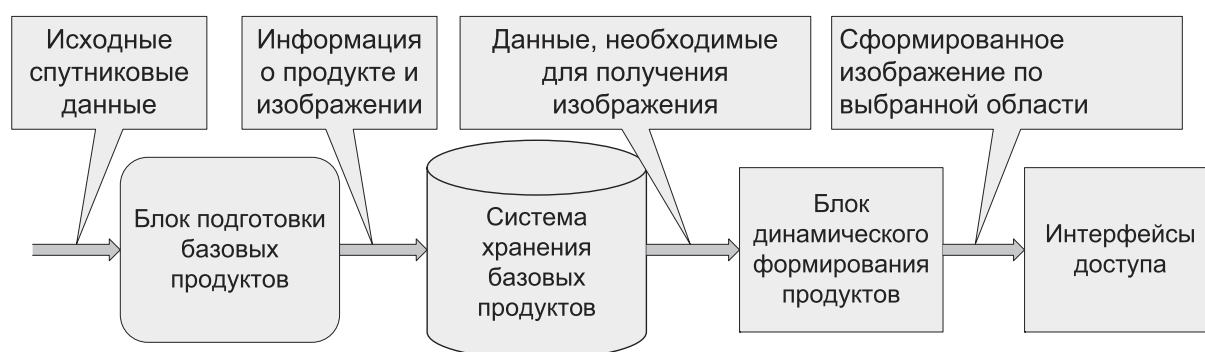


Рис.1. Принципиальная схема организации системы динамического формирования производных продуктов

Следует отметить, что функциональные блоки системы могут быть реализованы достаточно независимо, и использовать для работы различные вычислительные мощности, в том числе и распределенные. Это особенно важно, поскольку в рассматриваемой задаче мы имеем дело с большими объемами данных, которые принципиально не могут быть собраны, на одном, пусть и достаточно мощном вычислительном комплексе. Так, например, во многих случаях, особенно когда мы имеем дело с оперативными задачами, информация, необходимая для их решения, получается и обрабатывается в территориально распределенных центрах и не может быть оперативно собрана в единый архив данными из-за своих больших объемов. В этом случае, система динамического формирования продуктов должна обеспечивать возможность работы с территориально разнесенными блоками архива. Система должна обеспечивать генерацию продуктов на основе данных получаемых из различных блоков архива. Возможная схема реализации такой системы представлена на рисунке 2.

