

# **Предметные посредники в гибридной грид-инфраструктуре виртуальных обсерваторий для решения задач над неоднородными распределенными информационными ресурсами**

А.Е. Вовченко, В.Н. Захаров, Л.А. Калиниченко, С.А. Ступников,  
Н.А.Скворцов

## **Введение**

В различных областях науки наблюдается экспоненциальный рост объема получаемых экспериментальных данных. Число организаций, получающих данные наблюдений в отдельных областях науки, велико. Разнообразие (информационная несогласованность) получаемой информации вызывается, в частности, не только большим числом организаций, производящих наблюдения, и их независимостью, но и разнообразием объектов наблюдения, непрерывным и быстрым совершенствованием техники наблюдений, вызывающим адекватные изменения структуры и содержания накапливаемой информации. Разрыв между исследователями и информационными ресурсами приводит к необходимости поиска новых путей создания информационных систем, в которых особое внимание было бы сосредоточено на специальных средствах организации решения задач над множеством неоднородных распределенных информационных ресурсов (данных и сервисов), накапливаемых в научных центрах. Разрабатывается ряд инфраструктур, которые технически способствуют организации решения задач в такой среде: веб-сервисы, гриды данных, интероперабельные инфраструктуры промежуточного слоя и др. Вместе с тем проблемы семантики, возникающие при интеграции ресурсов, сложны и все еще далеки от решения.

Настоящая статья в качестве применений ограничивается рассмотрением инфраструктур информационных систем для науки, которые в последнее время приобретают вид международных виртуальных обсерваторий (ВО). Инфраструктура Российской виртуальной астрономической обсерватории (РВО) [1] основывается на стандартах Альянса международной виртуальной обсерватории IVOA [2], дополненных концепцией предметных посредников, позволяющих задавать определение прикладных областей для формулирования и решения классов научных задач в терминах понятий этих областей, структур информационных объектов, декларативно объявляемых

сервисов и процессов. Посредники располагаются между исследователями (формулирующими задачи в терминах посредников), и разнообразными распределенными информационными ресурсами (данными, сервисами, процессами), необходимыми для решения задач.

Предметные посредники играют ключевую роль для решения семантических проблем интеграции ресурсов. В частности, при интеграции неоднородных ресурсов в посреднике нужно уметь семантически отождествлять объекты, представленные в различных информационных моделях, и семантически правильно отображать схемы интегрируемых ресурсов в схему посредника. Поскольку в общем случае ресурсы неоднородны (представлены в различных моделях), при интеграции неоднородных ресурсов для однородного представления их семантики требуется приведение различных информационных моделей к унифицированному виду в рамках некоторой унифицирующей информационной модели, которая называется *канонической*. Методы отображения информационных моделей и синтеза расширяемых канонических информационных моделей для слоя предметных посредников рассмотрены в [3].

Другая совокупность семантических проблем создания посредников для ВО связана с определением предметного посредника и регистрации в нем релевантных посреднику информационных ресурсов. Определение предметного посредника для класса задач включает определение понятий предметной области, выражаемых соответствующими онтологическими спецификациями, спецификации классов объектов предметной области, спецификации типов экземпляров названных классов и их методов, определяющих их поведение, спецификации процессов решения задач данного класса как совмещенных во времени последовательностей действий, реализуемых методами классов, сервисами и другими процессами. Предполагается, что такие спецификации преобразуются в спецификации канонической модели, имеющей формальную семантику. Результат определения посредника, выполняемого заинтересованным научным сообществом, составляет спецификацию посредника, образуемую в результате достижения консенсуса в таком сообществе, а сама деятельность по спецификации посредника называется периодом его *консолидации*.

Регистрация релевантных посреднику ресурсов рассматривается как задача композиционного проектирования информационных систем (ИС). Регистрация ресурсов есть процесс трансформации спецификаций, включающий декомпозицию спецификаций посредника на непротиворечивые фрагменты, поиск среди спецификаций ресурсов подходящих типов данных - кандидатов для уточнения ими

спецификаций типов посредника, построение выражений, определяющих классы ресурсов в виде композиции классов посредника.

