

АстроГрид Российской Виртуальной Обсерватории как Центр коллективного пользования для решения астрономических задач над распределенными репозиториями информации, накапливаемыми в мире (аннонсирование готовности к использованию)

Российская Виртуальная Обсерватория (РВО) – составная часть Международной виртуальной обсерватории

Для решения проблемы интегрированного использования астрономических данных астрономическое сообщество разрабатывает новый подход к работе с результатами наблюдений, в основе которого находится концепция Виртуальной Обсерватории (ВО). Согласно этому подходу, разрабатываются специальные ИТ-методы и инструменты для интегрированного доступа к неоднородным распределенным архивам и каталогам астрономических данных, равно как и к вычислительным ресурсам и сервисам. Речь идет о разработке информационной инфраструктуры нового поколения, включающей средства ведения групповой научной деятельности учеными на основе средств поиска данных, манипулирования данными, извлечения знаний из данных, содержащихся в огромных, интегрированных распределенных репозиториях информации. Эта инфраструктура должна предоставлять ученым возможности образования новых сервисов анализа совместно используемой информации и реализации децентрализованных вычислительных моделей на основе таких сервисов.

Виртуальные обсерватории интенсивно разрабатываются в мире различными астрономическими институтами: так, к ним относятся Национальная Виртуальная Обсерватория США (NVO), Европейская Виртуальная Обсерватория (Euro-VO), создаваемая координированно рядом стран Европейского Союза, Виртуальная Обсерватория AstroGrid в Великобритании, Канадская Виртуальная Обсерватория (CVO), Японская Виртуальная Обсерватория (JVO), и многие другие. Объединенные усилия координируются Альянсом Международной Виртуальной Обсерватории (IVOA) и Комиссиями Международного Астрономического Союза (IAU).

В ВО работа ученых должна поддерживаться разного рода ИТ-инструментами, удовлетворяющими международным стандартам, разрабатываемым IVOA. В разных странах (США, Великобритании, Германии, Франции, Японии, Китае и других) интенсивно ведутся работы по созданию таких инструментов. Российская Федерация также объявила работы по Российской Виртуальной Обсерватории (РВО) и присоединилась к IVOA.

Информационная инфраструктура РВО

В начале 2005 г. в России объединенными силами ИПИ РАН, САО РАН и ИНАСАН был разработан Аванпроект Информационной Инфраструктуры РВО (РВОИИ) [1]. РВОИИ ориентирована на интегрированное представление информации в различных проблемных областях астрономии для обеспечения научных исследований. Аванпроект содержит анализ разнообразных видов накопленных астрономических информационных ресурсов в России и за рубежом; анализ технологических и архитектурных рекомендаций IVOA;

анализ классов астрономических задач, нуждающихся в средствах ВО; анализ соответствия и достаточности стандартов IVOA для различных видов деятельности РВО; анализ существующих сервисов, которые можно использовать при реализации РВОИИ. На основании проведенного анализа, разработан проект информационной инфраструктуры РВО. Стратегически программа разработки РВОИИ предполагает тесную координацию работ с деятельностью по созданию Международной ВО.

Согласно аванпроекту, РВОИИ должна иметь многоуровневую структуру, на каждом уровне которой должны быть представлены соответствующие этому уровню информационные объекты, средства доступа к объектам, средства их анализа и поддержки пользователей (структура РВОИИ при перечислении уровней от нижних к верхним включает уровень информационных ресурсов, уровень центров данных, уровень виртуальной обсерватории, уровень поддержки решения задач учеными на основе ВО).

Центр коллективного пользования для поддержки решения астрономических задач над распределенными репозиториями астрономической информации и назначение настоящего документа

Одним из первых шагов по реализации РВОИИ является введение в эксплуатацию Центра коллективного пользования для поддержки решения астрономических задач над распределенными репозиториями астрономической информации (содержащими данные наблюдений, результаты решения задач, сервисы анализа данных и знаний). Этот Центр соответствует верхнему уровню РВОИИ, что позволяет использовать его немедленно для решения задач исследователями в области астрономии.

