

Подход к предоставлению услуг по обработке спутниковых данных в Центре коллективного пользования регионального спутникового мониторинга окружающей среды ДВО РАН*

П.В. Бабяк, И.В. Недолужко, Е.В. Фомин

Институт автоматики и процессов управления ДВО РАН
paulb@dvo.ru, ilya@dvo.ru, john@satellite.dvo.ru

Аннотация

Работа посвящена созданию инфраструктуры, реализующей концепцию заказа на обработку, в Центре коллективного пользования регионального спутникового мониторинга окружающей среды ДВО РАН. Сформулированы функциональные требования, предъявляемые к реализации. Рассмотрены основные подходы и приемлемые технические средства, приведены текущие результаты реализации инфраструктуры, включая интеграцию в среду Service Support Environment Европейского космического агентства (ESA).

1. Введение

1.1. Проблемы предоставления спутниковых данных пользователю

В настоящий момент спутниковые данные широко применяются в различных отраслях исследовательской и хозяйственной деятельности человека. Эффективность решения прикладных и теоретических задач напрямую зависит от развитости средств доступа потребителей к продуктам обработки спутниковых данных. Классическим является подход, при котором спутниковый центр, занимающийся приёмом и обработкой спутниковых данных, обеспечивает построение только стандартной продукции, ориентированной на типовые задачи пользователей. Данная продукция накапливается в архивах и доступна для загрузки в виде файлов или отображения в качестве слоёв ГИС-систем. Как правило, присутствует каталог метаописаний, позволяющих пользователю получить информацию о присутствующих в архиве продуктах или коллекциях продуктов [10].

Недостатком классического подхода (обработка и поставка стандартной продукции) является его плохая ориентированность на исследовательские задачи. Рост числа спутников и типов измерительной аппаратуры в сочетании с увеличением числа алгоритмов обработки данных приводит к невозможности обеспечить пользователей всеми видами продуктов спутниковых данных. Для получения одного и того же продукта спутниковых данных могут быть использованы различные методики и пакеты программ, а также определённые параметры обработки. При этом выбор наиболее подходящих средств обработки сам по себе может являться темой для отдельного исследования. Таким образом, пользователь должен иметь возможность самостоятельного управления обработкой спутниковых данных. К сожалению, установка и настройка большинства современных пакетов обработки (например, SeaDAS [14, 6]) требует от пользователя наличия специальных знаний и навыков. Кроме того, пользователь может не располагать достаточными вычислительными мощностями для оперативного решения задач.

Выходом из сложившейся ситуации является обработка спутниковых данных на стороне спутникового центра и предоставление пользователю простых средств для удалённого управления основными параметрами этой обработки [3]. При выработке требований к инфраструктуре, обеспечивающей реализацию концепции, необходимо учитывать как собственный опыт спутникового центра по работе с пользователями, так и основные мировые тенденции в области доступа к данным дистанционного зондирования Земли (ДЗЗ). Одним из важнейших направлений работ в этой области является создание глобальных распределённых систем, позволяющих пользователю получать информацию о наличии определённых данных у конкретных спутниковых центров, и получать эти данные - бесплатно или на коммерческой основе. Важной особенностью таких

Труды XIV Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2011), Санкт-Петербург, Россия, 2011.

систем является возможность пользователя получать доступ к интересующим его ресурсам через единый интерфейс, вне зависимости от физического расположения поставщиков. Наиболее развитые глобальные информационные системы разрабатываются и поддерживаются NASA (США) и ESA (ЕС). В последние годы в работы в данной области получили развитие и в России [5, 9, 2, 1].

1.2. Постановка задачи

Центр коллективного пользования регионального спутникового мониторинга окружающей среды ДВО РАН создан на базе лаборатории спутникового мониторинга ИАПУ ДВО РАН. Центр обеспечивает информационную поддержку фундаментальных и прикладных научных исследований в области физики океана и атмосферы, океанологии, морской биологии и экологии. В Центре ведётся круглосуточный приём, обработка и архивирование данных со спутников AQUA, TERRA, MTSAT-1R, FY-1D и серии NOAA,. Общий объём накопленных архивов по состоянию на май 2011 года превышает 10 Тб. Данные хранятся как на RAID-массиве общим объёмом 46 Тб, так и в виде резервной копии на носителях DVD. Для обработки информации используется как общепризнанные программные пакеты, так и программное обеспечение, разработанное специалистами Центра на основе международно признанных и собственных методик. На текущем этапе развития Центра интеграция в глобальные системы обмена спутниковой информацией является одной из наиболее приоритетных задач. Приоритетной глобальной системой для Центра является среда ESA SSE (Services Support Environment), основанная на технологии веб-сервисов [8].