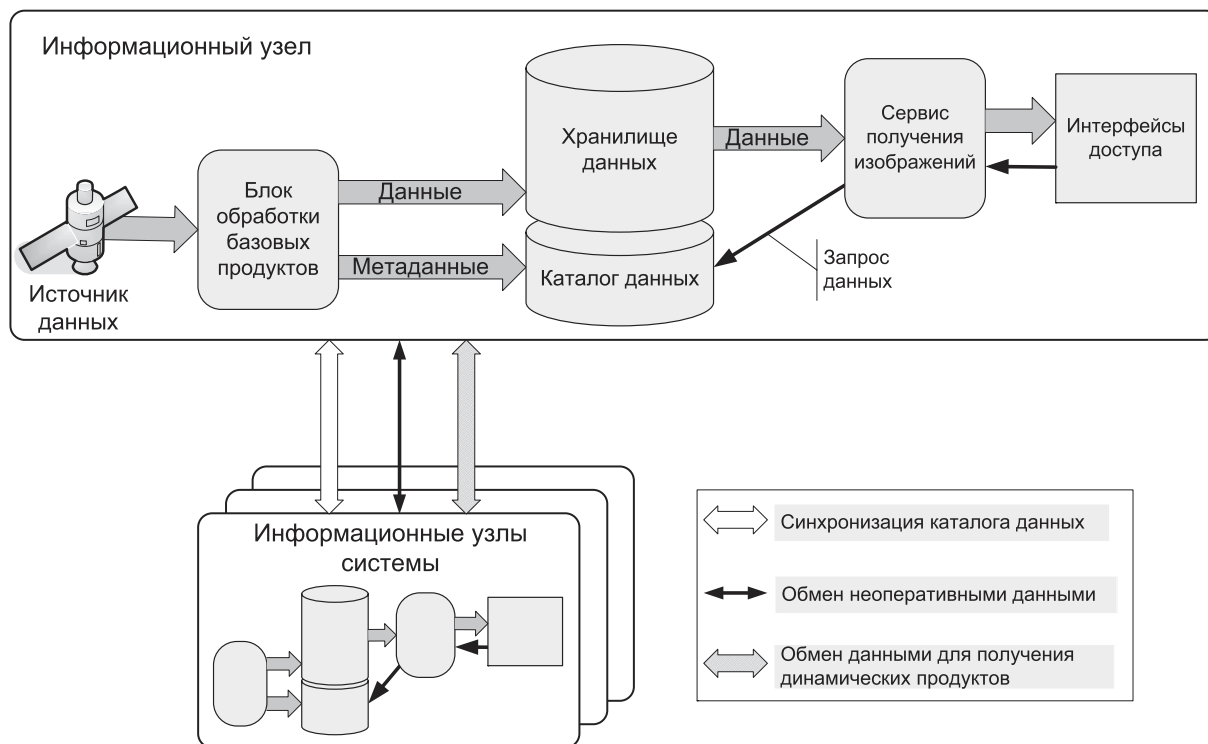


Рис. 2. Распределенная система динамического предоставления спутниковых данных

Базовым элементом данной схемы, фактически, является единый каталог данных, в котором хранится информация (метаданные) о состоянии архивов на всех узлах системы. По мере поступления информации в любой узел системы происходит ее обработка и наполнение локального архива узла, а все данные о поступивших изменениях отсылаются в центральный каталог. Поскольку передаваемые в каталог метаданные имеют небольшой размер, единый каталог данных имеет достаточно небольшой объем и может автоматически реплицироваться на все узлы системы, для использования в локальных блоках представления данных. Кроме того, во всех узлах системы должны быть организованы блоки, обеспечивающие динамическую подготовку продуктов на основе имеющихся в узлах архивов базовых продуктов. При поступлении запроса на подготовку конкретного продукта на основе информации, имеющейся в едином каталоге, запросы на производство фрагментов продукта направляются на соответствующие узлы системы. В узлах системы блоки подготовки продуктов формируют «свои» фрагменты и отсылают их либо на сервера – диспетчеры, где происходит «склейка» фрагментов в единый продукт, либо пользователю, в случае, если «склейка» фрагментов происходит на уровне интерфейсов работы с данными. Таким образом, пользователь получает доступ к любым данным в независимости от того, где они физически находятся.

Реализации системы

Безусловно, реализация систем динамического формирования информационных продуктов сильно зависит от конкретных задач, которые перед ними ставятся, типов данных, на которые они рассчитаны, видов информационных продуктов, требований к уровню оперативности и т.д. В то же время, для различных блоков, входящих в СДП, могут быть созданы достаточно стандартизованные базовые элементы, которые позволят сократить различные расходы при реализации конкретных систем. Такие базовые элементы были разработаны в ИКИ РАН на основе комплексов, использующихся для построения автоматизированных систем сбора, обработки, архивации и распространения спутниковых данных (см., в частности [3-5]). Эти комплексы использовались при построении различных систем дистанционного мониторинга. Комплексы, в частности, использовались при построении систем сбора, обработки, архивации и распространения данных в большинстве крупных российских центрах приема и обработки спутниковых данных.

