

Перспективы интеграции Центра спутникового мониторинга окружающей среды ДВО РАН в среду поставки и обработки данных Европейского космического агентства

И.В. Недолужко

*Институт автоматизации и процессов управления ДВО РАН
690041 Владивосток, Радио 5
E-mail: ilya@satellite.dvo.ru*

В работе рассматривается сервис-ориентированный подход к предоставлению ресурсов спутниковых центров пользователям и его возможности при интеграции центров в единую сеть. Описана среда SSE Европейского космического агентства, построенная на основе технологии веб-сервисов. Приведён опыт интеграции Центра спутникового мониторинга окружающей среды ДВО РАН в SSE и рассмотрены её дальнейшие перспективы.

Введение

В настоящий момент интенсивный обмен данными является одной из основных особенностей деятельности организаций, занимающихся вопросами дистанционного зондирования. Необходимость данного обмена связана с расширением сферы применения спутниковых данных, ростом числа типов спутников и способов обработки спутниковой информации. Результатом такого развития является невозможность решения всех сопутствующих задач в рамках единой организации, что и приводит к задаче обмена.

К основным проблемам поставки спутниковых данных следует отнести необходимость согласования принципов обмена и применяемых технических средств. Следует учитывать, что участник может преследовать собственные цели, обладая при этом собственными ресурсами, источниками данных, исторически сложившимися методиками их обработки. Ряд задач требует оперативности поставки, при этом должны приниматься во внимание ограниченная пропускная способность сетей и большой объём спутниковых изображений. Одной из наиболее важных проблем является возможность организации автоматической поставки данных, пригодность которых для решения конкретной задачи определяется также автоматически.

Основой для эффективной работы со спутниковыми данными в настоящий момент являются метаданные (структурированная описательная информация). Метаданные могут быть использованы [1]:

- для оценки пригодности (типа данных, пространственно-временных характеристик и атрибутов, характеризующих способ получения данных);
- для оценки качества (наличие сбойной информации и специфичные для задачи атрибуты, такие как характеристики облачности) данных;
- как информация о способах хранения и доступа к данным.

Выделяют применение метаданных как для внешних (реклама ресурса и каталог данных), так и для внутренних нужд организации (эффективное хранение и доступ к данным) [http://www.eoportal.org/documents/Metadata-document-2_1.doc]. Назначение метаданных определяет используемый набор полей и правила их заполнения. Строгая стандартизация содержимого полей метаданных даёт возможность их применения в автоматических системах обработки спутниковой информации.

На настоящий момент не существует единого стандарта и подхода к предоставлению пользователю метаданных о спутниковых данных. Задача выработки и применения единых стандартов, позволяющих интегрировать информацию различных спутниковых центров в единую глобальную распределённую систему, до сих пор не является решённой в полной мере. Глобализация позво-

ляет, с одной стороны, обеспечить доступ пользователя к данным через единую точку доступа, вне зависимости от физического расположения поставщика данных. С другой стороны, поставщик данных получает в своё распоряжение способ экспорта своих ресурсов большему числу клиентов с использованием уже созданного программного обеспечения.

Сложившейся практикой по обмену данными является создание каталогов (электронных библиотек) на основе баз данных. Хранимые в них записи метаданных позволяют отобразить необходимые гранулы (файлы или наборы файлов) в рамках отдельной коллекции (наборов однотипных данных). Регистрация коллекций и поставщиков данных в рамках отдельных каталогов глобальных информационных систем даёт возможность распределённого поиска. Непосредственный доступ к ресурсам (например, получение файлов спутниковых данных), список которых пользователь получает в результате поиска, является наиболее ценной возможностью каталогов.