Данная работа посвящена созданию инфраструктуры, реализующей концепцию заказа на обработку, в Центре коллективного пользования регионального спутникового мониторинга окружающей среды ДВО РАН.

2. Функциональные требования к инфраструктуре Центра

Подход к реализации предложенной концепции заказа на обработку предъявляет следующие базовые требования к инфраструктуре Центра:

- необходимо предоставить внешний доступ к ресурсам Центра как на верхнем (через веб-интерфейс), так и на нижнем уровне (удалённый вызов процедур, в т. ч. в автоматическом режиме);
- должна существовать возможность интеграции механизмов доступа в одну и более глобальных информационных систем (в первую очередь - среду SSE);
- для работы с файлами стандартной продукции, хранящимися в архиве Центра, пользователю должны быть предоставлены следующие возможности:

- поиск на основе набора критериев;
- просмотр метаинформации и обзорных изображений для найденных файлов;
- выбор интересующих файлов из списка найденных;
- загрузка или визуализация данных (в браузере или с использованием внешних средств, таких как Google Earth);
- при заказе на обработку пользователь, дополнительно к указанным возможностям, получит:
 - возможность загрузки собственных данных для обработки;
 - заказ на обработку данных с указанием технологии и параметров обработки, формата выходных данных и способа предоставления результата (выбор из списков predetermined значений);
 - возможность отслеживания статуса обработки;
 - по завершению обработки - получение уведомления, содержащего ссылку для загрузки или непосредственной визуализации полученного продукта;
- доступ к ресурсам может быть ограничен и предоставляться как в режиме подписки, так и в режиме единовременного заказа;
- должны присутствовать механизмы разграничения прав на получение тех или иных данных, заказа определённого вида обработки и т. д.

3. Подход к реализации требований

3.1. Общая стратегия

Основной стратегией реализации инфраструктуры Центра, позволяющей обеспечить перечисленные функции, является применение существующих решений и минимизация трудозатрат на каждом этапе её развития. Существующие решения могут включать как распространённые программные пакеты, так и собственные разработки Центра. При этом каждый компонент системы имеет чётко обозначенную функцию и может быть заменён либо продублирован другим, в зависимости от решаемой задачи. На рис. 1 приведена общая функциональная схема инфраструктуры. Основу инфраструктуры которой составляют следующие компоненты:

- Архив спутниковых данных и стандартной продукции; репозиторий как интерфейс для взаимодействия с ним.
- Веб-сервисы каталогов коллекций и продуктов спутниковых данных, зарегистрированные на портале SSE.
- Гетерогенная распределённая система обработки спутниковых данных.
- Веб-сервисы нижнего уровня для доступа к ней, реализованный на базе технологии Web Processing Services.
- Веб-сервис верхнего уровня для заказа обработки спутниковых данных.

- Веб-интерфейс, реализованный на базе средств сервера SSE Portal (где должны быть зарегистрированы сервисы каталога и заказа).
- Средства для предоставления пользователю конечного результата (FTP, WMS/WCS/WFS).
- Система аутентификации для внешних интерфейсов.

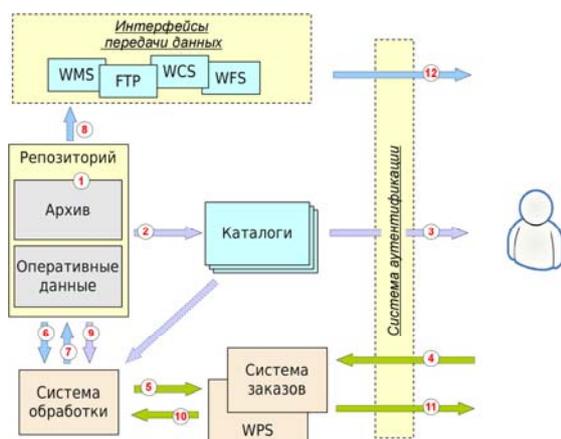


Рис. 1. Инфраструктура Центра

Рассмотрим типовой сценарий работы пользователя.

