

Программные системы и информационные ресурсы для обеспечения астрофизических исследований*

О.П. Желенкова, В.Н. Черненко, В.С. Шергин, Т.А. Пляскина, В.В. Витковский

Специальная астрофизическая обсерватория РАН
zhe@sao.ru, vch@sao.ru, vsher@sao.ru, taap@sao.ru, vvv@sao.ru

Аннотация

Рост объема цифровых данных, Интернет и активное внедрение современных информационных технологий в астрономические исследования привели к созданию инфраструктуры виртуальной обсерватории (ВО). За 10 лет существования ВО обеспечен удобный доступ к панхроматическим обзорам неба, центрам данных и архивам наблюдений наземных и космических обсерваторий с готовыми для научного анализа данными. Координирующая роль в разработках отведена Международному альянсу «Виртуальная обсерватория». Успех этого международного проекта определяет все большую интеграцию данных обсерваторий в инфраструктуру ВО. К инфраструктуре данных должна добавиться инфраструктура телескопов, что потребует стандартизации всех этапов наблюдательного цикла.

В CAO РАН разработаны и действуют несколько информационных систем, обеспечивающих поддержку наблюдательного цикла. Архитектура систем многослойная и они реализованы с использованием веб-технологий и технологий СУБД. Поддержка дистанционных наблюдений требует интеграции этих программных средств в единую систему с учетом тенденций в разработках ВО.

1. Текущий статус ресурсов виртуальной обсерватории

Количество цифровых данных и активное внедрение современных информационных технологий в астрономические исследования привели сначала к появлению концепции, а затем к созданию инфраструктуры виртуальной обсерватории (ВО) [1]. Для программных средств, обеспечивающих функционирование ВО, рабочими группами Международного альянса «Виртуальная обсерватория» (IVOA,

International Virtual Observatory Alliance) [2] ведутся разработка стандартов и рекомендаций более, чем по десятку направлений, в частности: по представлению и формализации данных и знаний предметной области, по разработке протоколов доступа к данным и обмена сообщениями для программных приложений ВО, по стандартам программ для распределенных вычислений (веб-сервисы и грид-сервисы), по описанию и публикации ресурсов, по формату обмена данными, языку запросов и, наконец, по поддержке сохранности данных. С момента появления альянса IVOA в 2003 году разработано около 30 спецификаций протоколов, форматов и соглашений ВО.

Инфраструктура ВО большей частью является сервисно-ориентированной. Веб-сервисы IVOA разделены на три группы: (1) обнаружение и публикация ресурсов, (2) передача и организация запросов данных, (3) сервисы для распределенных вычислений. Обнаружение ресурсов выполняется через регистры ВО. Ресурсами в регистрах считаются не только данные и сервисы, но и организации, проекты, программное обеспечение. Для регистров ВО и спецификации описания ресурсов были рассмотрены несколько промышленных стандартов, и выбран протокол OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). Для описания астрономических ресурсов в регистрах используется стандарт описания сетевых ресурсов Dublin Core (<http://dublincore.org/specifications/>).

Доступ к данным опирается на стандарты Data Access Layer (DAL), определяющие механизмы работы с распределенными астрономическими данными. Эти сервисы, кроме стандартных графических форматов (gif, jpeg) работают с двумя астрономическими форматами: FITS [3] и VOTable [4]. VOTable-формат используется для представления результатов запросов в ВО-сервисах. Его основой является промышленный стандарт XML и опыт разработок астрономических форматов FITS и CDS Astroles.

При обмене информацией между сервисами используется семантический описатель или дескриптор содержимого (Uniform Content Descriptors или UCDs), чтобы установить связь между наименованиями параметров и астрономическими понятиями и/или физическими величинами. Разработан словарь

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

дескрипторов UCDs [5], который поддерживается и контролируется IVOA.

Разработаны стандарты и для сложных сервисов, например, для объединенного запроса сразу к нескольким источникам данных. Обеспечивается такой запрос на базе интерфейса SkyNode [6]. Запрос формулируется Astronomical Data Query (ADQL) [7]. Оператор этого языка переводится в XML-представление. Чтобы выполнить объединенный ADQL-запрос к нескольким астрономическим ресурсам, надо, чтобы ресурсы поддерживали взаимодействие через SkyNode-интерфейс [8].