Протокол Z39.50 и унаследованные от него спецификации, широко использовавшийся ранее при создании распределённых информационных систем в задачах дистанционного зондирования, позволяет создание распределённых каталогов метаданных. Обеспечение распределённости может достигаться за счёт объединения групп баз данных под псевдонимом и переадресации запроса. В первом случае пользователь имеет возможность опроса единственной (фиктивной) базы данных, на самом деле являющейся списком отдельных баз данных. Переадресация позволяет серверу, при отсутствии у неё необходимой базы данных, выступить посредником и осуществить запрос к другому серверу, на котором требуемая база данных присутствует [2]. Для создания глобальной информационной системы по данным дистанционного зондирования CEOS (Committee on Earth Observation Satellites) был разработан протокол CIP (Catalogue Interoperability Protocol). Целью проекта являлось объединение каталогов метаданных различных стран, в том числе систем INFEO (ESA) и EOSDIS (NASA). Основой для интеграции является программный пакет, называемый шлюзом (gateway), в задачи которого входит трансляция CIP-запросов с запросы к СУБД.

Сервис-ориентированный подход к предоставлению ресурсов

Системы, основанные на протоколе Z39.50 существуют и продолжают развиваться [2, 3], однако в NASA и ESA было принято решение отказаться от него в пользу систем с сервис-ориентированной архитектурой. Так, для замены EDG (EOSDIS Data Gateway) предложено использовать клиенты ECHO и интерфейс TBD, основанный на технологиях XML и SOAP [http://edcdaac.usgs.gov/landdaac/presentations/sap_03-06/EOSDIS_Evolution_Willems.pdf]. В ESA для организации каталогов данных (<http://catalogues.eoportal.org>) была выработана спецификация EOLI-XML, в настоящий момент поддерживаемая организацией INTECS (Италия). Эта спецификация стала базовой для организации каталогов в системе SSE (Service Support Environment), активно развиваемой сейчас в Европе в качестве основы для интеграции служб наблюдения за Землёй. Основу среды, как и в системе ECHO, составляют службы (сервисы), основанные на технологии Web Services.

Отдельный сервис представляется в виде набора операций, каждая из которых может работать как синхронном (рис. 1), так и асинхронном режиме (рис. 2).

Синхронный режим подразумевает немедленный ответ сервера, тогда как при асинхронном используется отправка сообщения и его подтверждение – как со стороны клиента, так со стороны сервера. Для отсылки сообщений применяется протокол SOAP, основанный на формате XML. Синхронный вид взаимодействия, также называемый процедурным стилем SOAP, в SSE используется для создания каталогов. Асинхронный режим необходим в том случае, когда немедленное получение результата невозможно – при оформлении заказа на генерацию ещё не существующего продукта. При использовании асинхронного режима взаимодействия не исключается тот случай, когда часть операций выполняется персоналом организации. Это связано с тем, что время между сообщением подтверждения запроса и сообщением ответа технически никак не ограничивается. В общем случае, два описанных режима обмена сообщениями позволяют любое взаимодействие между сервис-провайдером (service-provider) и потребителем.