Отдельным вопросом реализации является разработка блока использования распределенных источников данных. Такой блок должен обеспечивать быстрое, что особенно актуально при динамическом формировании производных продуктов, получение информации из различных систем хранения. Одной из составляющих быстрого получения продуктов из большого числа источников является параллельное выполнение запросов. В ИКИ РАН был разработан модуль `fork_module`, обеспечивающий параллельное получение спутниковых данных с распределенных серверов хранения и предоставления информации. Другой составляющей, позволяющей быстро формировать производные продукты на основе распределенной информации, является оптимизация состава используемых узлов на основе данных о параметрах каналов связи между узлами и вычислительных ресурсах системы, а так же специфических для каждого информационного продукта показателей качества фрагментов, хранимых на узлах распределенной системы. Для реализации этой составляющей необходима подсистема, обеспечивающей оперативный мониторинг указанных показателей и предоставление интерфейса доступа к ним. Часть функций такой подсистемы была реализована в ИКИ в виде модуля `speed_monitoring`, ведущего базу данных по параметрам скорости каналов связи между узлами системы и загруженности серверов. Этот модуль позволил оптимизировать формирование производных информационных продуктов на основе распределенных данных по времени.

В качестве базовой СУБД при построении комплексов используется MySQL. Для организации архивов данных технология FDB (File Data Base), созданная в ИКИ РАН [4] на языке программирования Си. Базовые WEB интерфейсы реализованы на основе CGI скриптов, разработанных на языке программирования perl. В качестве WEB сервера используется ПО Apache. Для предоставления географически привязанных растровых и векторных данных используются программный пакет `Mapserver`, разрабатываемый в университете штата Миннесота и специализированная библиотека `MapImage`, разработанная в ИКИ РАН. Для высокопроизводительного склеивания отдельных фрагментов растровых продуктов применяются графические программные пакеты GD и GDAL. На клиентской стороне используются возможности языка сценариев JavaScript и, разработанные на нем, библиотеки для обработки событий и проведения запроса к БД без перезагрузки страницы (технология Ajax), а также свободно распространяемая библиотека `wz_dragdrop` (Walterzorn Drag'nDrop DHTML Library) для работы с графическими слоями.

Примеры реализации

Описанный выше подход и разработанные базовые элементы используются в настоящее время для построения систем работы с архивами обработки спутниковых данных и различных систем дистанционного мониторинга, активно развивающихся в России в настоящее время. В частности, на основе них реализованы системы удаленного доступа к данным системы дистанционного мониторинга лесных пожаров Федерального агентства лесного хозяйства (ИСДМ

Рослесхоз) [6] (<http://www.nffc.aviales.ru>) и системы дистанционного мониторинга земель Министерства сельского хозяйства РФ (СДМЗ МСХ РФ) [7] (<http://www.agrocosmos.gvc.ru>).

В качестве примера, иллюстрирующего использование предлагаемого подхода, рассмотрим реализацию блоков, обеспечивающих формирование композитных изображений, получаемых на основе данных разновременных наблюдений. Эти блоки сегодня активно используются в ИСДМ Рослесхоз и СДМЗ МСХ РФ.

Блоки используют оперативные, автоматически пополняющиеся архивы данных, расположенные в центрах приема, работающих в интересах ИСДМ Рослесхоз и СДМЗ МСХ РФ. При поступлении на один из серверов этих систем, запрос на получение композита по конкретному региону, исходя из наличия данных в архивах узлов систем мониторинга, автоматически определяется набор информации, который необходим для построения композита и, исходя из наличия информации и степени доступности отдельных серверов, запрашиваются фрагменты необходимого композита. Для повышения скорости получения информации от конкретных серверов могут быть запрошены не полные изображения по интересующей пользователя области, а только фрагменты, пересечение которых с данными предоставленными другими серверами будет минимальным. В частности, реализован вариант, когда на этапе подготовки композитного изображения с серверов запрашиваются изображения с указанием списка гранул, которые должны использоваться при его получении. Выполнение запросов к различным узлам системы на получение фрагментов информации происходит параллельно, с использованием разработанных модулей. После получения сервером «диспетчером» фрагментов информации из различных узлов системы, они склеиваются в единое изображение, на которое наносится также информация различных информационных слоев, выбранных пользователем, в том числе, векторными информационными продуктами, в необходимом порядке. Схема выполнения запросов на получения фрагментов продуктов и подготовки итогового продукта приведена на рис. 3.