Веб-сервисы ориентированы на то, чтобы операции над данными в сети выполнялись без участия человека. Повторное использование простых сервисов и комбинирование их для выполнения более сложных действий называется потоком работ. Поточное выполнение использует принципы интероперабельности, когда компоненты потока работ взаимодействуют друг с другом посредством протоколов, определяющих правила запуска сервиса и структуру входных и выходных данных. Реализация таких протоколов опирается на модели данных. Сериализация моделей позволяет провайдером согласованно описывать предоставляемые данные, а применение классов моделей данных в программных средствах ВО позволяет разрабатывать программное обеспечение, работающее со множеством вариантов представления данных без модификации структур данных и самих программ, что существенно упрощает разработку последних. Совместная работа сервисов требует наличия общей памяти, где происходит обмен данными. Такая область памяти называется VOSpace [9]. Эта виртуальная для пользователя, видимая в Интернете память выделяется автоматическим процессом для обмена между задачами.

Дальнейшее развитие инфраструктуры ВО направлено на реализацию методов асинхронных сообщений, авторизацию подписью и управление потоками работ посредством грид-технологий.

Стандарты IVOA используются для создания средств, обеспечивающих доступ, визуализацию и анализ астрономических данных в распределенной среде. Основные цели таких программных систем:

1. грид данных для ключевых панхроматических обзоров неба в астрономии;
2. программное обнаружение активных ресурсов;
3. публикация данных в Интернете;
4. программные интерфейсы пользователя на Java, Perl, Python и скриптовых командных языках bash, позволяющие загружать пользовательский код для запуска собственных алгоритмов;
5. обеспечение веб-интерфейсов пользователя для запроса к архивам и программному обеспечению по интеллектуальному анализу данных;
6. средства для интегрированного on-line анализа выбранных данных;

7. интерактивный анализ и исследования баз данных.

Примерами реализации таких систем являются: Astrogrid (www.astrogrid.org), Aladin (cdsweb.u-strasbg.fr), SDSS SkyServer (cas.sdss.org), SkyView (skyview.gsfc.nasa.gov) и др. Они имеют программные интерфейсы для расширения функциональности систем посредством встраивания пользовательских разработок. Для визуализации данных используются интерактивные программные пакеты ds9, Aladin и др. Редактирование, алгебраические операции со столбцами и их графическое представление удобно выполнять с помощью интерактивного редактора астрономических таблиц TOPCAT (www.star.bris.ac.uk/~mbt/topcat/). Эти инструменты используют протокол обмена сообщениями SAMP (Simple Application Messaging Protocol) [10], что позволяет приложениям обмениваться данными.

В астрономических исследованиях крайне важной является обработка и анализ наблюдательных данных, и системы, предназначенных для решения этой задачи, существовали еще до появления программных средств ВО. Чаще всего используют несколько универсальных систем редукиции астрономических данных, таких как MIDAS (www.eso.org/sci/software/esomidias), IRAF (iraf.noao.edu), AIPS (www.aips.nrao.edu), IDL (www.itvis.com). Они включают большой набор команд и специализированных пакетов, а также встроенный командный язык, который позволяет работать с системами в интерактивном и пакетном режимах, писать программы, приспособляющие возможности систем для обработки наблюдательных данных пользователя. В последнее время для этих систем разработаны программные интерфейсы к Python, Perl и Java в дополнение к уже существующим API для C и FORTRAN. С помощью этих интерфейсов можно интегрировать возможности систем обработки астрономических данных и программных средств ВО.

В инфраструктуру ВО входят центры данных а также архивы космических и наземных обсерваторий, поддерживающие совместимый со стандартами ВО доступ к информации. В качестве примера таких ресурсов можно привести базу данных астрономических каталогов Vizier, которая включает почти 9 тысяч каталогов, собранных Центром звездных данных в Страсбурге (CDS, Centre de Données de Strasbourg) и участвующими в этой работе институтами. В Vizier имеется несколько интерфейсов для запросов к данным: веб-интерфейс, доступ по ASU-протоколу и веб-сервисы. Имеется библиографический сервис, предоставляющий информацию об авторах и публикациях. CDS еще поддерживает базу данных астрономических объектов SIMBAD и портал (cdsportal.u-strasbg.fr) для совместного использования этих двух баз данных.

За десять лет существования IVOA обеспечен удобный доступ к данным панхроматических обзоров и информации центров данных, однако, для интеллектуального анализа данных, интегрированного on-line анализа выбранных пользователем данных,

интерактивного анализа и исследования баз данных пока нет удобных программных инструментов.