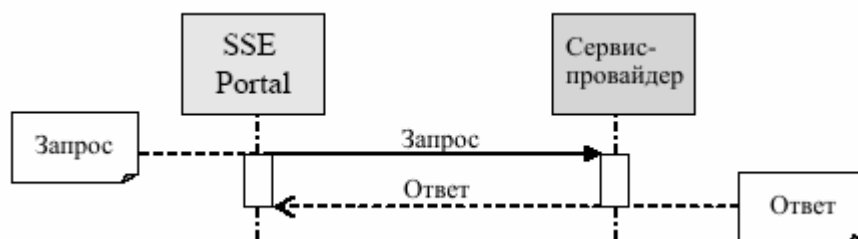


Рис. 1. Синхронный режим обмена сообщениями

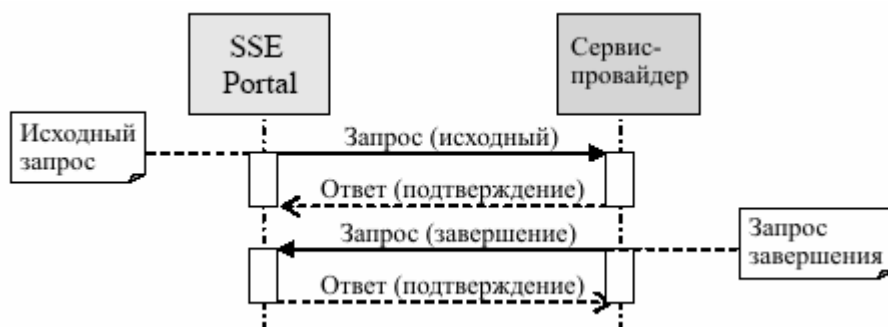


Рис. 2. Асинхронный режим обмена сообщениями

Сервис-ориентированный подход позволяет не только предоставить пользователю доступ к спутниковым данным, но и организовать механизмы удалённого вызова алгоритмов их обработки, функционирующих на стороне сервис-провайдера (спутникового центра). При данном подходе становится возможным предоставить пользователю механизмы управления обработкой, в то время как технические детали обработки и проприетарные алгоритмы могут быть скрыты. Не накладывается никаких ограничений не то, каким образом сервис-провайдер будет получать и обрабатывать информацию. Таким образом, используемые в обработке данные могут принадлежать сервис-провайдеру, либо получены от пользователя или другого сервис-провайдера. Более того, может происходить обращение к сервисам обработки других организаций. Сочленение веб-сервисов с использованием языка BPEL (Business Process Execution Language) позволяет создание полностью автоматических распределённых цепочек сбора, обработки спутниковой информации и принятия решений по её результатам. Такие цепочки могут функционировать как в рамках организации по локальной сети, так и на глобальном уровне через сеть Интернет, давая возможность международной кооперации.

Среда SSE Европейского космического агентства

Система SSE (Service Support Environment) является попыткой создания инфраструктуры, объединяющей организации стран Европы в единую информационную сеть, где каждый поставщик услуг равноправен. Разработка и реализация для ESRIN ведётся группой организаций под предводительством SPACEBEL (Бельгия), куда также входят GIM (Бельгия) и Intecs (Италия). Целью является реализация открытой сервис-ориентированной распределённой среды между потребителями и поставщиками, позволяющей интегрировать данные наблюдения за Землёй, метеорологические данные и данные геоинформационных систем. Прототип системы был отлажен в рамках проекта MASS-ENV, существующего с 2001 года. Основными задачами SSE являются [\[http://earth.esa.int/rtd/Documents/SSE_Whitepaper_2.pdf\]](http://earth.esa.int/rtd/Documents/SSE_Whitepaper_2.pdf) возможности:

- создания инфраструктуры, позволяющей взаимодействие между сервис-провайдерами (B2B) и с пользователями (B2C);

- сохранения базовых и сквозных служб на стороне сервис-провайдера;
- лёгкого подключения/отключения сервисов к/от среды;
- сочленения простых сервисов в составные;
- поддержки сервисов «с подпиской» (например, мониторинг пожаров и оповещение);
- развития и поддержки сервисов;
- простого поиска и доступа к требуемым сервисам, с возможностью наблюдения процесса вплоть до его завершения;
- интеграции сервисов из различных областей знаний.

Новым шагом развития системы является проект НМА, разрабатываемый 2006 года по инициативе GMES (Global Monitoring for Environment and Security) в сотрудничестве Комиссией европейских сообществ. НМА вводит ряд новых механизмов для обеспечения более тесной интеграции входящих в среду организаций. Подробный обзор назначения, архитектуры и технологической основы приведён в [http://epubs.cclrc.ac.uk/bitstream/1846/HMA_Final_Review_V10.pdf].