Существенным для данной работы является то обстоятельство, что в Великобритании развивается система АстроГрид [4] как полная инфраструктура для создания систем решения научных задач в астрономических ВО. Функционально сервисы АстроГрид в основном соответствуют требуемым для реализации инфраструктуры РВО. Вместе с тем, АстроГрид не предусматривает развитых средств решения задач над множествами неоднородных информационных ресурсов, подобных предметным посредникам. При поддержке РФФИ был инициирован проект создания гибридной архитектуры, объединяющей возможности АстроГрид с архитектурой предметных посредников. Целью настоящей статьи является обзор достигнутых в этом проекте результатов развития такой архитектуры. Ряд вопросов, затронутых в статье, более подробно рассматривается в работе [5].

Дальнейшее изложение материала в статье структурировано следующим образом. В разделе 2 рассматривается подход к решению задач над множеством неоднородных распределенных информационных ресурсов. В разделе 3 рассматриваются особенности инфраструктуры системы АстроГрид. В разделе 4 приведено краткое описание объединенной архитектуры АстроГрид и средств поддержки предметных посредников. Наконец, заключение статьи подводит итог обсуждению и намечает планы дальнейшего развития работы.

### **Решение задач над множеством неоднородных распределенных информационных ресурсов**

*Движимый приложениями* подход к представлению описания предметной области класса задач, развиваемый в гибридной архитектуре, отличается тем, что описание предметной области образуется независимо от ресурсов, а затем релевантные приложению ресурсы отображаются в это описание. Подход предполагает создание предметного посредника, который поддерживает взаимодействие между приложением и ресурсом на основе определения предметной области (определения посредника).

Для интеграции неоднородных грид-базируемых информационных ресурсов разработана инфраструктура предметных посредников, а также специализированный промежуточный слой (middleware), включающий:

- расширяемую каноническую информационную модель для унифицированного определения посредников и неоднородных информационных ресурсов;
- Унификатор неоднородных информационных моделей (для построения канонических информационных моделей);

- средства идентификации информационных ресурсов, релевантных посреднику и их семантической интеграции в посреднике;
- средства определения (спецификации) прикладных задач;
- средства интерпретации задач в контексте посредника над зарегистрированными ресурсами.

Ниже рассматриваются основные компоненты инфраструктуры предметных посредников.

Основной принцип синтеза канонической информационной модели для состоит в расширяемости ее ядра в разнородной среде, включающей различные информационные модели, используемые для представления ресурсов. Ядро канонической модели фиксируется. Для каждой конкретной информационной модели  $R$  определяется расширение ядра канонической модели так, чтобы оно вместе с ядром *уточнялось* бы моделью  $R$ . Такая уточняющая трансформация моделей должна быть доказуемо правильной. Для формализации понятия уточнения используется нотация абстрактных машин и В-технология [6]. Каноническая модель среды синтезируется как объединение расширений, образованных для моделей ресурсов.

Разработан прототип Унификатора неоднородных информационных моделей, предназначенный для частичной автоматизации методов синтеза канонических моделей [3]. Одной из практических задач Унификатора является поддержка интеграции неоднородных ресурсов в посредниках. Именно поэтому информационная модель предметных посредников (язык СИНТЕЗ [7] – гибридная слабоструктурированная и объектно-ориентированная информационная модель) была выбрана в качестве ядра канонической информационной модели.

Целью Унификатора моделей является унификация множества информационных моделей (называемых *исходными*), совместно использующихся в некоторой ИС. Унификация исходной модели  $R$  есть приведение ее к канонической информационной модели  $C$ , т.е. создание такого расширения  $E$  канонической модели и такого отображения  $M$  исходной модели в расширенную каноническую, что исходная модель *уточняет* расширенную каноническую модель. Уточнение моделей означает, что для любой допустимой спецификации  $r$  модели  $R$  ее образ  $M(r)$  при отображении  $M$  уточняется спецификацией  $r$ . В результате унификации также должна быть получена возможность доказывать уточнение произвольной конкретной спецификацией  $r$  модели  $R$  ее образа  $M(r)$ . Верификация уточнения моделей осуществляется на наборе образцов спецификаций исходной модели.