2. Информационное обеспечение наблюдений

В ведущих обсерваториях мира достаточно давно используется понятие наблюдательного цикла. Он рассматривается как единый технологический процесс, состоящий из отдельных взаимосвязанных этапов. Сюда входит подача заявок на наблюдательное время, составление расписания, подготовка к наблюдениям, сам процесс наблюдений, архивирование необработанных данных, подготовка и проверка калибровочного материала, верификация правильности и полноты заполнения заголовков файлов, обработка, архивирование научных данных. Чем выше степень автоматизации и интеграции отдельных частей, тем эффективнее используется наблюдательное время телескопа.

В течение ряда лет в Специальной астрофизической обсерватории РАН ведутся работы по созданию информационных систем по поддержке отдельных этапов наблюдений, а именно: архивная система, on-line подача заявок и составления расписания наблюдений, система поддержки электронного журнала наблюдений, система координатной привязки наблюдательных файлов и др.

Информационно-поисковая система общего архива наблюдательных данных

Архивная система общего архива наблюдательных данных [10] состоит из хранилища цифровых данных и информационно-поисковой системы (ИПС) на базе СУБД PostgreSQL, которая поддерживает 16 цифровых коллекций. Для обеспечения сохранности данных при возникающих дефектах на носителях или сбоях оборудования имеются несколько копий данных — две копии на CD/DVD-дисках и две копии архива на жестких дисках. Одна реплицирует оптические диски (A0), а другая копия (A1), несколько по структуре каталогов модифицированная, собственно используется ИПС. В настоящее время все цифровые носители и устройства чтения-записи данных нельзя отнести к устройствам длительного хранения, поэтому для обеспечения сохранности цифровых данных на долгий срок необходимо планировать периодическое переписывание информации. Копии архива A0 и A1 поддерживаются также и для такой процедуры переписывания. Вариант архивных данных (A1) повторен на USB-диске с добавлением дампа таблиц ИПС и программ. Он является резервной копией для восстановления информационной системы при аварийных ситуациях на сервере и/или переносе системы на другую платформу.

Информационно-поисковая система общего архива реализована в трехзвенной архитектуре «клиент»-«сервер приложений»-«сервер БД» и обеспечивает открытый веб-доступ к наблюдательным данным в соответствии с резолюцией МАС о досту-

пе к архивам наблюдений [11]. Каждый файл с наблюдениями описывается в таблицах базы данных 60 параметрами. Они используются для динамического формирования веб-интерфейса, установления соответствия параметров FITS-заголовка и атрибутов таблиц ИПС, идентификации файлов, определения типа файла и т.п.

Часть таблиц схемы базы данных являются справочниками (словарями) и содержат информацию, собранную при анализе данных локальных архивов. Она позволяет интегрировать разнородные цифровые коллекции в одной ИПС. Другая часть таблиц содержит параметры, описывающие каждый наблюдательный файл, включенный в архив. Таблицы пополняются по мере поступления оптических дисков с новыми данными. В архиве нет жестких ограничений на формат файлов, поэтому добавление новых локальных архивов не вызывает трудностей при соблюдении достаточно простых правил.

Для проведения модернизации ИПС поддерживаются две схемы базы данных — тестовая и рабочая. Разработки ведутся на тестовой схеме. После ее проверки ИПС переключается на обновленную версию БД. Оригинальная копия A0 и наличие тестовой схемы позволяют выполнять модернизацию ИПС даже на уровне таблиц.

On-line система подачи заявок и составления расписания наблюдений

Планирование эксперимента - один из этапов наблюдательного цикла, включающий в себя подачу конкурсных заявок и составление расписания наблюдений. Из рассмотренных и принятых КТБТ (Комитет по тематике больших телескопов) заявок составляется календарный план наблюдений на полгода. Для автоматизации этого этапа создана информационная система [12], реализованная на такой же технологической основе, как и ИПС архива. Для нее определены три ролевые группы, различающиеся по сервисам и правам доступа к данным. Роль «пользователь» дает право на чтение/запись/редактирование только своих заявок, регистрацию и настройку персонального бюджета, просмотр расписания наблюдений, получение краткой информации об инструментах. Выполняется контроль по времени при редактировании подаваемой заявки, поскольку заявочная кампания имеет определенные сроки. Роль «эксперт» включает все привилегии и функции «пользователя» и еще просмотр базы данных заявок, получение статистики и генерацию отчетов о программах, участвующих в наблюдениях, затребованном количестве ночей, программах наблюдений, публикациях. Роль эксперта предоставляется членам КТБТ. Для роли «администратор» разрешены все функции экспертов. В дополнение к ним администратор может редактировать базу данных заявок, управлять бюджетами пользователей, заносить и редактировать заявки, поданные в текстовом виде, составлять расписание наблюдений, управлять бюджетами пользователей, выполнять профилактику базы данных и др.

