

Состояние и перспективы развития архива наблюдений обсерватории

О.П.Желенкова, В.В.Витковский, Г.А.Малькова, Т.А.Пляскина, В.С.Шергин

Специальная астрофизическая обсерватория РАН, Нижний Архыз, 369167, Россия

Поступила в редакцию 15.07.2004; принята к печати 23.07.2004.

В статье рассматривается текущее состояние архива наблюдений САО РАН. На примере входящих в него локальных архивов проведен анализ параметров, описывающих наблюдения. Рассмотрена стратегия поэтапного включения архива в информационную среду виртуальной обсерватории (ВО) на основе архитектуры, протоколов, интерфейсов, форматов Международного альянса IVOA (International Virtual Observatory Alliance).

Ключевые слова: астрономические базы данных, архивы наблюдений

THE SAO RAS OBSERVATION ARCHIVE STATE AND PROSPECTS OF DEVELOPMENT, by O. P. Zhelenkova, V. V. Vitkovsky, G. A. Mal'kova, T. A. Plyaskina, V. S. Shergin. The paper considers the current state of the SAO RAS observational archives. The parameters describing observations were analyzed on the basis of included local archives. The strategy of stepwise inclusion of the archives into information environment of the virtual observatory (VO) on the base of architecture, protocols, interfaces and formats of the International Virtual Observatory Alliance is considered.

Key words: astronomical data bases, observation archives

1. Введение

Многочисленные астрономические проекты, телескопы и космические миссии дают огромный поток информации для изучения Вселенной. Можно существенно углубить и расширить знания о физических свойствах небесных объектов, изучая их в разных спектральных диапазонах и исследуя долговременные ряды наблюдений с целью поиска переменности. В свободном web-доступе имеются архивы, содержащие наблюдения миллионов объектов. Примерами таких ресурсов являются архивы космических телескопов Хаббла и Чандра, Слоуновский цифровой обзор (SDSS), 2-микронный обзор неба (2MASS) и цифровой Паломарский обзор (DSS). Количество и объемы каталогов, обзоров, архивов настолько велики, что поиск необходимых данных в них занимает много времени и нет возможности полностью копировать информацию на компьютер пользователя. Чтобы решить задачу объединенного использования данных, астрономическое сообщество активно разрабатывает концепцию виртуальной обсерватории (ВО). Более полтора десятков национальных проектов, связанных с созданием виртуальных обсерваторий, объедине-

ны в Международный альянс IVOA (International Virtual Observatory Alliance, <http://www.ivoa.net>).

В IVOA функционирует несколько рабочих групп, которые занимаются разработкой архитектуры ВО (Уильямс и др., 2004), технических основ для публикации астрономических ресурсов, обнаружения и доступа к ним, а также протоколов, интерфейсов, форматов и стандартного языка запросов для обмена и работы с данными. Общепринятые астрономические программные пакеты, такие как IRAF, IDL, MIDAS дают астрономам возможность создавать на их основе инструментарий для обработки и анализа данных. Ядро этих программных систем состоит из набора команд и специального языка, позволяющего строить из последовательности команд более сложные процедуры, учитывая особенности обработки данных. Методология построения ВО опирается на схожую архитектуру. Имеется набор простых базовых функций, реализованных как web-сервисы, из которых пользователь может создавать, как из кирпичиков, новые сервисы для своих целей. Ядро ВО составляют следующие функции: публикация и обнаружение имеющихся в наличии ресурсов, доступ к данным (изображения, каталоги, словари)

посредством простых запросов. Уже сейчас провайдеры больших по объему данных предоставляют к ним web-доступ с помощью протоколов ВО. Для небольших по объему данных разрабатываются программные средства, которые позволяют провайдерам публиковать информацию в Интернете в виде, совместимом со стандартами ВО.

САО РАН обладает двумя крупнейшими в России телескопами — БТА и РАТАН-600, предназначенными для выполнения исследований по приоритетным проблемам астрофизики. В обсерватории более 20-ти лет ведется цифровой архив наблюдательных данных. В конце 80-х в обсерватории была разработана концепция Банка астрономических данных (БАД) САО РАН (Кононов и др., 1990), объединяющая в одну информационную систему наблюдательные данные, астрономические каталоги, цифровые обзоры, технические документы, статьи.