1. Хранимые спутниковые данные доступны через средства репозитория.
2. Из данных извлекается метаинформация, которая помещается в каталоги спутниковых данных.
3. Результатом поиска по каталогу являются идентификаторы коллекции и продуктов.
4. Информация о выбранных пользователем продуктах передаётся в систему заказа.
5. Система заказов формирует задание для распределённой системы обработки
6. В рамках системы заказов происходит запрос на извлечение нужных файлов по идентификаторам
7. Результат обработки помещается в репозиторий
8. Репозиторий сохраняет результат и делает его доступным через один из интерфейсов предоставления данных
9. Идентификатор результата сообщается системе обработки
10. Система обработки сообщает о завершении задания
11. Система заказов уведомляет пользователя
12. Пользователь получает доступ к результату обработки. Стандартная продукция Центра может быть доступна через те же механизмы изначально.

3.2. Репозиторий спутниковых данных

Опыт эксплуатации архива говорит о необходимости создания дополнительного уровня абстракции в задаче доступа к архиву, по следующим причинам:

- Программе, сохраняющей или извлекающей данные из архива необходимо иметь точную информацию о файловой структуре хранилища. В случае возможной реорганизации хранилища может потребоваться дополнительная настройка этих программ.
- Взаимодействие между программами обработки спутниковых данных, разрабатываемыми разными разработчиками, требует дополнительных соглашений на размещение файлов данных.
- Работа с внешним пользователем требует абстрагирования от понятия физического размещения файла, поскольку внутренняя структура хранилища по известным причинам скрыта от пользователя.

Вместо физического пути в репозитории используется идентификатор коллекции (определяющий тип файла) и идентификатор продукта (определяет файл или группу файлов). В настоящий момент в Центре реализован прототип репозитория. Его реализация включающий базу данных и набор утилит для сохранения и извлечения информации.

3.3. Каталоги спутниковых данных

Каталоги традиционно используются как основа инфраструктур пространственных данных. К сожалению, значительное число существующих каталогов содержит весьма общее описание пространственных данных, а занесение новой метаинформации производится вручную. Это позволяет пользователю выйти на нужного поставщика данных либо, либо получить доступ к определённому ресурсу - однако не даёт возможности непосредственного ознакомления с содержимым архива. Такой уровень абстракции в среде SSE называется каталогом коллекций (либо сервисов - в зависимости от типа описываемых ресурсов). Он может быть реализован с применением популярного пакета Geonetwork Opensource [11].

Для описания отдельных продукта обработки спутниковых данных, как правило соответствующего отдельной сессии приёма спутниковых данных, в среде SSE используется понятие каталога продуктов. Для реализации каталога продуктов [8] был использован стандарт EOLI, используемый в ESA и предложенный в качестве минимального профиля OGC для поиска и заказа продуктов обработки спутниковых данных [12].

3.4. Обработка спутниковых данных

Для обеспечения оперативной обработки спутниковых данных и обработки по заказу в Центре разработана распределённая система обработки. Основной её задачей является интеграция в единую гетерогенную систему различных программных комплексов обработки данных, как разработанных в рамках Центра, так и пакетов сторонних производителей. Так для обработки данных MODIS используется пакет SeaDAS и IMAPP, а для обработки данных спутников серии NOAA используются пакеты AAPP, RTTOV, MetOffice 1D-VAR и пакет программного обеспечения разработанный в рамках центра.

Обработка данных со спутников Aqua и Terra производится с помощью программного комплекса SeaDAS, свободно-распространяемого NASA, позволяющего строить по данным радиометра MODIS множество видов продукции. Данный программный комплекс выбран для обработки данных радиометра MODIS потому, что широко распространён, активно развивается NASA и в котором широко представлены разнообразные алгоритмы тематической обработки данных цветности. Ещё одним преимуществом SeaDAS [14] является его модульная архитектура: алгоритмическая часть SeaDAS представляет собой набор командно-строчных утилит, а встроенный в него интерпретатор языка IDL позволяет создавать собственные программы обработки спутниковых данных.

В процессе опытной эксплуатации SeaDAS были выявлены определенные трудности: использование SeaDAS требует от оператора-обработчика не только знаний функциональных возможностей самого ПО, но и умения работать в командной оболочке Linux. При работе с SeaDAS в режиме потока значительное время уходит на повторение одних и тех же действий для каждого обрабатываемого файла, так как обработка каждого файла представляет собой последовательный вызов отдельных утилит с заданием параметров для каждой из них.