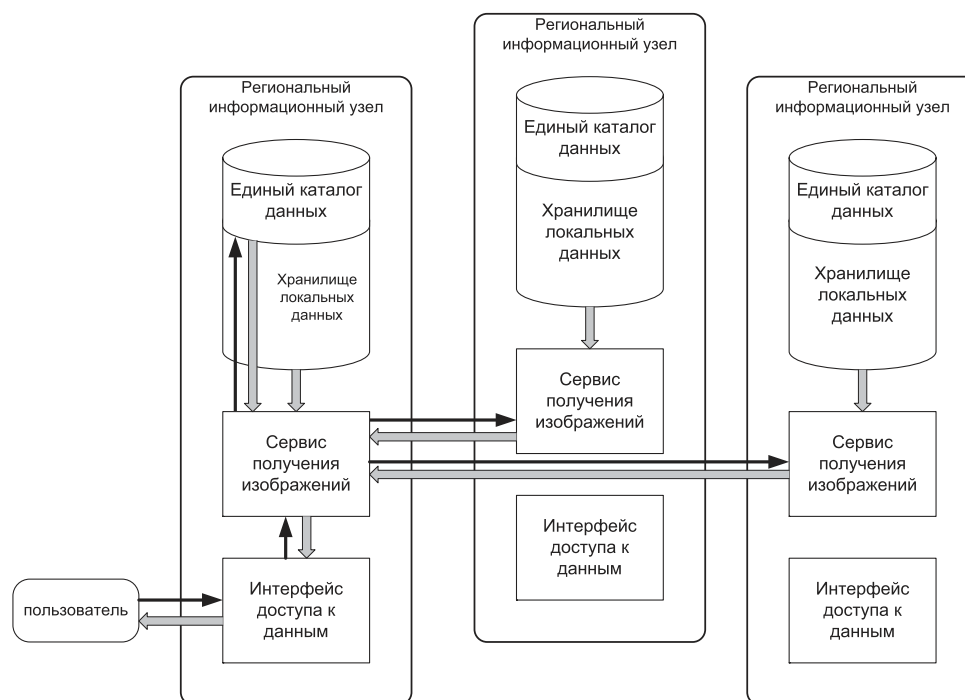


Рис. 3. Схема работы интерфейса получения композитного изображения по данным распределенного архива

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 узел «диспетчер», где формируется интересующее пользователя результирующее изображение.

Для обеспечения работы с динамическими продуктами для удаленных пользователей в системах реализованы специальные интерфейсы. Эти интерфейсы, в частности, позволяют выбрать тип продукта, область по которой он должен быть сформирован, время, за который формируется продукт, список различных слоев, которые будут использованы при его формировании (различные картографические слои, информация о пожарах, горях, повреждениях лесного покрова) и т.д. Следует отметить, для интерфейсов, предназначенных для работы с различными типами информационных продуктов, например, оперативными данными и данными архивов высокого разрешения, использована одинаковая схема реализации. Пример интерфейса, предназначенного для работы с оперативными данными, получаемыми в ИСДМ Рослесхоз на основе информации приборов AVHRR (спутники NOAA) и MODIS (спутники TERRA/AQUA) приведен на рис. 4. С примерами интерфейсов, обеспечивающих работу с данными высокого пространственного разрешения и описанием их основных характеристик, можно ознакомиться в работе [8].