Телескопы обсерватории являются инструментами коллективного пользования, что предполагает интенсивный обмен информацией и требует организации оперативного доступа к ней. Реализация web-доступа к разнородным электронным коллекциям и объединение информационных ресурсов обсерватории — это задачи Интегрированного банка данных (ИБД) САО, который развивает концепцию БАД с учетом достижений информационных и компьютерных технологий.

Для вхождения (интеграции) ИБД в информационную среду ВО можно предложить ряд последовательных шагов, используя спецификации сервисов, протоколы и стандарты, предлагаемые IVOA. Первоначально этот подход применяется к данным общего архива обсерватории, затем его можно будет распространить на другие коллекции (каталоги, обзоры и т.д.), составляющие ИБД.

2. Об архитектуре виртуальной обсерватории

Чтобы перейти к обсуждению архитектуры ВО рассмотрим, какие технологические приемы используются в коммуникационной среде Интернета для вычислений и работы с данными. Напомним, что WWW — это глобальная сеть, посредством которой миллионы компьютеров обмениваются информацией, используя стандартные протоколы. Самым простым и наиболее распространенным способом обмена является комбинация из HTML и HTTP. HyperText Markup Language — специализированный язык, предназначенный для описания гипертекстовых документов, а HyperText Transfer Protocol — протокол передачи гипертекста. Работа в сети заключается во вза-

имодействии между двумя компьютерами с помощью программных инструментов (web-браузер, web-сервер) для передачи информации по этому протоколу и отображения гипертекстовых документов в графическом или текстовом виде. Для распространения информации протоколом HTTP используется модель “клиент-сервер” или принцип “запрос-ответ”. Это означает, что информация хранится централизованно (на сервере), а доступ к ней обеспечивается программой-клиентом по запросу пользователя. Для этого по протоколу HTTP отправляются сообщения от клиента серверу и наоборот.

Более сложным уровнем взаимодействия в сети являются web-сервисы — это приложения (задачи), которые могут быть опубликованы, обнаружены и запущены в Интернете. Они обмениваются сообщениями с узлами (серверами). Это позволяет выполнять процедуры на удаленных компьютерах. Сообщения создаются на XML (eXtensible Markup Language (Extensible ..., 2004)). Это более универсальный, чем HTML, язык для создания гипертекстовых документов, используемый для коммуникационной связи между любыми web-приложениями. В качестве примера можно привести сервисы на основе протокола SOAP (Simple Object Access Protocol). Этот протокол на нижнем своем уровне в качестве транспортного средства использует HTTP. С его помощью передается текстовая информация, в которой по правилам SOAP-протокола с использованием XML описываются структура данных, запрос, правила использования и запуска сервиса. Клиентская программа осуществляет поиск нужной службы SOAP и запускает ее. Поскольку web-сервисы чаще всего выполняются в сети с разными операционными системами и компьютерами, то протоколы, используемые для переноса данных и запуска процедур, должны быть независимыми от среды выполнения. SOAP обладает такими свойствами. Web-узлы, поддерживающие этот протокол, предоставляют web-сервисы, которые взаимодействуют между собой программным путем и не требуют участия человека.

Вычислительные мощности и системы хранения данных на предприятиях и в учреждениях редко полностью загружены. По статистике UNIX-системы бездействуют почти 90% времени, а персональные компьютеры — 95%. Можно, не приобретая дополнительных машин, повысить эффективность работы существующего парка, “научив” связанные сетью компьютеры работать сообща. В этом состоит идея распределенных вычислений. Метакомпьютинг, или распределенные вычисления, — это использование вычислительных ресурсов посредством коммуникационной среды. Такой

уровень взаимодействия между компьютерами в сети реализуется с помощью активно развиваемой технологии Grid (Четти и Бьюа, 2002). Grid — метод использования распределенных процессорных мощностей и распределенных систем хранения, позволяющий загружать работой незанятые вычислительные ресурсы. Эта технология реализуется с помощью набора стандартизированных сервисов, обеспечивающих надежную совместную работу географически распределенных ресурсов, объединяя компьютеры, кластеры, хранилища информации, сети, научный инструментарий и т.д. Пользователя не должно интересовать, где находятся используемые им ресурсы. Важнейшим компонентом Grid-инфраструктуры является промежуточное программное обеспечение (middleware), которое управляет заданиями, обеспечивает безопасный доступ, перемещает и тиражирует данные из одного географически удаленного узла в другой и организует синхронизацию их копий. Следующее поколение Интернета позволит выполнять и нечто большее: со своего компьютера вы сможете заставить работать на себя сеть компьютеров. Роль универсального языка для такого взаимодействия отводится OGSA (Open Grid Service Architecture), открытой архитектуре Grid-сервисов (Фостер и др., 2002).

