

Технологии и стандарты интеграции сервисов, каталогов и баз данных дистанционного исследования Земли из космоса*

© Е.Б. Кудашев

А.Н. Филонов

Институт космических исследований РАН, Институт проблем информатики РАН

kudashev@iki.rssi.ru, filonov_a@inbox.ru

Аннотация

Статья посвящена вопросам интеграции в мировые системы спутниковых исследований Земли и, в первую очередь, интеграции информационных ресурсов ДЗЗ в систему EoPortal Европейского космического агентства (ESA) на основе современных стандартов OpenGIS и международных технологий обмена метаданными глобальных поисковых систем. Исследуются технологии интеграции, обеспечивающие возможность объединения информационных ресурсов космического экологического мониторинга в единую информационную среду.

1. Введение

В настоящее время развиваются системы, обеспечивающие отображение информационных ресурсов ДЗЗ и спутниковой информации экологического мониторинга в Интернет. В качестве примера можно привести разработки NASA-EOSDIS и европейскую информационную систему INFEO [1,2,3]. Многолетний опыт получения, хранения и обработки спутниковых данных показывает неуклонно возрастающее количество потенциальных пользователей спутниковой информации, и все большее разнообразие запросов каждого конкретного пользователя на тот или иной вид обработки спутниковых данных. Эта тенденция определяется появлением нового поколения спутников и спутниковых приборов, позволяющих пользователю привлекать для решения задач новые виды спутниковой информации.

Существует, однако, ряд причин, которые делают задачу предоставления пользователю спутниковой информации достаточно сложной. Главными из них являются огромные объемы этой информации и отсутствие информационной инфраструктуры. Разнообразие методов и

алгоритмов обработки спутниковой информации не позволяет каждому конкретному пользователю обрабатывать эту информацию самостоятельно в соответствии со своими требованиями

Значение этой проблемы будет со временем лишь возрастать и перспективный способ ее решения состоит в том, чтобы перенести значительную часть обработки спутниковой информации со стороны пользователя на сторону спутникового центра, одновременно предоставив пользователю гибкую систему запросов, с помощью которой он мог бы формулировать и выбирать доступный для себя уровень обработки данных. Развитие распределенной обработки спутниковой информации, объединение информационных ресурсов в единую информационную среду, оперативное управление информацией, хранящейся в разнородных и распределенных архивах, является актуальной задачей, особенно для спутниковой информации.

2. Распределенная информационная система первого поколения

В данный момент успешно внедрена и выполняет свои функции распределенная информационная система первого поколения, построенная на основе программных средств и технологий системы INFEO.

Международная Информационная система INFEO (Information about Earth Observation) – инициатива Европейского Космического Агентства (ESA), призванная решить проблему доступа к распределенным данным дистанционного зондирования Земли (ДЗЗ). основополагающим принципом системы является обеспечение интероперабельности – возможности взаимодействия гетерогенных сервисов и каталогов данных в рамках единой информационной системы. Основная функция такой системы – создание единой информационной среды, предоставление единой точки доступа для поиска и получения информации по всем архивам и каталогам, зарегистрированным в системе, независимо от их географического расположения и внутреннего формата данных.

В основе распределенной информационной системы первого поколения лежит сеть узлов, так