Обработка данных спутников серии NOAA представляет собой набор цепочек последовательного получения продуктов разнообразными программными средствами. Так, для получения профилей температуры и влажности атмосферы по данным радиометров ATOVS используются пакеты AAPP, RTTOV и MetOffice-1Dvar, разработанные в английском метеорологическом агентстве MetOffice. Для получения температурных полей океана, карт термических структур, скоростей поверхностных течений используется пакет прикладных программ разработанный в рамках центра. Данные геостационарного спутника MTSAT-1R используются для получения температурных полей океана, полей льда, потоков водяного пара.

Кроме того, получение некоторых продуктов обработки спутниковых данных является достаточно вычислительноёмкой задачей и для их получения в оперативном режиме вычислительных ресурсов центра явно недостаточно. Так расчёт ДОТК (доминантных ориентаций термических контрастов) по одному району в зависимости от параметров может занимать более 24-х часов, а расчёт тепловых профилей и профилей влажности по региону за месячный период занимает 3-4 дня. Поэтому ещё одной задачей создаваемой информационной системы является интеграция с ЦКП «Дальневосточный Вычислительный Ресурс» (ДВВР) [4] для получения доступа к его вычислительным ресурсам (рис. 2).

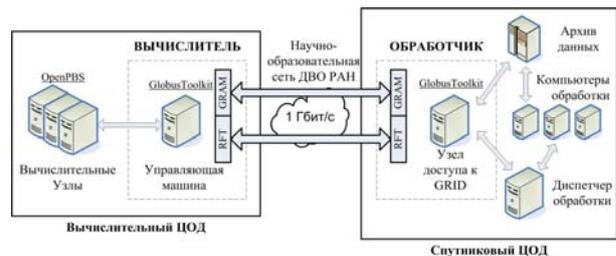


Рис. 2. Интеграция центров обработки данных ЦКП ДВВР и ЦКП РСМ ОС ДВО РАН.

3.5. Веб-сервисы обработки и заказа

Имея в наличии развитые средства для обработки спутниковых данных, необходимо обеспечить доступ к последним. Наиболее популярным специализированным средством такого доступа является интерфейс Web Processing Service (WPS) Открытого геопространственного консорциума (OGC). Данный интерфейс может быть реализован на основе протоколов HTTP и SOAP (в том числе и в виде веб-сервиса) и позволяет организовать синхронный и асинхронный вызов обработки спутниковых данных с указанием параметров. Поверх реализованного таким образом веб-сервиса может быть реализован веб-сервис более высокого уровня, являющийся интерфейсом заказа спутниковых данных среды SSE [13]. Сервис заказа допускает задание дополнительных параметров, которые могут быть использованы для выбора способа и параметров обработки. Сервис заказа может быть интегрирован с каталогом продуктов, что позволяет пользователю осуществлять заказ обработки данных, выбранных по каталогу. Отдельная подсистема должна обеспечивать аутентификацию и авторизацию пользователей для обоих видов сервисов.

В рамках решения данной задачи в Центре был реализован прототип нового интерфейса OGC 06-141 v1.0.4 Draft, используемого для заказа спутниковых данных в среде SSE [7].

Заключение

Описанный подход даёт возможность создать в Центре инфраструктуру, реализующую обработку спутниковых данных в соответствии с параметрами, задаваемыми внешним пользователем. При этом доступ пользователя к ресурсам такой инфраструктуре может быть организован как через портал среды SSE Европейского космического агентства, так и в автоматическом (на уровне вызова веб-сервисов). Отдельные компоненты (службы обработки, каталог продуктов) описанной инфраструктуры уже функционируют в оперативном режиме, ряд компонентов (каталог коллекций, репозиторий, система заказов) требует доработки для соответствия обозначенным требованиям. Основными текущими направлениями разработки являются создание репозитория и системы заказов на обработку, реализующей взаимодействие с пользователем через интерфейсы SSE (напрямую или через портал).