После этого краткого обзора web-технологий перейдем к архитектуре ВО, предлагаемой IVOA (Уильямс и др., 2004). Основные цели, к которым стремятся разработчики ВО, — это открытие новых закономерностей и новых классов объектов, реализация распределенных вычислений, публикация в сети имеющихся в наличии информационных ресурсов и результатов вычислений, а также обеспечение совместной работы научных групп, занимающихся одной проблематикой. Это реализуется с помощью порталов (web-сайты со специализированной настройкой на свою аудиторию, выполняющие роль отправной точки или входа при работе с информацией), унифицированных пользовательских интерфейсов и специализированного программного обеспечения. В качестве примера можно привести уже реализованные программные пакеты и системы: VOPlot, Mirage, Topcat и пользовательские интерфейсы и порталы: SkyQuery, Aladin, Oasis. Взаимодействие с разнородными источниками данных осуществляется через интерфейс с помощью web-сервисов, опирающихся на HTTP, SOAP и Grid стандарты W3C консорциума. Эти сервисы выбраны альянсом IVOA в качестве основных технологий реализации ВО.

HTTP, SOAP и Grid обеспечивают разную степень производительности и функциональности при работе с информацией. HTTP-сервисы являются самым простым и широко распространенным

средством передачи данных, но они не обеспечивают необходимой для ВО гибкости и масштабируемости при запросе информации. SOAP расширяет возможности HTTP, что позволяет организовывать межресурсное взаимодействие без участия человека. Grid используется для управляемой и авторизованной передачи больших объемов информации и распределенных вычислений.

Сервисы ВО разделены на три широких класса: сервисы регистров для обнаружения и публикации ресурсов, сервисы для передачи и организации запросов данных и сервисы для вычислений и объединения данных. Остановимся более детально на каждом из них.

Обнаружение данных выполняется через регистры ВО (Плейнт и др., 2004), которые содержат описания ресурсов и являются важной составляющей в эффективном поиске информации в распределенной среде. Регистр — это описания ресурсов, представленные в виде структурированных метаданных, чтобы обеспечить автоматизированный поиск и обработку запросов. Ресурсами считаются данные и сервисы, но можно рассматривать и другие объекты, к примеру, организации, проекты, программное обеспечение. Запрос к ресурсу может выполняться по следующим параметрам: тип (каталог, архив изображений, образовательный ресурс), область неба, время, длина волны, а также источники данных. Модель, разработанная IVOA, содержит три типа регистров. Поисковый регистр предназначен для пользовательских приложений и содержит все описания ресурсов, имеющих в ВО. Для этого типа регистра выполняется специальный процесс сбора информации о ресурсах (harvesting) из множества регистров. Регистры для публикации отличаются от поисковых тем, что они не поддерживают поиск, они просто выставляют в Интернете описания ресурсов. Они могут многократно пополнять поисковые регистры с помощью процесса, который синхронизирует их содержание. Третий тип — это специализированные поисковые регистры. Они не содержат всей информации ВО. Специализированные регистры предназначаются для особого типа ресурсов или научной тематики, к примеру, связанной с исследованием сверхновых. Они выполняют селективный сбор по другим регистрам, обновляя имеющуюся информацию.