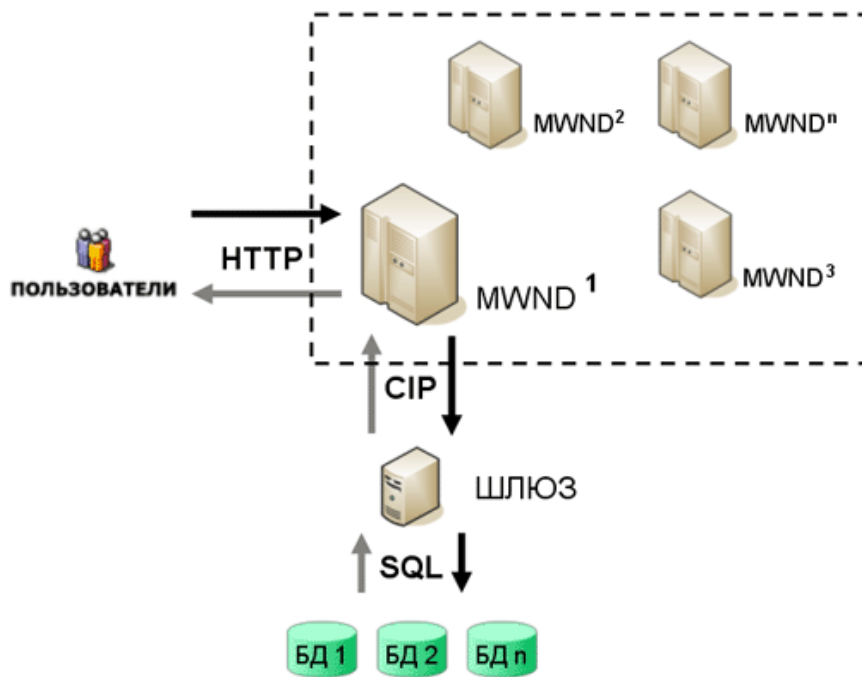


Рис 1. Схема работы распределенной информационной системы

называемых Middleware Nodes (MWNDs) (рис. 1).

Каждый из узлов содержит метаданные коллекций, описывающие данные, доступ к которым предоставляется через Шлюзы (Gateways). Пользователь для поиска информации обращается к системе через WWW-сервер на одном из узлов MWND, где формируется распределенный запрос, рассылаемый релевантным шлюзам. Результаты выполнения запроса объединяются и возвращаются пользователю.

Шлюз представляет собой промежуточное звено между узлом MWND и каталогом или архивом поставщика данных. Шлюз обменивается сообщениями по протоколу CIP (Catalogue Interoperability Protocol) с узлом MWND, получая от него поисковые запросы и возвращая ему результаты поиска. Запросы передаются транслятору RDBMS, который преобразует их в SQL-запросы для непосредственного поиска в базе данных, либо формирует запрос для обращения к предоставляемому сервису (Web Map Server и т.п.).

В основе распределенной информационной системы лежит протокол CIP [4], регламентирующий правила взаимодействия пользователей и каталогов данных ДЗЗ. Для поддержки одновременного доступа пользователя к множеству каталогов используется трехуровневый принцип распределения запросов. Пользователь через web-интерфейс задает поисковый запрос и посылает его к узлу MWND, который, в свою очередь, перенаправляет запрос множеству серверов каталогов данных. Сервера, имеющие данные, удовлетворяющие критериям поиска, возвращают ответы на узел MWND, через который пользователь получает объединенный результат

поиска.

Существует несколько стандартов, пригодных для поиска и извлечения информации в распределенных информационных системах (HTTP, Gopher и т.п.). Протокол Z39.50 [5,6,7] имеет важное преимущество, ставшее решающим при выборе базового протокола для CIP. Преимущество заключается в том, что форма представления результатов поиска, так же, как и способ ввода поисковых запросов, не зависит от источника информации. Это определяется тем, что Z39.50 использует качественно иную модель взаимодействия пользователя и поисковой системы, позволяющую абстрагироваться от способа организации конкретной базы данных, сделать доступ к ее содержанию независимым от формы хранения последнего. Поисковые запросы всегда адресованы не к реальной базе данных, а к абстрактной. Эта абстрактная БД не имеет структуры и характеризуется только поисковыми атрибутами. При таком подходе к процедуре поиска все базы данных становятся для клиента одинаковыми, если поддерживают один и тот же набор поисковых атрибутов.

Различные вычислительные системы используют различные средства и методы описания, хранения и доступа к данным, протокол Z39.50 избавляет пользователя от необходимости адаптации к каждому из источников информации. Z39.50 представляет собой общую абстрактную структуру для поиска и получения информации, требующую дополнительной настройки для каждой конкретной области применения. Возможность осуществления таких настроек, так называемых профилей, является одним из основных преимуществ протокола. С

учетом вышесказанного, протокол SIP, по сути, является профилем Z39.50, ориентированным на работу с данными ДЗЗ.