Центральным узлом среды является сервер SSE Portal (<http://services.eoportal.org>), предоставляющий пользователю веб-интерфейс для работы с системой. Предполагается создание и региональных порталов, построенных по той же технологии. Пользователь работает с инструментами портала, не задумываясь о расположении сервис-провайдера, ресурсы которого он использует. Для выдачи результатов могут быть использованы сервера WCS и WFS, HTTP или FTP доступ к файлам данных, или само SOAP-сообщение. На форматы файлов, доступных по FTP и HTTP специальных ограничений не накладывается.

SSE Portal преобразует пользовательские запросы в SOAP-сообщения, которые отсылаются к соответствующему сервис-провайдеру. Обратное взаимодействие реализовано также. Форматы сообщений и способы обмена задаются профилями, основанными на xsd-схемах. Тип профиля определяется назначением сервиса. Так, для создания каталогов могут быть использованы профили (предлагаются порталом на момент написания статьи) EOLI, EO Application Profile for CSW 2.0, ISO Application Profile for CSW 2.0, EO Products Extension Package for ebRIM и ряд других. Ряд этих профилей предложен в качестве стандартов OGC. Сервис-провайдер может выбрать любой из них, руководствуясь собственными потребностями. Портал предоставляет пользователю интерфейс для работы с любым из них, включая поиск по каталогу, оформление и отслеживание заказа и др. Следует также отметить, что требуемый для заполнения провайдером набор полей метаданных в большинстве из них значительно уже используемого в протоколе SIP. Такой подход существенно упрощает создание каталогов, в то время как ограниченный набор полей достаточен для отсеивания значительной части непригодных для применения пользователем спутниковых данных.

Основным программным средством для интеграции в среду является пакет SSE Toolbox, распространяемый по лицензии GNU General Public License. Пакет основан на технологии Java Enterprise Edition с применением сервера приложений Apache Tomcat. На сайте пакета (<http://toolbox.pisa.intecs.it>) доступны сборки для Windows и Linux. Toolbox предоставляет сервис-провайдеру инструменты для развёртывания веб-сервисов в соответствии с требованиями профилей и их локальной отладки. Ключевой особенностью пакета является использование специализированного языка скриптов, предназначенных для обработки SOAP-запросов. Возможен вызов других скриптов, написанных на встроенном языке, а также языке Jelly. Наиболее ценной возможностью является работа со стандартными или пользовательскими классами Java. В язык скриптов встроен набор функций для работы с базами данных через интерфейс JDBC, средства для работы с XML, SOAP, а также e-mail и FTP. Набор встроенных функций позволяет упростить процесс создания промежуточного программного слоя между SOAP-клиентом (в частности, сервером SSE Portal) и инфраструктурой сервис-провайдера, от которой не требуется никаких изменений.

Несмотря на то, что SSE разрабатывалась в первую очередь для решения задач, стоящих перед Европейскими службами наблюдения за Землёй, возможно участие организаций других

стран. Так, есть прецедент регистрации на портале сервисов ECHO (США) и JAXA (Япония). По состоянию на август 2007 на портале зарегистрировано свыше 500 пользователей и порядка 50 сервис-провайдеров. Зарегистрировано около 250 сервисов, треть из которых доступно для пользователей.

К недостаткам системы можно отнести отставание реализации части описанных в документации механизмов и протоколов, а также отсутствие единого профиля для каталогов. Все существующие недостатки можно объяснить быстрой модернизацией технической основы системы.

Интеграция Центра в систему SSE

Интеграция Центра коллективного пользования регионального спутникового мониторинга окружающей среды ДВО РАН в глобальные системы обмена данными является одной из наиболее приоритетных задач, наряду с включением в обработку новых типов данных, разработкой новых методик и обработки, расширением ассортимента получаемой продукции.

Система SSE позволяет решить задачи Центра по поставке данных на глобальном уровне и имеет возможность организации доступа к системе распределённой обработки спутниковых данных, в соответствии с развиваемой в Центре концепцией заказа на обработку. Преимуществом системы SSE для Центра является использование открытых и документированных стандартов и наличием готового программного обеспечения. При этом реализация пользовательского интерфейса для доступа к системе не требуется, так как большинство функций уже реализовано на самом портале.