Сервисы данных обеспечивают запрос и доступ к информации. Самый распространенный запрос — поиск информации о выбранном участке неба (указывается направление и размер области) и получение изображений, спектров или списка объектов, туда попадающих. Запрос выполняется с помощью протокола Simple Image Access (SIA) (Тоди и Плейнт, 2004), который имеет определенный

формат обращения к ресурсам, поддерживающим сервис, и получает изображения указанной области неба. Это простой сервис, и он обращается к одному источнику данных. Сервисы данных, кроме стандартных графических форматов (gif, jpeg), работают с двумя астрономическими форматами. Один из них — это Flexible Image Transport System (FITS) (Уелс и др., 1981), являющийся астрономическим стандартом для хранения и обмена данными, и второй — VOTable формат (Ошенбайн и др., 2004), используемый для вывода полученных при запросе данных. VOTable соответствует стандарту XML (Extensible ..., 2004) и предназначен для описания и представления астрономических таблиц. Астрономы используют разные названия для обозначения одних и тех же физических величин и параметров. Чтобы избежать неоднозначности при интерпретации величины, необходимо определить, что именно обозначают разные названия. Особенно это важно при обмене информацией между сервисами. Поэтому для определения семантического типа величины предлагается использовать разрабатываемый и контролируемый IVOA-словарь, который называется Unified Content Descriptor (UCD) (Дерьер и др., 2004). В словаре закрепляются фиксированные наименования физических величин и даются их определения. При разработке спецификаций сложных запросов на более высоком семантическом уровне UCD будет важной частью в них.

IVOA предполагает разработку стандартов для сложных сервисов, к примеру, для федерации (организации объединенных запросов к нескольким источникам данных) и поиска информации в каталогах, обработки изображений и обнаружения объектов, статистического анализа и визуализации сложных по структуре данных. Для запроса к реляционной базе данных или к нескольким базам данных разрабатывается OpenSkyQuery-протокол (IVOA SkyNode Interface..., 2004). В этом случае запрос записывается на языке, близком к Structured Query Language (SQL), который обычно используется при работе с базами данных. Этот Astronomical Data Query Language (ADQL) (Исуда и др., 2004) представляет собой упрощенный вариант SQL. Оператор этого языка переводится в XML-представление. Если имеется несколько астрономических ресурсов, которые поддерживают этот протокол и обеспечивают web-сервис, то можно сразу выполнить к ним объединенный ADQL-запрос.

Третья группа сервисов связана с вычислениями. К примеру, мы получаем по запросу из цифрового обзора или архива наблюдений изображение участка неба. Для этого может быть необходимо построить мозаику из нескольких изоб-

ражений, выровнять фон, развернуть и сложить отдельные изображения для совмещения, то есть выполнить несколько простых операций. Желательно, чтобы в процессе формирования ответа на запрос эти операции выполнялись без участия человека. Такие действия называются потоками или потоками работ/задач (workflow). В разработках IVOA используется программное обеспечение для обработки информации, представляющее собой связанный набор распределенных сервисов с потоковым принципом выполнения. Поточковая парадигма для работы с распределенными данными состоит в повторном использовании простых сервисов для построения сложных приложений. Причем компоненты приложения изолированы друг от друга посредством хорошо определенных протоколов, а именно определены правила для запуска, а также структура входных и выходных данных. Аналогичный принцип можно использовать для виртуальных данных, то есть таких данных, которые динамически создаются в случае необходимости и сохраняются некоторое время во временной памяти (кэше), что позволяет использовать их повторно при аналогичных запросах.

3. Текущее состояние общего архива наблюдений САО

Архив САО за время своего существования прошел несколько этапов представления и хранения данных, предоставления сервисных функций. Эти этапы тесно связаны с развитием компьютерных и информационных технологий, а также с необходимостью интеграции с астрономическим сообществом в проведении научных исследований.

В табл. 1 приведена информация по объему оптических и радиоданных. На рис. 1 показано изменение общего объема архива наблюдений (гигабайты) САО с 1996 по 2003 год, а также ежегодный темп прироста.

Можно рассмотреть этапы существования архива наблюдений в виде цепочки:

локальный архив → **основной архив** → **общий архив** → **IVOA совместимый архив**,

что позволяет проследить процесс объединения разнородных по форматам и структуре наборов данных, а также организацию доступа к файлам с наблюдениями.