Протокол SIP поддерживает концепцию параллельных операций. Это позволяет пользователю инициировать несколько поисковых запросов одновременно. То есть, клиенту нет необходимости дожидаться окончания выполнения предыдущей операции (а именно, получения ответа, соответствующего запросу), чтобы сделать следующий запрос.

В протоколе SIP определены разные принципы представления метаданных для пользователей и для поставщиков данных. Для пользователей базовой единицей структуры данных является понятие архива, для поставщиков – понятие коллекции. Архив может содержать различные типы данных: спутниковые изображения, результаты обработки изображений, температурные карты, статистические данные, алгоритмы их обработки, документацию и т.д. Коллекция – группа данных, объединенных на основе каких-либо общих атрибутов. Эти атрибуты являются описанием коллекции, которое может быть использовано для формирования дополнительных критериев поиска. Коллекция может содержать как описания (дескрипторы) конечных данных, так и описания других коллекций, образуя сложную иерархию коллекций.

Все типы коллекций поддерживают поисковые методы, определенные стандартом SIP, а именно:

- Поиск коллекций. Этот метод используется для поиска коллекций, удовлетворяющих заданным условиям (время, координаты, тип датчика и т.п.) Результат поиска может быть использован для дальнейшего поиска с измененными условиями.
- Поиск данных. Этот метод используется для поиска дескрипторов, указывающих на конечный продукт – данные ДЗЗ.

Пользователь может дополнительно задавать область поиска: выполнять поиск только в одном каталоге (локальный поиск) или задать распространение поискового запроса на другие каталоги и архивы (распределенный поиск).

3. Реализация распределенной информационной системы

Рассмотрим пример реализации системы, использующей принципы распределенного хранения и поиска информации. Распределенная информационная система спутниковых архивов была реализована в сотрудничестве ИКИ РАН и ИАПУ ДВО РАН. В качестве поставщика данных дистанционного зондирования Земли выступает Лаборатория спутникового мониторинга ИАПУ Дальневосточного Отделения РАН (<http://satellite.dvo.ru>). Шлюз установлен на сервере, физически размещенном в Москве, в Институте Космических Исследований РАН

(<http://iris.iki.rssi.ru>). Узел Middleware Node находится в Институте Европейского Космического Агентства ESA/ESRIN (<http://www.esa.int>). В настоящее время в лаборатории спутникового мониторинга ИАПУ ежедневно принимается, обрабатывается, поставляется и архивируется информация со спутников NOAA (более 1500 сеансов в год) и информация с геостационарных спутников: японского GMS-5 и китайского FY-2B (с середины 2003г. круглосуточно). Базу для основных выходных продуктов составляют следующие категории спутниковой информации.

- Каталоги – наборы описаний всей принятой информации с указанием даты и времени приема, спутника, спектральных каналов и координат обзора.
- Цифровые мгновенные калиброванные изображения в меркаторской проекции с пространственным разрешением в 1.1 км и разрешением по температуре для ИК каналов – одна восьмая градуса.
- Карты температуры поверхности океана (ТПО) в изотермах, осредненные по времени (5-10 дней) и пространству (10-100км).
- Карты скоростей поверхностных течений и ветра, строящиеся на основе метода морских маркеров (МММ).
- Карты доминантных ориентаций термических контрастов (ДОТК), которые являются альтернативой векторам скоростей течений, строятся автоматически на основе метода ориентированных текстур.

На основе принятой информации в соответствии со стандартом для описания географических метаданных ISO19115 [8] в автоматическом режиме генерируются метаданные коллекций. Раз в сутки, через ftp-соединение метаданные отправляются на сервер со шлюзом, где они размещаются в базе данных шлюза. С этого момента очередная коллекция данных становится доступна любому пользователю, зарегистрированному в системе. В данной схеме работы в качестве поставщиков данных может выступать не одна, а несколько независимых организаций, причем данные могут быть самыми разнообразными – начиная от температурных карт поверхности океана и карт течений и заканчивая спутниковыми снимками лесных пожаров и зон неблагоприятной экологической обстановки мегаполисов. Единственное условие, которое необходимо выполнить поставщику данных – подготовить метаданные, описывающие коллекцию, в соответствии со стандартом SIP. В свою очередь, оператору Шлюза необходимо выполнить настройку системы для каждого поставщика: задать отображение (mapping) получаемых данных на предопределенную стандартом схему, описывающую коллекцию.