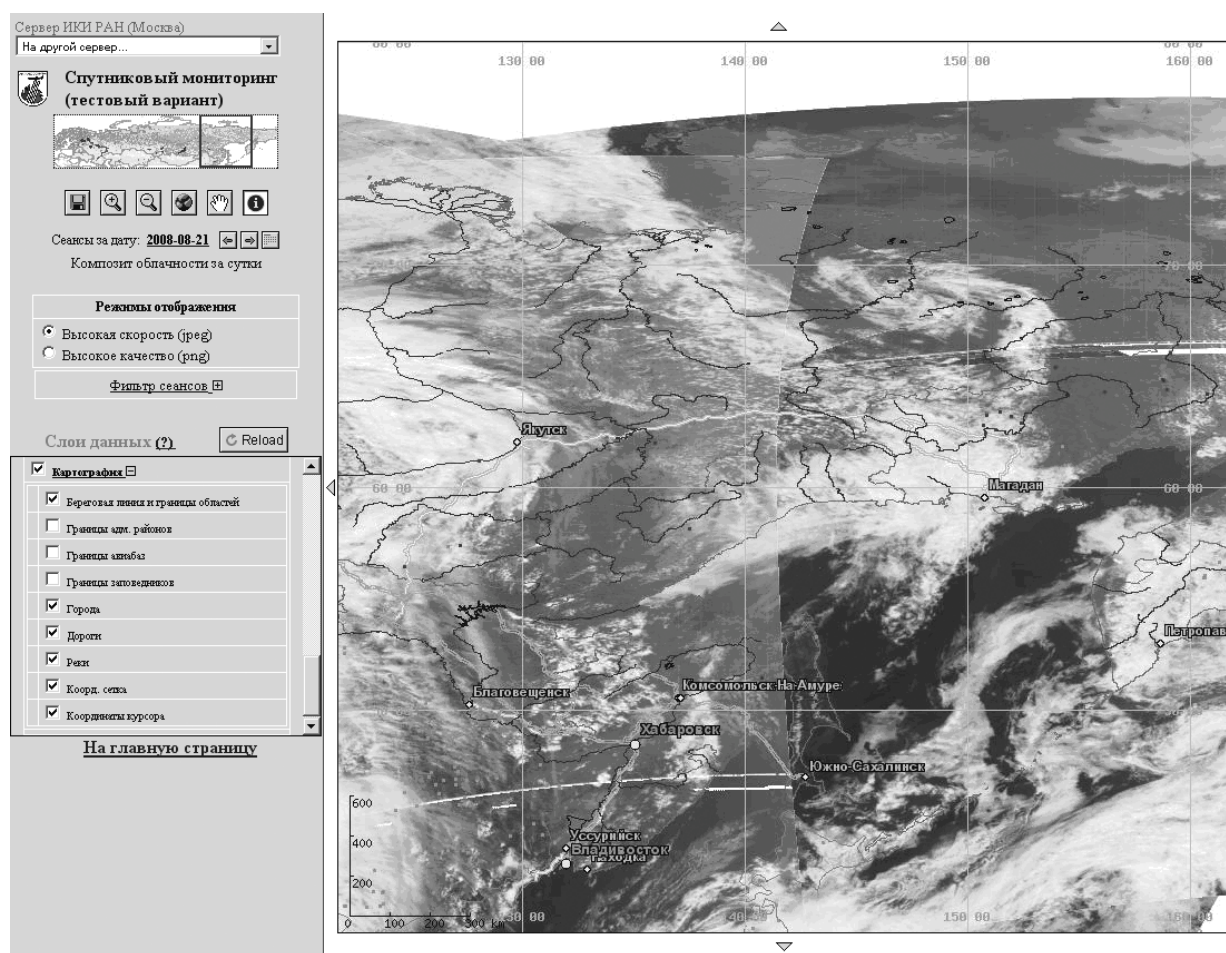


Рис. 4. Внешний вид динамического интерфейса ИСДМ «Рослесхоз»

Заключение

Опыт применения рассмотренного в настоящей статье подхода к построению систем динамического формирования комплексных информационных продуктов показал, что данный

подход, а так же созданные элементы для его реализации, могут быть использованы при решении достаточно широкого спектра задач, связанных с хранением и обработкой спутниковых данных.

Основными достоинствами описанного подхода можно считать:

- Гибкость и масштабируемость при разработке систем хранения спутниковых данных;
- Возможность применения распределенных схем хранения данных, а так же, кластерных технологий.