Локальный архив — это цифровая коллекция данных, получаемых одним или несколькими сходными методами наблюдений. С 1988 г. для описания и хранения наблюдательных данных в обсерватории используется FITS-формат (Витковский и др., 1988). FITS-файл состоит из блоков с символьной и двоичной информацией. Блоки

Таблица 1: Архив наблюдательных данных САО РАН (CD-диски) на 15.12.04

Архивы	Число CD-дисков	Среднесуточный поток данных	Объем архива	Число файлов
Оптика	111+111 копий	85.2МВ	165.4ГБ	136589
Радио	7	4.3МВ	3.6ГБ	45921
Архивы пользователей	30 копий			

SAO archive data volume

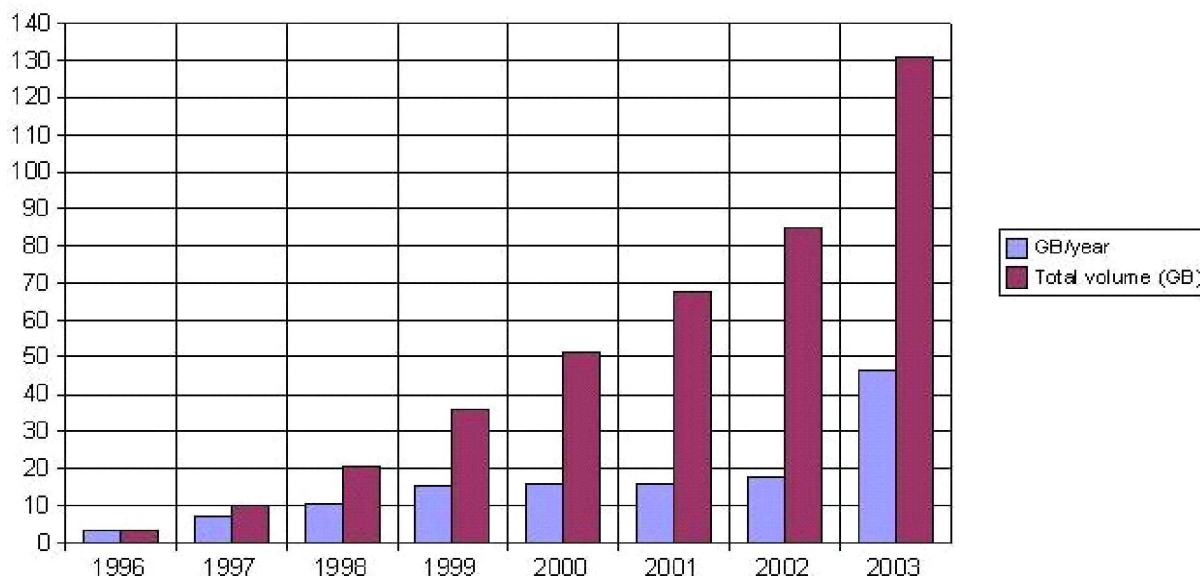


Рис. 1: Изменение объема архива наблюдений САО РАН.

с символьной информацией содержат набор пар “ключевое слово — его величина”, которые описывают параметры файла. Собственно наблюдение — это двоичные данные. Ключевые слова FITS-формата можно считать метаданными при описании двоичного содержимого файла. Для наблюдательных данных обсерватории был определен набор ключевых слов, а также зафиксирована и отражена в заголовке структура двоичной части файлов для разных локальных архивов. Следствием многопрограммности телескопов является существование различающихся между собой по многим параметрам методов наблюдений. Каждый метод связан с определенным компьютерно-аппаратным комплексом — системой сбора. Формат цифровых данных для системы сбора разрабатывается группой, занимающейся созданием прибора, поэтому в файлах разных методов нет единого набора ключевых слов для описания параметров наблюдений. Использование FITS-формата для локальных архивов можно назвать первым шагом к объединению разнородных данных.

Основной архив — это коллекция локальных

архивов. С 1994 года наблюдательные данные, полученные на инструментах обсерватории, хранятся на оптических дисках. Физическая единица хранения в архиве — оптический диск. Наблюдатель определяет, что относится к наблюдениям и должно сохраняться в архиве. Администратор формирует образ диска, при необходимости может внести дополнительную информацию и записывает архивные данные на оптический диск. В итоге на диски записываются: файлы с наблюдениями объектов, сервисные файлы, используемые для коррекции инструментальных ошибок и калибровок, журналы наблюдений, сопутствующая информация, подготовленная наблюдателем, а также дополнительная информация — текстовые справочные файлы, информация, используемая для идентификации диска, программное обеспечение, связанное с содержимым диска, контрольные суммы. В настоящее время в основной архив входят 16 локальных архивов, которые различаются по форматам, ключевым словам и структурам данных. В табл. 2 и 3 приведены основные характеристики локальных архивов.