Рис 2. Поисковый интерфейс пользователя

Для поиска необходимой информации пользователь может воспользоваться поисковым интерфейсом по адресу <http://catalogues.eoportal.org/eoli.html>. Интерфейс выполнен в виде Java-апплета (рис. 2).

Существует возможность ограничивать поиск либо по поставщикам данных (панель Find Products), либо по коллекциям данных (Find Collections). В раздел Find Products входят крупные поставщики, имеющие несколько собственных обширных коллекций. Например, ESA (коллекции Radar Imagery, Radar Altimetry, Optical/Multispectral Imagery), DLR (коллекции Atmospheric Sensors, Thematic Maps, Optical Sensors, Digital Elevation Model, Synthetic Aperture Radar Data) и др. После выбора типа данных, по которым будет ограничен поиск, следует выбрать площадь земной поверхности. Для более точного поиска информации можно воспользоваться режимом расширенного поиска Advanced Query Mode. Этот режим позволяет дополнительно задавать ключевые слова для поиска, а также выбирать спутник и сенсор, с помощью которого были получены данные. Ключевые слова служат для отбора данных, поле описания которых содержит строку, заданную в поле Free Text.

Опыт внедрения системы и ее эксплуатации в

течение значительного периода позволяет говорить и о выявленных недостатках. В первую очередь они относятся к базовому протоколу системы – Z39.50. Z39.50 – устаревающий стандарт, в последнее время практически не развивающийся, что налагает существенные ограничения на интеграцию дополнительных сервисов, помимо каталогов и архивов.

Необходимо упомянуть определенные трудности, возникающие при переносе Gateway на другие *nix системы, кроме поддерживаемых SunOS и HP-UX, а также выявленные сложности настройки СУБД и программного обеспечения INFEO/Gateway на стороне поставщика данных.

4. Распределенная информационная среда второго поколения EOLI-XML

ESA (Европейское космическое агентство) начинает разработку технологий интеграции второго поколения. Это вызвано стремлением расширить участие ESA в глобальном экологическом мониторинге GMES, улучшить в задачах охраны окружающей среды взаимодействие с мировыми информационными системами ДЗЗ [9] и необходимостью поддержать требования

Директивы ЕС – INSPIRE [10] о совместимости с другими информационными системами. Исследования технологий интеграции второго поколения направлены на миграцию от подхода, основанного на технологии Z39.50/Gateways, к использованию веб-сервисов (XML над SOAP/HTTP) для связи внешних каталогов с каталогами INFEO. Технологии интеграции второго поколения, получившие название EOLI-XML, разрабатываются ESA в кооперации с национальными космическими Агентствами: Франции Centre National d'Etudes Spatiales (CNRS), Италии Italian Space Agency (ASI), Германии German Space Agency (DLR), European Union Satellite Centre (EUCS), Joint Research Centre (JRC) при участии ряда партнеров в промышленности SPOT Image, Siemens, Spacebel и др. EOLI-XML представляет собой внешний интерфейс для обмена сообщениями между сервером и клиентом на основе протокола SOAP [11].

4.1 Система SSE

Технология EOLI-XML лежит в основе системы SSE (Service Support Environment) [12], разрабатываемой для решения указанных задач с помощью создания открытой, сервис-ориентированной, распределенной среды. Система SSE является инфраструктурой, единой средой для потребителей и провайдеров данных и сервисов. SSE объединяет своих пользователей на основе стандарта XML с использованием протоколов обмена сообщениями SOAP и языка описания веб-сервисов WSDL (Web Service Definition Language) [13]. Таким образом, любой сервис, интегрируемый в среду SSE, должен предоставить SOAP-интерфейс, в соответствии с SSE ICD (Interface Control Document) [14].

Итак, основные цели создания инфраструктуры SSE:

- Предоставить среду, максимально упрощающую взаимодействие типа «сервис-провайдер – пользователь» и «сервис-провайдер – сервис-провайдер».
- Упростить интеграцию уже существующих сервисов, предоставляя универсальный интерфейс на основе XML, позволяющий не изменять их структуры.
- Предоставить пользователям единую точку входа – Интернет-портал системы.