Литература

- [1] Антонов, А.В. Построение объединенного каталога распределенных архивов спутниковых данных различных центров / Антонов А.В., Бурцев М.А., Ефремов В.Ю., Крамарева Л.С., Калашников А.В., Крашенинникова Ю.С., Лупян Е.А., Матвеев А.М., Прошин А.А., Флитман Е.В. // Современные проблемы дистанционного зондирования Земли из космоса. 2010. Т.7. № 2. С.84-89.
- [2] Атаева, О.М. Портал интеграции пространственных данных «Геомета» / Атаева О.М., Кузнецов К.А., Серебряков В.А., Филиппов В.И. // Труды 12й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2010, Казань, Россия, 2010.
- [3] Бабяк, П.В. Система удаленного доступа и управления обработкой спутниковой информации / Бабяк П.В., Недолужко И.В. // Тезисы докладов Дальневосточной математической школы-семинара имени академика Е.В. Золотова, Владивосток, 6-11 сентября 2004. Владивосток: Изд-во ДВГУ, 2004. С. 141-142.
- [4] Бабяк, П.В. Опыт использования Grid-Технологий В Системе Обработки Данных Спутникового Центра ДВО РАН / Бабяк П.В., Тарасов Г.В. // Современные проблемы дистанционного зондирования Земли из космоса. 2009. Т. 6. № 1. С. 71-80.
- [5] Белов, С.В. Интеграция информационных ресурсов в задачах исследования морской среды / Белов С.В., Бритков В.Б. // Информационные технологии и вычислительные системы. 2008. № 1. С. 73-81.
- [6] Левин, В.А. Разработка технологий спутникового мониторинга окружающей среды по данным метеорологических спутников / Левин В.А., Алексанин А.И., Алексанина М.Г., Дьяков С.Е., Недолужко И.В., Фомин Е.В. // Открытое образование. 2010. № 5. С. 41-49.
- [7] Недолужко, И.В. Заказ спутниковых данных среды SSE: реализация в ЦКП регионального спутникового мониторинга окружающей среды ДВО РАН // Современные проблемы дистанционного зондирования Земли из космоса. 2010. Т. 7, № 2. С. 561-567.
- [8] Недолужко, И.В. Интеграция ресурсов Центра коллективного пользования регионального спутникового мониторинга окружающей среды ДВО РАН в среду SSE Европейского космического агентства // Вычислительные технологии. 2010. Т. 15, № 4. С. 116-130.
- [9] Шокин, Ю.И. Распределенная информационно-аналитическая система для поиска, обработки и анализа пространственных данных / Шокин Ю.И., Жижимов О.Л., Пестунов И.А., Синявский Ю.Н., Смирнов В.В. // Вычислительные технологии. 2007. Т. 12. Специальный выпуск 3. ГИС- и веб-технологии в междисциплинарных исследованиях. Материалы Междисциплинарной программы СО РАН 4.5.2. Выпуск I. С. 108-115
- [10] Шокин, Ю.И. Распределенная информационная система сбора, хранения и обработки спутниковых данных для мониторинга территорий Сибири и Дальнего Востока / Шокин Ю.И., Пестунов И.А., Смирнов В.В., Синявский Ю.Н., Скачкова А.П., Дубров И.С., Левин В.А., Алексанин А.И., Алексанина М. Г., Бабяк П.В., Громов А. В., Недолужко И.В. // Научный журнал Сибирского федерального университета, серия «Техника и технологии», 2008. Т. 1. № 4. С. 291-314.
- [11] GeoNetwork opensource [Электронный ресурс]. — Режим доступа: <http://geonetwork-opensource.org/>.
- [12] Minimal Profile for EO products using WSDL and SOAP [Электронный ресурс]. — Режим доступа: <http://earth.esa.int/XML/eoli/documents/EOPProfile.doc>.
- [13] Service Support Environment. Interface Control Document. [Электронный ресурс]. — Режим доступа: <http://services.eoportal.org/massRef/documentation/icd.pdf/>.
- [14] The SeaWiFS Data Analysis System (SeaDAS) [Электронный ресурс]. — Режим доступа: <http://seadas.gsfc.nasa.gov/>.

**An approach used to provide satellite data
processing services at Multiple Access
Centre for Regional Satellite Monitoring of
Environment FEB RAS**

P.V. Babyak, I.V. Nedoluzhko, E.V. Fomin

The paper deals with infrastructure that implements "ordering for processing" concept at Multiple Access Centre for Regional Satellite Monitoring of Environment FEB RAS. Functional requirements for the implementation are defined. Basic approaches and technical means are discussed, up-to-date results of the infrastructure implementation are presented, including Integration into Service Support Environment of ESA (European Space Agency).

* Работа поддержана грантом РФФИ 11-07-00511-а