Центр образован в ИПИ РАН на основе системы AstroGrid [2], недавно разработанной в Великобритании и любезно предоставленной авторами для нужд РВО.

Следует иметь ввиду, что АстроГрид – сложная система, которая находится в процессе развития, документация по системе [2] все еще недостаточна, система может содержать некоторые ошибки.

Кроме того, следует понимать, что настоящий документ не является ни описанием системы АстроГрид РВО, ни инструкцией по пользованию системой. Это лишь объявление о готовности системы к использованию астрономическим сообществом России, содержащее краткую аннотацию основных возможностей, предоставляемых АстроГрид, и пример задачи, проектирование которой было выполнено при помощи АстроГрид.

Существенно, что настоящий документ содержит необходимые инструкции для регистрации пользователей в системе и контактные адреса для получения необходимых консультаций при необходимости.

Средства АстроГрид РВО

Основными средствами АстроГрид РВО являются:

- средства поддержки распределенной инфраструктуры РВО на основе технологии Веб-сервисов и в ближайшей перспективе гибридной (информационно-вычислительной) грид-технологии;
- средства поддержки решения задач учеными над накопленной в ВО информацией (включая средства управления потоками работ, средства управления общими пространствами данных при групповом решении задач, специальные средства анализа астрономических данных);
- средства образования новых сервисов анализа информации и реализации децентрализованных вычислительных моделей на основе таких сервисов;
- средства размещения систем управления данными и наборов данных, поддерживаемых такими системами, в центре коллективного пользования;
- средства поиска в каталогах Международной виртуальной обсерватории на основе стандартов IVOA;
- средства публикации результатов астрономических наблюдений, результатов решения задач, программных сервисов в РВО в реестрах метаданных в соответствии со стандартами IVOA;
- средства администрирования сообществами пользователей, принадлежащих различным организациям, исследовательским группам, проектам, и т.п.

Компоненты АстроГрид РВО

Виртуальная обсерватория (ВО) - это не монолитный комплекс программных средств, а открытая система стандартов и интероперабельных модулей, которые можно комбинировать для совместной работы разными способами. Разработчики инструментальных средств следуют стандартным интерфейсам, так чтобы их можно было включить (plug-in) в состав приложения, а центры данных предоставляют возможность доступа к своим данным стандартным образом, равно как и сервисы данных, уподобляемые запросам к базам данных. Основная функция АстроГрид заключается в предоставлении инфраструктурных программных средств, которые могут быть развернуты различным образом. АстроГрид РВО представляет собой специальную систему, которую астрономы могут использовать для выполнения нужной работы по анализу и обработке информации. Все, что требуется для этого - это наличие современного Веб-браузера (Mozilla Firefox, Internet Explorer и другие) и Java 2 Platform (версия 1.4.1 или более поздняя, но не 1.5.x) на машине пользователя. Все программные средства, зарегистрированные в АстроГрид, выполняются либо в Центре коллективного пользования, либо как Java Web Start приложения, которые автоматически можно запускать на машине пользователя.

Существуют четыре основных компонента АстроГрид РВО для пользователей, которые не являются разработчиками новых наборов данных или программных сервисов.

PORTAL обеспечивает средства взаимодействия с системой. Это Веб-страница, которая позволяет войти в систему, получить доступ к остальным компонентам, осуществлять мониторинг задач, и пр. С помощью портала можно также реализовать переход к

произвольным приложениям, разработанным третьей стороной (например, использование средств визуализации изображений Aladin [1]).

Центральную часть АстроГрида составляет WORKFLOW -компонент спецификации и управления потоками работ. Он позволяет образовать композицию сложной последовательности задач (tasks), которые можно выполнять одновременно (например, запрос к базе данных для получения изображений в некотором диапазоне, выбор одного из изображений по некоторому критерию, применение к нему SExtractor [1], слияние каталогов объектов, представление результата в графической форме). Задачи могут выполняться последовательно, совмещенно во времени, циклически. Для спецификации потоков работ предоставляется графическое средство. Система разыскивает необходимые сервисы в реестре и передает им заданные параметры. Потоки работ представляются в XML на основе языка, производного от BPEL. Спецификации потоков работ могут сохраняться, извлекаться с изменением параметров и передаваться системе исполнения заданий (Job Execution System, JES). На каждом шаге потока работ осуществляется вызов ресурсов данных или программных сервисов, зарегистрированных в АстроГрид.