Система поддержки электронного журнала наблюдений

Система предназначена для интерактивного заполнения журналов наблюдений на 6-м оптическом телескопе. Электронный журнал наблюдений реализован в трехзвенной архитектуре, где в качестве клиента используется стандартный браузер, сервера приложений – сервер Apache и сервер базы данных под управлением СУБД PostgreSQL.

Заполнение журнала имеет подготовительный этап. После утверждения расписания наблюдений БТА подготавливаются шаблоны журналов с частично внесенной информацией из базы данных расписания наблюдений (дата наблюдения, название программы, заявитель, ответственный наблюдатель), чтобы облегчить работу наблюдателя и уменьшить количество ошибок при внесении информации.

В системе определены следующие роли:

- *пользователи* — регистрация, вход по паролю, просмотр и поиск журналов наблюдений по дате;
- *ответственные наблюдатели* — регистрация, вход по паролю, просмотр и поиск журналов, заполнение журналов, подпись журнала. Это ограниченная поименным списком группа пользователей;
- *администраторы* — регистрация, вход по паролю, просмотр и поиск журналов, предварительная подготовка текстовых шаблонов протоколов наблюдений, управление информацией о регистрации пользователей: просмотр данных, удаление пользователей, изменение списка ответственных наблюдателей, напоминание пароля.

Система координатной привязки прямых снимков

На настоящее время более половины наблюдений на 6-м оптическом телескопе проводятся дистанционно. При удаленных наблюдениях важно оперативно контролировать наведения телескопа, а также отслеживать наблюдательные условия (оценка прозрачности, качество изображений) для выбора объекта из заявленного программой списка.

Для контроля наведения телескопа и уточнения астрометрической калибровки разработана технология автоматизированной координатной привязки прямых изображений. Для 145-мм рефрактора и 200-мм гида Цейсс-1000 реализован и внедрен программный комплекс экспресс-обработки изображений с использованием авторского пакета программ по масштабно-инвариантному и корреляционному методам привязки координат в соответствии с требованиями FITS-формата. Комплекс использует программные средства виртуальной обсерватории. Система экспресс-обработки построена на принципах многослойной архитектуры для реализации взаимодействия в распределенной среде между клиентом, сервером приложений и сервером-SkyNode, обеспечивающим доступ к базе данных астрономических каталогов.

Алгоритм потока задач протестирован на основе наблюдательного материала вспомогательных телескопов и по архивным прямым снимкам БТА. Он показывает устойчивые результаты для разных по качеству и динамическому диапазону данных. Проведенные разработки могут применяться для подготовки научных данных, коррекции координатных параметров наблюдательных файлов, а также для контроля наведения телескопа.

3. Перспективы интеграции информационных систем поддержки наблюдений

Дистанционные наблюдения требуют все большей автоматизации отдельных частей наблюдательного цикла. Успешная деятельность альянса IVOA в создании инфраструктуры виртуальной обсерватории, поддержанная МАС, направлена на объединение всех имеющихся астрономических данных, включая и архивы обсерваторий. Такое «отчуждение» данных предъявляет более жесткие требования как к организации данных, так и к описанию наблюдений, чем было ранее, когда данные являлись чаще всего собственностью автора программы. Понятно, что без семантического описания невозможно воспользоваться цифровыми данными, собственно представляющими наблюдение. Доставка параметров от разных информационных систем наблюдательного цикла к файлу с наблюдением предполагает проектирование потоков информации. Судя по востребованности астрономических ресурсов, доступных в настоящее время пользователю с помощью средств ВО, все больше обсерваторий будет подготавливать свои данные до совместимости со стандартами IVOA. Еще нет стандартов на отдельные этапы наблюдательного цикла, но они появятся в недалеком будущем стараниями разных рабочих групп IVOA.

В САО РАН, да и вообще в других российских астрономических институтах, не поддерживалась стратегия наблюдения как единого технологического цикла от постановки задачи до получения научного результата, поскольку инструменты обсерватории были введены в строй в 70-е годы прошлого века. В САО РАН активно разрабатывались программные средства, обеспечивающие отдельные этапы наблюдательного цикла. Эти системы приходилось неоднократно менять и модернизировать из-за быстрого развития компьютерных и цифровых технологий. В настоящее время, когда ВО четко обозначены направляющие тенденции, гораздо проще будет определить связующие звенья между разрозненными частями наблюдательного цикла и произвести интеграцию уже разработанных информационных систем.