Таблица 2: Локальные архивы, входящие в состав общего архива наблюдательных данных САО

Локальный архив	Телескоп	Число CD	Период
LYNX	6m	7	1996.02 - 2001.12
NES	6m	23	2001.07 - 2004.05
PFES	6m	1	1999.12 - 2001.01
MPFS	6m	17	1996.08 - 2004.02
IFP	6m	2	1997.05 - 2000.03
MOFS	6m	1	1997.03 - 2001.08
SCORPIO	6m	18	2000.09 - 2004.04
UAGS	6m	8	1994.11-2003.09
CCD	6m	4	1996.02 - 2000.04
SP124	6m	2	1996.02 - 1999.05
MSS	6m	1	1996.05 - 2001.01
ZMCCD	z1000	18	1996.12 - 2004.04
ZUAGS	z1000	3	1998.04-2000.09
CEGS	z1000	1	1997.03 - 2000.12
Z600	z600	8	1996.01 - 2001.06
RATAN	RATAN-600	7	1996.06 - 1999.01

Из опыта работы с такой коллекцией были определены правила, которыми предполагается руководствоваться при работе с архивом. При этом необходимо учитывать возможность добавления новых цифровых коллекций, старых наблюдательных данных и наблюдений, полученных в других обсерваториях. Вот эти правила:

- архив является прозрачным для пользователя, не меняет форматы и параметры хранящихся данных; в каком формате поступили данные на вход архива, в таком они выдаются при запросе; дополнительно имеется возможность получения данных в FITS-формате;
- в архиве хранятся необработанные данные, записанные на оптических дисках; доступ предоставляется ко всему, что записано на дисках;
- логической единицей хранения в архиве является наблюдение;
- доступ к данным разрешается на основании Положения об архиве наблюдательных данных САО РАН (http://www.sao.ru/Doc-k8/Science/arch_regul.html).

Общий архив — это основной архив с организацией web-доступа. Для реализации такого доступа к локальным архивам проанализированы разные типы файлов с наблюдениями, записанные на оптические диски. Основная проблема, возникающая при объединении архива, — это различие форматов, параметров и структуры данных. Для ее решения используется информационно-поисковая система (ИПС) на основе реляционной системы управления базами данных (СУБД) (Витковский

и др., 2000).

Основным источником информации о наблюдении является заголовок файла. Для описания наблюдения разные системы сбора в оптике используют от 0 до 75 параметров, в радиодиапазоне — до 289. Величины этих параметров формируются в системе управления телескопом, инструментом и в системе сбора. Часть параметров поступает в заголовок файла автоматически, часть заносится наблюдателем. Поскольку системы сбора модернизируются, то меняются форматы данных и параметры, описывающие наблюдение, поэтому каждый архив включает в себя не одну версию формата.

Из анализа данных, записанных на оптический диск, можно получить информацию о параметрах наблюдения. Их значения имеются в полном имени файла, в тексте, полученном при выполнении системной команды `ls`, ключевых словах (дескрипторах) из заголовка файла, протоколах ночи, `log-файлах`, расписании наблюдений.

Информация может многократно дублироваться, но иногда ее может дать или уточнить только автор программы наблюдений. Для некоторых наблюдений имеется только полное имя файла. По этому имени, в которое включен путь к файлу от корневого каталога оптического диска, можно определить дату наблюдения, метод компрессии, тип изображения, фильтр, формат. Системная команда `ls` дает информацию о размере файла, о времени записи (копирование на оптический диск).