С компонентом WORKFLOW связан интерпретатор потоков работ - Job Execution Server (JES). Спецификации потоков работ обрабатываются асинхронно, возможно, в порядке некоторой очереди. При обработке спецификации потока работ JES принимает ее от WORKFLOW, планирует следующий шаг потока работ, определяет при помощи реестра АстроГрид ресурсы, требуемые для исполнения шага, и в случае успешного его завершения, приступает к планированию последующего шага. В качестве шагов потока работ могут выступать вызовы Java-приложений, приложений командной строки UNIX, вызовы Веб-сервисов, а также запросы к базам данных. Вместе с последовательным или параллельным выполнением шагов, JES осуществляет диспетчеризацию выполнения спланированных шагов, аннотирует выполненные шаги.

Следующим важным компонентом является REGISTRY (реестр) астрономических ресурсов – не только каталогов и архивов изображений, но и приложений, готовых к использованию, и др. К этим ресурсам можно задавать запросы, ресурсы можно просматривать, но, что более важно, программные средства, использующие комбинации ресурсов, могут обеспечивать поиск требуемых ресурсов, используя реестр.

Различаются следующие виды реестров.

Поисковый реестр позволяет производить поиск записей на основе критериев, формулируемых относительно метаданных реестра. Поисковый реестр собирает содержащиеся в нем описания используя процесс, именуемый *harvesting* («сбор урожая»).

Реестр публикации используется для предоставления описаний ресурсов в среде виртуальной обсерватории так, что потом эти описания могут быть использованы при «сборе урожая».

Другим отличительным признаком реестра является его отнесение к *полному* или *локальному* виду. Полный реестр содержит все ресурсы, содержащиеся в ВО. Полным может быть только поисковый реестр. Возможно наличие нескольких таких реестров, каждый из которых соответствует некоторому большому проекту. Локальный реестр содержит лишь подмножество известных ресурсов. Практически, все реестры публикации являются локальными, хотя не исключается существование локальных поисковых реестров, например, ориентированных на определенную научную область.

Наконец, четвертым компонентом АстроГрид является MYSPACE. Это некоторый вид пространства хранения промежуточных результатов, файлов, спецификаций потоков работ, запросов, позволяющий работать в группе и повторно использовать ранее полученные результаты и спецификации.

Для пользователей, которые намереваются разрабатывать новые ресурсы (наборы данных и программные сервисы), дополнительно предоставляются следующие компоненты.

Конструирование приложения: Common Execution Architecture (CEA)

В качестве приложения может выступать любой процесс обработки, который потребляет или продуцирует данные, что технически может включать:

- приложение командной строки UNIX (UNIX command line application);
- запрос к базе данных;
- вызов Веб-сервиса;
- Java-приложение.

CEA можно представлять себе в виде двух основных частей:

1. Модель приложения и его параметры. Сюда включается описание приложения как ресурса в Реестре ВО, так и описание приложения и его параметров для вызова при выполнении. Эти описания представляются в XML Schema.
2. Определение интерфейсов, необходимых для асинхронного вызова приложения. Такие интерфейсы определяются при помощи WSDL.

Доступ к наборам данных: Dataset Access (DSA)

DSA – это компонент, необходимый для включения наборов данных (баз данных) в систему АстроГрид. DSA предоставляет пользователям АстроГрид доступ к базе данных посредством запроса на ADQL [1] или посредством ConeSearch [1]. Также DSA предоставляет набор Веб-сервисов и, что наиболее важно, CEA-приложение, выполняющее запросы к базе данных.