Система SSE состоит из двух основных компонент: SSE Portal Server и AOI Server, образующих вместе Интернет-портал, с которым взаимодействует конечный пользователь.

SSE Portal Server предоставляет web-интерфейс для доступа пользователей к portalу. Сервер построен с использованием промышленной технологии J2EE, которая удовлетворяет

требованиям SSE: хорошая масштабируемость, интеграция с существующими информационными системами, гибкая политика безопасности, поддержка стандартных протоколов и языков, переносимость компонент без необходимости перекомпиляции. Выполнение этих требований является обязательным для успешного построения информационной инфраструктуры. В качестве сервера приложений выбран сервер JBoss с открытым исходным кодом.

AOI Server работает в связке с SSE Portal Server и служит для предоставления для сервис-провайдеров функций визуального выделения области на карте (area of interest, AOI) при задании критериев поиска, а также визуализации результатов поиска.

Так как большинство сервисов в качестве одного из входных параметров требуют указание географической области, SSE Portal предоставляет специальную поддержку этой возможности. При обращении к сервису пользователь должен указать географическую область и сделать это он может разными способами: выбрать из списка; указать область на карте; загрузить файл, описывающий AOI; задать AOI с помощью указания координат.

При обращении пользователя к сервису, SSE Portal использует апплет, подгружаемый с сервера AOI для визуального отображения выделенной области на карте. SSE позволяет использовать для визуализации как локально хранимую информацию, так и подключаться к OGC WMS серверам для загрузки дополнительных слоев карты [15]. Выделенная таким образом на карте область затем преобразуется в описание в формате GML (Geography Markup Language) [16] описание вместе с запросом отправляется сервис-провайдеру в виде SOAP сообщения.

4.2 Интеграция сервиса

Компонент системы, устанавливаемый на стороне сервис-провайдера, имеет условное название Toolbox и предназначен для упрощения процедуры конвертации уже существующего сервиса в сервис, позволяющий принимать запросы и отвечать на них в формате SOAP сообщений. Toolbox состоит из двух приложений: среды разработки и среды выполнения. Среда разработки представляет собой плагин для Eclipse [17], предоставляющий инструментарий для создания XML скриптов. Среда выполнения (Toolbox Runtime Environment) это Веб-приложение, позволяющее размещать SOAP интерфейсы, совместимые с SSE, посредством XML скриптов, описывающих операции, которые должны быть выполнены при вызове сервиса. Toolbox может работать на любой платформе, поддерживающей технологию Java Virtual Machine. Toolbox это открытый программный продукт, свободно распространяемый по лицензии GNU General Public License [18].

Для выполнения процедуры интеграции сервиса сервис-провайдеру необходимо выполнить лишь

несколько базовых операций:

- Задать XML-схему сервиса
- Выбрать один из механизмов интеграции, предоставляемых SSE Toolbox.
- Интегрировать сервис с помощью SSE Toolbox.
- Зарегистрировать вновь созданный сервис в системе SSE.

В соответствии с SSE ICD, среда SSE поддерживает две модели обмена информацией. На этапе регистрации сервиса, сервис-провайдер должен решить, какая из моделей будет использоваться для каждой из базовых операций, т.е. поиск, запрос детальной информации, заказ информации. Инфраструктура SSE не накладывает ограничений на время, необходимое для обработки результатов запроса. Для получения результатов некоторых запросов могут потребоваться часы, тогда как другие обрабатываются практически мгновенно. Портал SSE позволяет хранить результаты запросов до тех пор, пока пользователь не получит их. На стороне сервис-провайдера результаты запроса обычно сохраняются в течение нескольких дней для получения пользователем.

- *Асинхронная модель*: используется, когда выполнение операций занимает достаточно большой интервал времени, что не дает возможности пользователю незамедлительно увидеть результат запроса. Результат сохраняется в базе данных SSE; у пользователя есть возможность посмотреть их в удобное ему время, когда он снова зайдет на портал. SSE портал может исполнять роль как SOAP-клиента, так и SOAP-сервера.
- *Синхронная модель*: используется, когда сервис возвращает результат запроса незамедлительно. Таким образом, пользователь может увидеть результат на той же странице, где вводил критерии поиска. SSE-портал выступает в роли SOAP клиента и Toolbox RE работает только как SOAP сервер, отвечая на SOAP запросы. В рамках одной HTTP-сессии клиенту (SSE-портал) отправляется XML-ответ.