Из интервью с пользователями определено несколько типов запросов архивных данных. В

Таблица 3: Характеристики локальных архивов

Локальный архив	Объем (МВ)	Число файлов	Средний размер файла (КВ)	Число ночей	Среднесуточный поток данных (МВ)
LYNX	8790	5397	1700	291	30
NES	28334	4529	6400	137	207
PFES	1248	771	1700	36	35
MPFS	20830	10819	2000	186	112
IFP	2180	4502	500	28	78
MOFS	1286	1059	1200	26	49
SCORPIO	21730	21274	1000	189	115
UAGS	8678	19809	450	397	22
CCD	4988	6012	800	127	39
SP124	2338	8162	300	173	14
MSS	1284	2986	400	153	8
ZMCCD	22958	21737	1100	345	67
ZUAGS	364	3855	100	107	3
CEGS	462	220	2200	18	26
Z600	10134	15968	650	415	24
RATAN	3554	45921	80	861	4

дальнейшем эти типы назовем стандартными. К ним относятся запросы по дате наблюдения, прибору, типам файлов, координатам наблюдаемого поля/объекта, имени астрономического объекта, программе наблюдений, автору программы и наблюдателям. Для выполнения стандартных запросов необходимо наличие соответствующих параметров в заголовке файла. Локальные архивы были проверены на достаточность ключевых слов в заголовке, на необходимость привлечения других источников и каких именно, определены правила (связь между ключевыми словами), по которым возможен поиск данных, соответствующих запросу.

Из-за ошибок заполнения заголовков наблюдательных файлов, неполного количества ключевых слов, отсутствия необходимых параметров нет возможности организовать стандартные запросы ко всем без исключения архивным файлам, кроме одного типа запроса – по дате наблюдения. Другие типы запросов можно организовать только к части локальных архивов.

Теперь рассмотрим процедуру занесения оптического диска в архив. Сначала диск копируется на винчестер архивного сервера и просматривается администратором на предмет особенностей записи. Данные на винчестер с оптического диска записываются неким стандартным образом, а именно наблюдения, полученные в течение одной ночи, записываются в один каталог, в названии которого имеется дата. Логической единицей хранения в ИПС считается одно наблюдение, поэтому несколько наблюдений, собранных в один архив-

ный файл, разделяются. Действия, которые выполняются с диском для приведения его к стандартному виду, записываются в командный файл или программу подготовки CD-диска.

Администратор заносит в ИПС данные о новом диске, следит за состоянием сервера базы данных, оптических и жестких дисков с архивной информацией. Для администратора необходима информация о каждом оптическом диске: размер диска, число файлов и каталогов, полные имена файлов, контрольные суммы. Пользователи архива имеют ftp-доступ к наблюдательным данным. Администратор, кроме этого, имеет еще доступ к данным другого типа (программное обеспечение, справочная и служебная информация на дисках).

Для архива наблюдений используется специализированный сервер, на котором размещены информационно-поисковая система на основе СУБД Oracle и хранилище файлов. Реализована структура таблиц для запроса по дате, который можно выполнить для всех архивов. Наполнение таблиц выполняется программами, написанными на языке Perl. Этими программами выполняется анализ текстовых файлов, полученных с помощью команды ls, примененной к оптическим дискам. Интерфейс пользователя для доступа к данным реализован на основе спецификаций CGI-интерфейса (Common Gateway Interface), DBD (DataBase Driver) и DBI (DataBase Interface) интерфейсов к СУБД. Таблицы пополняются по мере поступления CD-дисков с новыми данными. Схема таблиц приведена на рис. 2.