Литература

- [1] Hanisch, R. J. Prologue: The Virtual Observatory - A New Environment for Astronomical Research /

- Hanisch R. J., De Young D. // *The National Virtual Observatory: Tools and Techniques for Astronomical Research*. ASP Conference Series. 2007. V. 382. С. 1-7.
- [2] Quinn, P.J. The International Virtual Observatory Alliance: recent technical developments and the road ahead / Quinn, P.J.; Barnes, D.G.; Csabai, I.; Cui, C.; Genova, F.; Hanisch, B.; Kembhavi, A.; Kim, S.-C.; Lawrence, A.; Malkov, O.; Ohishi, M.; Pasian, F.; Schade, D.; Voges, W. // *Proceedings of the SPIE*. 2004. V. 5493. P. 137-145.
- [3] Wells, D.C. FITS: A Flexible Image Transport System / Wells, D. C., Greisen, E. W., and Harten, R. H. // *Astronomy & Astrophysics Supplement Series*. 1981. V. 44. P. 363-370.
- [4] VOTable Format Definition, Version 1.2. IVOA Recommendation 30 November 2009 // International Virtual Observatory Alliance, IVOA Documents [Электронный ресурс]. – Режим доступа: <http://www.ivoa.net/Documents/VOTable/20091130/REC-VOTable-1.2.html>.
- [5] The UCD1+ controlled vocabulary, Version 1.23. IVOA Recommendation 02 April 2007 // International Virtual Observatory Alliance, IVOA Documents [Электронный ресурс]. – Режим доступа: <http://www.ivoa.net/Documents/REC/UCD/UCDlist-20070402.html>.
- [6] IVOA SkyNode Interface, Version 1.01. IVOA Working Draft 24 June 2005 // International Virtual Observatory Alliance, IVOA Documents [Электронный ресурс]. – Режим доступа: <http://www.ivoa.net/Documents/latest/SNI.html>.
- [7] IVOA Astronomical Data Query Language, Version 2.00. IVOA Recommendation 30 October 2008 // International Virtual Observatory Alliance, IVOA Documents [Электронный ресурс]. – Режим доступа: <http://www.ivoa.net/Documents/latest/ADQL.html>.
- [8] VOSpace specification, Version 2.0. IVOA Working Draft 28 June 2011 // International Virtual Observatory Alliance, IVOA Documents [Электронный ресурс]. – Режим доступа: <http://www.ivoa.net/Documents/VOSpace/20110628/WD-VOSpace-2.0-20110628.html>.
- [9] Simple Application Messaging Protocol, Version 1.3. IVOA Working Draft 28 July 2011 // International Virtual Observatory Alliance, IVOA Documents [Электронный ресурс]. – Режим доступа: <http://www.ivoa.net/Documents/SAMP/20110728/WD-SAMP-1.3-20110728.html>.
- [10] Витковский, В.В. Состояние и перспективы развития архива наблюдений обсерватории / Витковский В.В., Желенкова О.П., Малькова Г.А., Пляскина Т.А., Шергин В.С. // *Bulletin of the Special Astrophysical Observatory (Известия САО)*. 2005. Т. 58. С. 52-63.
- [11] Public Access to Astronomical Archives // The Resolution of 5 Commission of IAU. 2003 [Электронный ресурс]. – Режим доступа: <http://www.atnf.csiro.au/people/morris/WGAD/Resolution.htm>.
- [12] Витковский, В.В. Комплексная web-система подачи заявки на наблюдательное время и работы с расписанием наблюдений на телескопах САО РАН / Витковский В.В., Желенкова О.П., Малхасян С.В. // *Научно-технический отчет САО РАН N293*. 2003. С. 23.

Software systems and information resources for astrophysical researches

O.P. Zhelenkova, V.N. Chernenkov, V.S. Shergin,
T.A. Plyaskina, V.V. Vitkovskij

The growth of digital data volume, Internet, and to actively introduce modern information technologies in astronomy led to the creation of virtual observatory (VO) infrastructure. Over 10 years of VO existence is easily accessible panchromatic sky surveys, data centers and archives of observations of space and ground-based observatories with science-ready data. Coordinating role in these developments assigned the International Alliance "Virtual Observatory". The success of this international project defines all the great data integration in the VO infrastructure. By the infrastructure of data, must adds the infrastructure of telescopes, which will require standardization of all phases of the observation cycle.

A number of information systems that support the observation cycle has developed in the SAO RAS. The systems have multi-tiers architecture. They are implemented using web and database technology. Support for remote observations requires integration of these software tools into a unified system, taking into account trends in the VO development. 2011. Recommended size is about half of column.

* Работа поддержана грантами РФФИ № 10-07-00412 и № 11-07-00108.