Таким образом, DSA дает возможность пользователю:

- получить в свое распоряжение CEA-приложение, выполняющее запросы к базе данных, которое он может использовать в своем потоке работ;
- надлежащим образом зарегистрировать DSA-компонент в реестре для организации доступа к своим наборам данных, так, что ими могут воспользоваться любые участники АстроГрид.

Модель администрирования

- права пользователя могут назначаться как группам пользователей, так и отдельным индивидуумам;
- создаются сообщества (community) пользователей и группы пользователей;
- пользователь может быть зарегистрирован только в одном сообществе, но может принадлежать разным группам (включая группы, принадлежащие другому сообществу);
- «владелец» группы может назначать права членам группы, включая права доступа к данным, соответствующие группе;

- при вызове задачи инкапсулируются команды, идентичность пользователя, его права по отношению к сервисам или данным;
- провайдеры сервисов могут предоставлять их:
 - о пользователям при помощи портала;
 - о JES при помощи реестров.

Безопасность обеспечивается так, что аутентификация пользователя производится единожды, и она связывается с любой деятельностью, реализуемой этой личностью (в рамках одной сессии). Например, для пользователя, запускающего задание, его идентичность, установленная в результате аутентификации, связывается с каждой задачей вместе с соответствующими правами.

Пример использования системы АстроГрид РВО при решении астрономической задачи дан в [3].

Инструкция о том, как стать пользователем АстроГрид РВО

Для того, чтобы стать пользователем АстроГрид РВО, необходимо послать электронное письмо по адресу ssa@ipi.ac.ru. Письмо должно содержать следующую информацию:

- имя, фамилия, e-mail, [почтовый адрес, телефон] контактного лица, желающего зарегистрировать пользователя или группу пользователей в АстроГрид РВО;
- наименование, [почтовый адрес, телефон] организации, которую представляет контактное лицо;
- список имен пользователей АстроГрид РВО (имя пользователя должно состоять из символов латинского алфавита, числом не менее 5);
- список желаемых имен групп и желаемое разбиение пользователей на группы.

В случае, если вся необходимая информация предоставлена в корректном виде, пользователи (группы пользователей) будут зарегистрированы в сообществе ipi.ac.ru АстроГрид РВО. По указанному контактному электронному адресу будут высланы пароли для входа в портал АстроГрид РВО. Вход в портал осуществляется по адресу

<http://astrogrid.ipi.ac.ru:8080/astrogrid-portal/>

Базы данных, а также программные сервисы могут быть размещены на серверах АстроГрид РВО в ИПИ РАН. Заявку на размещение необходимо послать по адресу ssa@ipi.ac.ru. Заявка должно содержать следующую информацию:

- имя, фамилия, e-mail, [почтовый адрес, телефон] контактного лица, желающего разместить сервис или базу данных на сервере АстроГрид РВО;
- наименование, [почтовый адрес, телефон] организации, которую представляет контактное лицо;
- краткое описание сервиса или базы данных, предлагаемых для размещения.

Дальнейшее согласование по вопросам размещения будет проводиться по электронной почте.

Согласование о порядке проведения возможных консультаций по пользованию АстроГрид РВО осуществляется по электронной почте ssa@ipi.ac.ru или телефонам в Москве (495) 1292098, 1291843.

Литература

1. Briukhov D.O., Kalinichenko L.A., Zakharov V.N., Panchuk V.E., Vitkovsky V.V., Zhelenkova O.P., Dluzhnevskaya O.B., Malkov O.Yu., Kovaleva D.A Information Infrastructure of the Russian Virtual Observatory (RVO). Second Edition. IPI RAN, 2005, 173 p.
2. AstroGrid Release 1.1., <http://software.astrogrid.org>
3. Briukhov D., Kalinichenko L., Martynov D., Stupnikov S., Vovchenko A. (IPI RAS) Zhelenkova O. (SAO RAS). Distant galaxy discovery problem design in AstroGrid and Aladin at IPI RAS in Moscow, <http://synthesis.ipi.ac.ru/synthesis/projects/astromedia/distgal>