С точки зрения Toolbox, сервис это набор определенных операций, выполняемых при получении соответствующего запроса; результат выполнения операций в случае успешного завершения отправляется пользователю. Каждая операция связана с определенным набором XML-скриптов в зависимости от принятой модели работы. Количество одновременно поддерживаемых сервисов ограничено только производительностью сервера, на котором выполняется Toolbox RE.

В зависимости от сервиса, результатом его выполнения могут быть файлы растровых изображений большого объема, небольшие

векторные файлы или текстовая информация. SSE не налагает ограничений на тип результирующей информации, предоставляя сервис-провайдеру свободу выбора.

Среда SSE поддерживает следующие механизмы передачи данных от провайдера пользователю:

- Передача данных в теле SOAP сообщения: в текстовом виде или векторной графике (SVG)
- Передача файлов через FTP: пользователь получает ссылку через SOAP сообщение
- Передача файлов через HTTP: пользователь получает ссылку через SOAP сообщение
- Получение векторных изображений с помощью сервиса OGC WFS [16]
- Получение растровых изображений с помощью сервиса OGC WCS [16]

4.3 Архитектура Toolbox

Итак, Toolbox – программный комплекс на стороне провайдера, функциональный аналог шлюза в инфраструктуре первого поколения, связующее звено между провайдером и SSE порталом. Основная его задача – преобразование входной и выходной информации уже существующего сервиса к формату сообщений системы (SOAP сообщения). На рис. 3 показана структурная схема Toolbox.

Toolbox поддерживает широкий спектр механизмов обмена информацией с внутренним интерфейсом сервисов: FTP, Java API, JDBC, SOAP, файловый обмен и т.п. Это позволяет интегрировать в SSE большинство существующих сервисов без дополнительной настройки внутреннего интерфейса.

Типичная схема работы выглядит следующим образом. Пользователь через интерфейс SSE портала формирует запрос к нужному сервису, выделив на карте область АОИ и указав характеристики искомой информации (см. п. 4.1). После этого сформированный в формате SOAP-сообщения запрос пересылается на URL сервиса, указанный провайдером при регистрации. Здесь сообщение поступает на вход сервера веб-приложений, где в зависимости от поступившего запроса запускается соответствующий сервлет, который занимается дальнейшей обработкой запроса и передачей его на внутренний интерфейс сервиса. Для выполнения сервлетов сервис-провайдеру вместе с Toolbox следует установить web-сервер с поддержкой сервлетов (рекомендуется к установке веб-сервер Tomcat).

Обработка входного запроса и формирование ответа на запрос выполняется на основе WSDL документа в соответствии с XML схемой сервиса. При установке Toolbox уже имеется предопределенная схема, тем не менее, некоторые сервисы могут иметь специфические параметры