В настоящее время проводится ряд работ по совершенствованию рассмотренной в настоящей статье технологии, в частности:

- проводится оптимизация программного обеспечения для повышения скорости работы СДП;
- разрабатывается блок динамического преобразования растровых данных в различные географические проекции;
- ведутся работы по созданию системы параллельного выполнения операций подготовки динамических продуктов, в том числе, с возможностью использования кластерных технологий.

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 явлений и объектов. Сб. научных статей. Выпуск 4. М.: ООО «Азбука-2000», 2007. Т. 1. С. 125-132.

2. *Балашов И.В., Бурцев М.Ю., Ефремов В.Ю., Луян 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: Физические основы, методы и технологии мониторинга окружающей среды, потенциально опасных явлений и объектов. Сб. научных статей. Выпуск 5. М.: ООО «Азбука-2000», 2008. Т.1. С. 26-31.

3. *Луян Е.А., Мазуров А.А., Назиров Р.Р., Прошин А.А., Флитман Е.В.* Технология построения автоматизированных информационных систем сбора обработки и хранения спутниковых данных для решения научных и прикладных задач // В кн.: Аэрозоли Сибири / Под ред. К.П. Куценого. Новосибирск: Изд-во СО РАН, 2006. С. 458-470.

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. Сб. научных статей. М.: ООО "Полиграф сервис", 2004. С. 437-443.

5. *Bourtsev M., Loupian E., Mazurov A., Nazirov R., Proshin A., Flitman E.* Technology of automated satellite data archiving and representing systems development // Proceedings of the 7th International Symposium on Reducing Costs of Spacecraft Ground Systems and Operations (RCSGSO), 11-15 June 2007, Moscow, Russian Federation (ESA SP-648, July 2007) p. 3b2-1 - 3b2-4.

6. *Беляев А.И., Коровин Г.Н., Луян Е.А.* Состояние и перспективы развития Российской системы дистанционного мониторинга лесных пожаров // Современные проблемы дистанционного зондирования Земли из космоса: Физические основы, методы и технологии мониторинга окружающей среды, потенциально опасных объектов и явлений. Сб. научных статей. М.: ООО «Азбука-2000», 2006. Т. I. С. 341-350.

7. *Барталев С.А., Бурцев М. А., Еришов Д.В, Ефремов В.Ю., Ильин В.В, Луян Е.А., Мазуров А.А., Мельник Н.Н., Нейштадт И.А., Полищук А.А., Столпаков А.В., Прошин А.А., Темников В.А., Флитман Е.В.* Система автоматизированного сбора, обработки и распространения спутниковых данных для мониторинга сельскохозяйственных земель // Современные проблемы

дистанционного зондирования Земли из космоса: Физические основы, методы и технологии мониторинга окружающей среды, потенциально опасных объектов и явлений. Сб. научных статей. М.: GRANP polygraph, 2005. Т. I. С.131-139.

8. Ефремов В.Ю., Луян Е.А., Мазуров А.А., Матвеев А., Прошин А.А., Флитман Е.В. Организация работы с данными высокого пространственного разрешения в информационной системе дистанционного мониторинга лесных пожаров // Настоящий сборник.

Development of remote sensing complex information products dynamic creation systems

I.V. Balashov, V.Yu. Efremov, E.A. Loupian, A.A. Proshin, V.A. Tolpin

Space Research Institute of RAS
E-mail: ivbalashov@smis.iki.rssi.ru

Due to expansion of the satellite information usage, there is a need for a sufficiently large number of different products based on remote sensing of the Earth. This actually leads to explosive-like grow of satellite data archives volume and their processing time. At the same time, in many cases, entire classes of data products can be produced quickly enough on some basic products derived from raw satellite data. On-the-fly creation of derivative products allows escape from the necessity of their storage and significantly reduces the size of archives. To build systems that provide the dynamic creation of complex information products, you can use a special technology of storing and providing information, discussion of which is devoted to this report.

Keywords: Earth remote sensing, satellite data processing, complex information products, storing and providing satellite data systems.