Процесс регистрации неоднородных информационных ресурсов в предметном посреднике в подходе, движимом приложениями, основан на технике GLAV [8], комбинирующей два подхода – LAV (Local As

View) и GAV (Global As View). Согласно LAV, схемы регистрируемых ресурсов рассматриваются как материализованные взгляды над виртуальными классами посредника. В этом случае GAV взгляды служат для разрешения различных конфликтов между спецификациями ресурсов и посредника и обеспечивают правила трансформации результатов запроса от ресурса в представление в посреднике. Подобная техника регистрации обеспечивает стабильность спецификации приложения ВО при изменении конкретных информационных ресурсов и их фактического присутствия (удаление ресурса, добавление новых ресурсов, и пр.), а также масштабируемость посредников по отношению к числу ресурсов, регистрируемых в них. Принципиальным моментом в этой схеме является реализация доказательства уточнения фрагментов спецификаций посредника спецификациями ресурсов в процессе построения отображений таких спецификаций. Данные методы составляют основу прототипа инструментария идентификации и регистрации информационных ресурсов в посреднике.

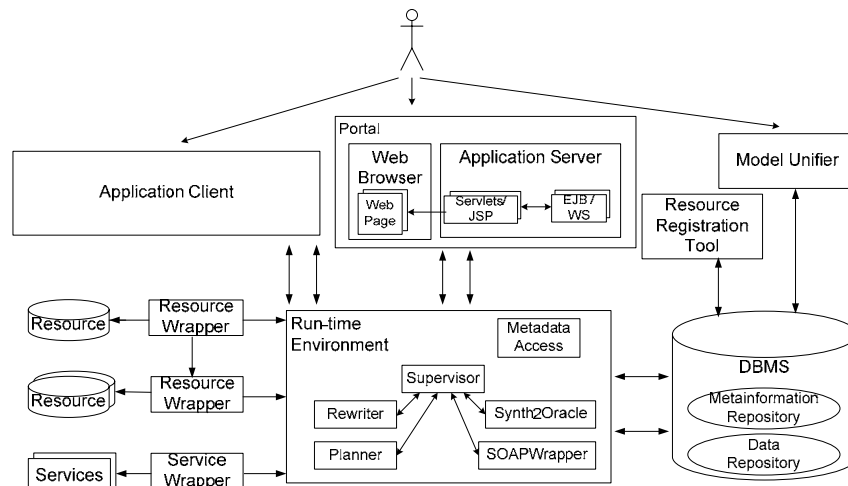


Рисунок 1. ПО инфраструктуры посредников

ПО поддержки посредников (Рис. 1) включает также средства описания задач и их интерпретации в контексте посредника над зарегистрированными ресурсами. Ввиду ограниченного объема статьи, детальные описания инструментария регистрации и Унификатора не приводятся.

*Application Client* – это графический интерфейс пользователя к посреднику, поддерживающий язык программ (запросов) посредника и

обеспечивающий удобное представление результатов запросов. *Portal* является Интернет-аналогом компонента *Application Client*.

*Run-time environment* (среда исполнения посредников) предназначена для обработки программ (запросов) пользователя, переписывания программ в запросы над зарегистрированными ресурсами, передачи переписанных запросов адаптерам, получения результатов от адаптеров и передачи их пользователю.

*Адаптеры* (*wrappers*) подключают посредник к конкретным ресурсам. Адаптеры преобразуют запросы на языке посредника в запросы на языке ресурсов а также результаты запросов, представленных в формате ресурсов, в формат посредника.

*Metainformation and Data repository* используется в качестве хранилища метайнформации посредника (используемой для преобразования запросов), а также для хранения промежуточных результатов, возвращаемых ресурсами.

### **АстроГрид как основа инфраструктуры РВО**

Система АстроГрид имеет целью поддержку инфраструктуры для решения научных задач в астрономических виртуальных обсерваториях (ВО). Для решения научных задач АстроГрид предоставляет средства доступа к астрономическим каталогам данных или архивам изображений, а также к реестрам метаданных (в которых регистрируются ресурсы). Система АстроГрид построена на основе архитектуры веб-сервисов. АстроГрид включает следующие основные компоненты:

- Компонент Registry (реестр) представляет собой коллекцию метаданных - XML-документов, описывающих ресурсы, которые могут использоваться при решении задач с помощью ВО. Registry реализован на основе стандарта OAI PМН, специализированного IVOA для нужд ВО. В качестве ресурсов могут выступать приложения, сервисы, информационные ресурсы, компоненты АстроГрида. Registry поддерживает функцию сбора информации о ресурсах, зарегистрированных в других реестрах Международной Виртуальной Обсерватории (*harvesting*).
- Компонент VOSpace представляет собой виртуальное хранилище данных, к которым могут иметь доступ все сервисы системы АстроГрид. Компонент MySpace используется для хранения и передачи файлов между сервисами в рамках одной задачи, и между задачами, решаемыми при помощи ВО.
- Компонент Common Execution Architecture (CEA - Общая Исполнительная Архитектура) определяет способ, каким приложения могут быть оформлены как сервисы. CEA-приложение может быть вызвано из потоков работ и как отдельное приложение,

может быть зарегистрировано и найдено в реестрах АстроГрид; может считывать и записывать данные в VOspace, и т.д.

- Компонент DataSet Access (DSA), предоставляет возможность подключать базы данных к системе АстроГрид. DSA предоставляет два интерфейса доступа: CEA-приложение, которое выполняет запросы к базе данных; и Cone Search, который позволяет производить поиск в некоторой заданной области неба.

Анализ показывает, что использование АстроГрида в качестве основы инфраструктуры РВО позволяет реализовать перечисленные принципы (поддержка грид-интероперабельных сервисов, модульность архитектуры, возможность повторного использования и композиции сервисов, создание многослойной архитектуры). Компоненты АстроГрида являются непосредственно применимыми в качестве ядра архитектуры РВО.

### Гибридная грид-инфраструктура для решения задач

При разработке гибридной архитектуры (Рис. 2) во главу угла ставилась задача обеспечения интероперабельности компонентов слоя предметных посредников и АстроГрида.

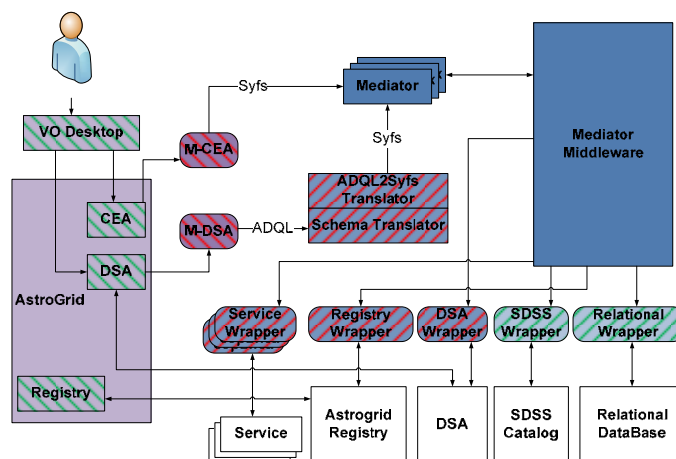


Рисунок 2. Гибридная архитектура АстроГрид и посредников

В качестве необходимых, были выделены следующие возможности гибридной архитектуры:

- регистрация и поиск посредников в реестрах АстроГрида;
- просмотр пользователем схем посредников и другой информации о посредниках;

- задание запросов к посредникам как на языке посредников Syfs [7], так и на языке ADQL [9];
- использование программ посредников как шагов потоков работ в системе АстроГрид;
- получение результатов программ посредников в среде АстроГрид.

Посредник может регистрироваться в реестрах АстроГрид в двух вариантах: посредник как СЕА-приложение и/или как DSA-компонент. Посредник, зарегистрированный как СЕА-приложение, выполняет запрос на языке посредников (Syfs) над объектной схемой посредника.

Посредник, зарегистрированный как DSA, выполняет запрос на языке ADQL [9] над упрощенной (реляционной) схемой посредника. В обоих случаях результат возвращается в формате, совместимом с АстроГрид. Для реализации этой возможности были разработаны компоненты М-СЕА и М-DSA, которые инкапсулируют конкретный экземпляр посредника. Компоненты *Schema Translator* и *ADQL2Syfs Translator* разработаны для обеспечения возможности написания запросов на реляционном языке ADQL над объектно-ориентированной схемой посредника. Компонент *Schema Translator* используется для преобразования схемы посредника в реляционный вид, а *ADQL2Syfs Translator* – для преобразования запросов на языке ADQL (в терминах реляционной формы схемы посредника) в запросы на языке Syfs (в терминах исходной схемы посредника). Оформление посредника как DSA-компонента важно, поскольку предоставляет пользователям АстроГрид видеть посредник как обычный DSA-источник.