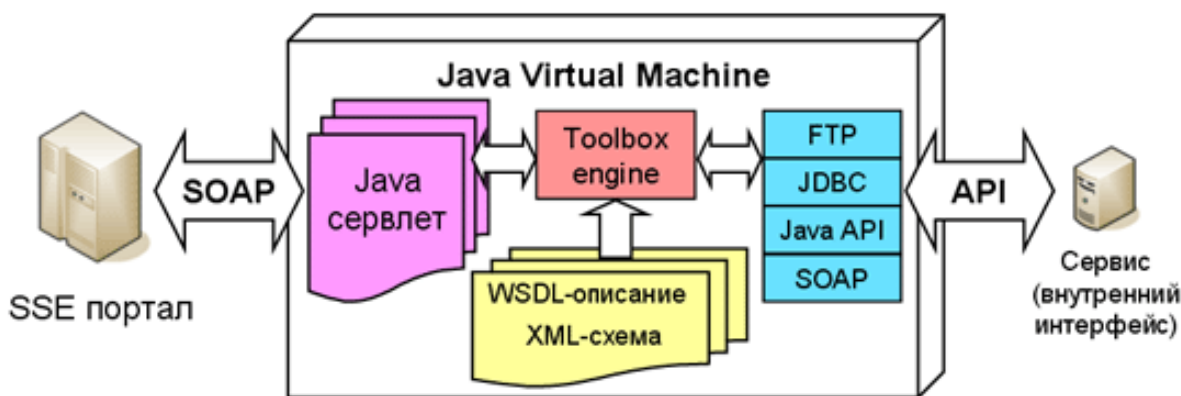


Рис. 3. Схема работы Toolbox

входных и выходных сообщений, что можно легко исправить, отредактировав соответствующие файлы XML схем.

Литература

- 1] Kudashev E.B. Digital Library: Improving the accessibility to Russian satellite data for mitigating natural disaster. // Proceedings of ESA Workshop PV-2004. Ensuring the Long-Term Preservation and Adding Value to the Scientific and Technical Data – Session 4 Future prospects. Paper 4.01. p. 233-240 – ESA / ESRIN, Frascati, Italy. 2004.
- 2] Кудашев Е.Б., Балашов А.Д., Филонов А.Н. От Спутникового Экологического Мониторинга к Электронным Библиотекам спутниковой информации в образовании и исследовании Земли из космоса. *Вычислительные технологии*. 2004. Т. 9. Специальный выпуск. С. 50-62
- 3] Кудашев Е.Б., Филонов А.Н. Организация информационной распределенной среды и интеграция спутниковых архивов. *Труды 7ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2005*, Ярославль, Россия, 2005.
- 4] Catalog Interoperability Protocol (CIP) Specification – Release B-24, 2005. http://wgiss.ceos.org/ics/documents/cip2.4/cipspec-2_4_75_6.pdf
- 5] Registry of Z39.50 Object Identifiers, 2002. <http://www.loc.gov/z3950/agency/defns/oids.html>
- 6] Жижимов О.Л., Мазов Н.А. Принципы построения распределенных информационных систем на основе протокола Z39.50. *ОИГТМ СО РАН*, Новосибирск, ИВТ СО РАН, 2004, стр 361
- 7] Капустин В.А., Смирнов В.М. Протокол Z39.50: обзор. *Электронные библиотеки: перспективные методы и технологии, электронные коллекции - Труды Первой Всероссийской научной конференции*. С.-Петербург, 1999 г., СПб, стр. 124-127, 1999.
- 8] ISO/TC 211 Geographic information/Geomatics home page, 2003. <http://www.isotc211.org/>
- 9] National Spatial Data Infrastructure (NSDI) home page, 2007. <http://www.fgdc.gov/nsdi>
- 10] INSPIRE home page, 2007. <http://www.ec-gis.org/inspire>
- 11] Earthnet On-line XML Front-End Interface Control Document, EOLI-XML-006-ICD, Issue 1.7, 2003. <http://earth.esa.int/rtd/Documents>
- 12] Service Support Environment Architecture, Model and Standards, 2007. http://earth.esa.int/services/esa_doc/doc_sse.html
- 13] Web Services Description Language (WSDL) 1.1, W3C Note, 2001. <http://www.w3.org/TR/wsdl>
- 14] Interface Control Document (ICD) for the Service Support Environment (SSE), 2007. <http://services.eoportal.org/massRef/documentation/icd.pdf>
- 15] The Open Geospatial Consortium, 2007. <http://www.opengeospatial.org/>
- 16] ISO/TC 211/WG 4/PT 19136 Geographic information – Geography Markup Language (GML), 2004. http://portal.opengeospatial.org/files/?artifact_id=4700
- 17] Eclipse home page, 2007. <http://www.eclipse.org/>
- 18] GNU GENERAL PUBLIC LICENSE, Version 2, 1991. <http://www.gnu.org/licenses/gpl.txt>

Technologies and standards for services, catalogues and databases integration in Earth Observation programs

Kudashev E.B., Filonov A.N.

The paper surveys emerging standards designed for building distributed service-oriented environments. The project of integration Russian satellite data archive in international system SSE supported by ESA is observed as an example.

* Работа проводится при поддержке РФФИ (проект № 05-07-90021)