Базовым видом интеграции в среду является сервис каталога данных. Наиболее устоявшейся спецификацией является EOLI-XML [<http://earth.esa.int/rtd/Documents/EOLI-XML-ICD.pdf>]. Этот профиль содержит минимальный набор полей и наиболее прост для реализации. Несмотря на то, что он не предусматривает заказ данных, поставка всё же возможна через задание URL файла наряду с другими полями метаданных. Таким образом, пользователь получить файл, найденный в каталоге.

Профиль EOLI-XML использует опыт таких стандартов как CIP, FGDC, GML, Open GIS и стандартов ISO. Набор полей метаданных определяется представлением (presentation) и назначением SOAP-сообщения. Хотя на поля метаданных накладываются строгие ограничения по типу, содержимое многих из них произвольно. Так, профиль EOLI-XML использует около двух десятков полей и список допустимых значений (Valid Values) описывает только четыре поля. Профиль включает в себя два операции: search и present. Первая предназначена для поиска по каталогу, вторая позволяет получить расширенную информацию о продукте (файле данных).

Для реализации каталога были созданы скрипты, взаимодействующие с существующей базой метаданных Центра. Произведена локальное тестирование каталога средствами Toolbox, после чего он был развёрнут на одном из серверов Центра. Были получены права сервис-провайдера на сервере SSE Test Portal, предназначенном для прототипирования. В настоящий момент созданный каталог для файлов данных спутников NOAA (рис. 3) интегрирован в тестовом режиме (недоступен для других пользователей). Ведётся работа по созданию и интеграции каталогов для данных спутников AQUA и некоторых видов стандартной продукции Центра.

Рассмотрим работу такой системы на примере созданного каталога данных NOAA (рис. 4).

1. Пользователь выбирает коллекцию; задаёт географический регион, используя механизм AOI (Area of Interest), предоставляемый порталом; указывает временной промежуток и запускает поиск.

2. SSE Portal формирует запрос EOLI search к серверу Центра

3. SSE Toolbox, функционирующий на сервере, вызывает скрипт, сконфигурированный для обработки запроса search

4. Скрипт производит действия по разбору параметров поиска и отправляет запрос к СУБД PostgreSQL, где функционирует база метаданных, используя интерфейс JDBC



Рис. 3. Использование интерфейса SSE Portal для доступа к каталогу данных Центра

5. На основе полученного результата формируется SOAP-сообщение и работа скрипта завершается
6. Toolbox отсылает результирующее сообщение на SSE Portal
7. Портал отображает информацию о найденных файлах

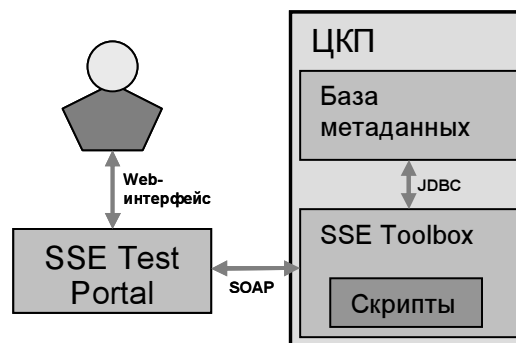


Рис. 4. Взаимодействие с каталогом данных Центра

При щелчке на файл пользователь получает более полную информацию, получение которой происходит через запрос present, обрабатываемый аналогичным образом. Обзорное изображение передаётся в виде динамической ссылки, обрабатываемой Java-сервлетом и извлекающим изображение. В подробную информацию может входить URL файла, используя который пользователь может его получить. Доступ к файлу, находящемуся на FTP-сервер можно ограничить, что даёт возможность коммерциализации поставки в режиме доступа по подписке.