Адаптеры обеспечивают преобразование запросов на языке посредника в запросы на языке ресурса, получение результата запроса от другого ресурса, преобразование объектов результирующего множества в объекты схемы посредника. В частности, разработаны и включены в прототип гибридной грид-инфраструктуры адаптер DSA-ресурсов (*DSA Wrapper*) для подключения их к посреднику и адаптер реестров АстроГрида (*Registry Wrapper*) для реализации поиска ресурсов по метаданным. Разработаны также адаптер реляционных баз данных, обеспечивающий подключение к посреднику различных видов реляционных СУБД, и адаптер к астрономическому каталогу данных SDSS с поддержкой возможности выполнения кросс-идентификации на сервере SDSS.

### **Заключение**

В настоящей статье рассматриваются первые результаты создания промежуточного слоя предметных посредников в гибридной грид-инфраструктуре виртуальной обсерватории для решения научных задач над множеством интегрируемых посредниками неоднородных распределенных информационных ресурсов. Введение такого



промежуточного слоя призвано решить ряд семантических проблем взаимодействия ученого, решающего задачу в некоторой предметной области, и разнообразных релевантных задаче результатов наблюдений и средств их обработки. Гибридная архитектура ВО реализована как соединение инфраструктуры системы поддержки ВО АстроГрид, разработанной в Великобритании, и средств поддержки предметных посредников, созданных в ИПИ РАН. В исследованной архитектуре предметных посредников реализован подход, движимый приложениями, при котором для класса приложений формируется спецификация предметной области независимо от существующих информационных ресурсов.

Полученные результаты свидетельствуют о перспективности исследованного подхода. В настоящее время они используются при решении нескольких задач Российской Виртуальной Обсерватории.

#### **Литература**

- 1 Briukhov D.O., Kalinichenko L.A., Zakharov V.N., Zhelenkova O.P., Malkov O.Yu., Kovaleva D.A. Information Infrastructure of the Russian Virtual Observatory (RVO). Moscow, Russia: IPI RAN. – 2005.
- 2 Williams R., Hanisch B., Linde T. Virtual Observatory Architecture Overview. <http://www.ivoa.net/Documents/Notes/IVOArch/IVOArch-20040615.html> - 2004.
- 3 Kalinichenko L.A., Stupnikov S.A. Constructing of Mappings of Heterogeneous Information Models into the Canonical Models of Integrated Information Systems. Advances in Databases and Information Systems: Proc. of the 9th East European Conference. 2008. - pp. 106-122.
- 4 АстроГрид Project. Retrieved October 24, 2008 from the WWW: <http://www.astrogrid.org> – 2008.
- 5 Брюхов Д.О., Вовченко А.Е., Захаров В.Н., Желенкова О.П., Калининко Л.А., Мартынов Д.О., Скворцов Н.А., Ступников С.А. Архитектура промежуточного слоя предметных посредников для решения задач над множеством интегрируемых неоднородных распределенных информационных ресурсов в гибридной грид-инфраструктуре виртуальных обсерваторий // Информатика и ее применения. – М., 2008. – Т. 2, Вып. 1. – С. 2-34.
- 6 Abrial J.-R. The B-Book: Assigning Programs to Meanings. Cambridge, UK: Cambridge University Press. - 1996.
- 7 Kalinichenko L.A., Stupnikov S. A., Martynov D. O. SYNTHESIS: a Language for Canonical Information Modeling and Mediator Definition for Problem Solving in Heterogeneous Information Resource Environments, M.: IPI RAS. – 2007.
- 8 Briukhov D.O., Kalinichenko L.A., Martynov D.O. Source Registration and Query Rewriting Applying LAV/GAV Techniques in a Typed Subject Mediator. Proc. of the Ninth Russian Conference on Digital Libraries RCDL'2007. Pereslavl-Zalesskij, Russia. – 2007. - pp. 253-262.
- 9 Ohishi M., Szalay A. IVOA Astronomical Data Query Language. <http://www.ivoa.net/Documents/WD/ADQL/ADQL-20050624.pdf> - pdf. 2005.