В дальнейшие планы входит интеграция в среду системы распределённой обработки Центра, однако это требует пересмотра организации информационных потоков внутри Центра (рис. 5). При этом следует учитывать возможность поставки данных в другие системы, а также развитие самой среды. Наиболее важной задачей является развитие базы метаданных и интерфейсов доступа к ней через интерфейсы SSE и из распределённой системы обработки спутниковых данных, разрабатываемой в Центре. Необходима также модернизация системы заказов на обработку, с которой в настоящий момент работают пользователи Центра.

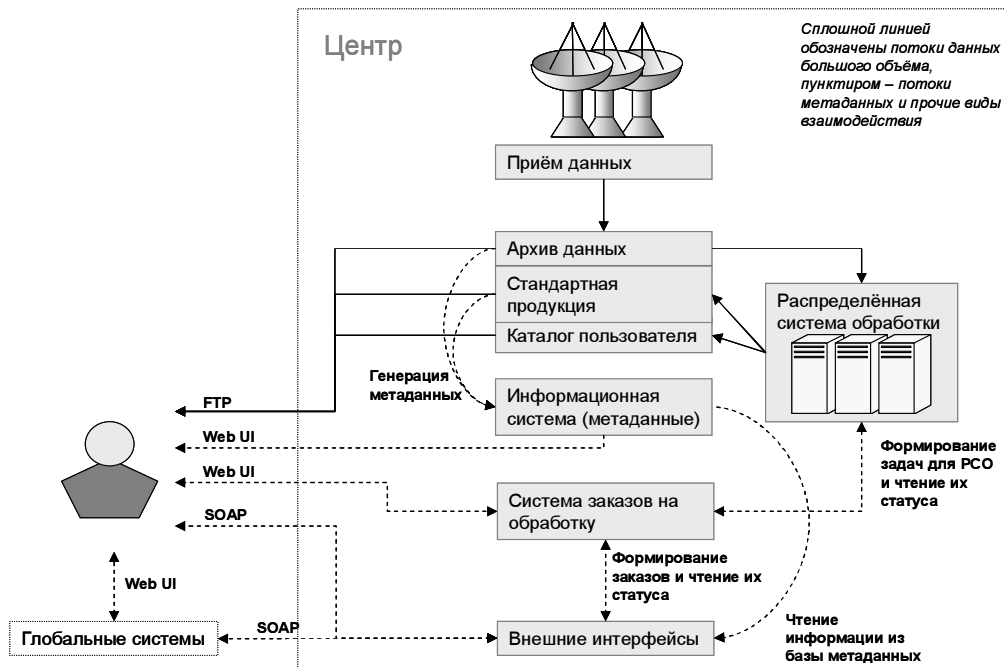


Рис. 5. Предложенная организация потоков внутри Центра

Выводы

Система SSE является наиболее предпочтительным выбором для интеграции, а используемые в её рамках технические средства могут быть применены для поставки данных и удалённого вызова обработки на ресурсах Центра в рамках других проектов. Полученный опыт свидетельствует о возможности такой интеграции, однако ставит перед Центром вопрос о реорганизации внутренних потоков работы с данными.

Литература

1. Саворский В.П., Лупян Е.А., Тищенко Ю.Г. Базовые спецификации запросов, обслуживаемых информационной системой космических данных // Современные проблемы дистанционного зондирования Земли из космоса: Физические основы, методы и технологии мониторинга окружающей среды, потенциально опасных объектов и явлений. Сб. научн. статей. М.: ООО "Полиграф сервис", 2004. С.482-490.
2. Жижимов О.Л., Мазов Н.А. Принципы построения распределённых информационных систем на основе протокола Z39.50 // Новосибирск: Изд-во ИВТ СО РАН, 2004. 361 с.
3. Шокин Ю.И., Жижимов О.Л., Пестунов И.А., Синявский Ю.Н., Смирнов В.В. Распределённая информационно-аналитическая система для поиска, обработки и анализа пространственных данных // 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.