Сервисные функции, предоставляемые ИПС

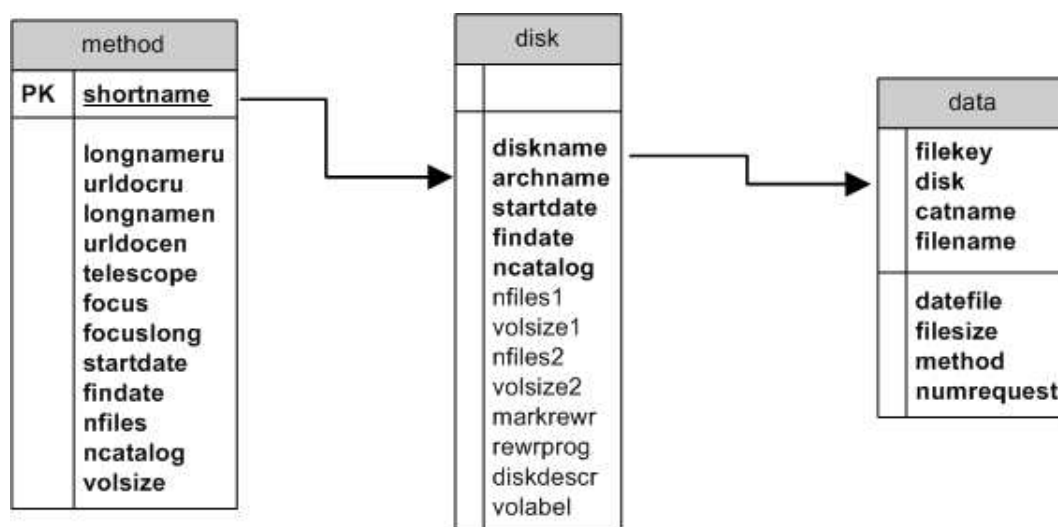


Рис. 2: Схема таблиц ИПС для запроса по дате.

пользователю, следующие: интерфейс с использованием web-браузеров, получение выбранных данных по сети, запросы по дате, методу и дате, просмотр заголовка FITS-файла, предварительный просмотр данных файлов в FITS-формате (генерируется черно-белая картинка в jpeg-формате), получение данных наблюдательного сети из архива расписаний (по названию программы наблюдений). Предполагается расширять сервисные функции ИПС, добавляя новые типы запросов к общему архиву. Разработанная информационная система объединила разнородные цифровые коллекции, которыми являются локальные архивы, и дала возможность обращаться к ним с помощью общего web-интерфейса.

4. Архив обсерватории как информационный ресурс, совместимый с требованиями IVOA

Астрономическое сообщество активно использует новые компьютерные технологии для создания среды, которая объединила бы все имеющиеся данные в единое хранилище информации. С момента появления проектов, связанных с ВО, разрабатываются и уточняются спецификации протоколов, форматов и типов запросов, используемые для создания программных продуктов, позволяющих работать с астрономическими данными в Интернете. Эти стандарты уже применяются в ряде центров астрономических данных для доступа к архивам наблюдений, обзорам, каталогам. В ежеквартальном отчете NVO (Building . . . , 2004) отмечено, что в регистре ВО сейчас зарегистрировано несколько тысяч астрономических ресурсов. Через три-

четыре года публикация данных в Интернете с использованием IVOA-стандартов станет обычной и необходимой процедурой. Важно, чтобы архив CAO вошел в информационную структуру виртуальной обсерватории. Это требует выбора спецификаций и тех способов и методов включения, которые подходят для нашей коллекции.

Напомним, что информационная структура или, иными словами, среда, инфраструктура виртуальной обсерватории включает в себя компьютеры и хранилища данных, соединенные WWW, и несколько взаимопроникающих слоев программного обеспечения, базирующихся на общепринятых и специализированных спецификациях и стандартах, для передачи информации по сети, работы с базами данных, а также интерфейсы пользователей для доступа, визуализации и анализа данных. Для включения астрономических данных в эту структуру необходимо решить несколько задач. Сначала надо получить или представить информацию в цифровом виде, затем ввести набор данных в компьютер, что можно делать разными способами: от простого — запись файлов на жесткий диск, до сложного, но более гибкого по функциональности и надежности — создание базы данных. После создания web-интерфейса к данным требуется “объявить” сообществу о своем существовании, чтобы web-сервисы, выполняющие запрос данных, могли обнаружить ресурс без посредничества человека. Для этого цифровая коллекция заносится в ВО-регистр, к которому будут обращаться сервисы, обслуживающие интерфейс пользователя.

В проекте NVO процесс регистрации астрономических ресурсов еще только тестируется. Имеются два регистра, в которых публикуются астро-

номические ресурсы — California Institute of Technology (Caltech) и National Center for Supercomputing Applications (NCSA) (Плейнт и др., 2004). На каждом, заполняя web-форму, можно зарегистрировать свой источник данных. Описания ресурсов, введенные пользователями, хранятся как XML-документы. Поисковые регистры используют для сбора информации специальный процесс (harvesting) на основе протокола Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (The Open Archives..., 2002).

Если владелец данных имеет несколько статичных, редко изменяемых коллекций, то проще всего воспользоваться существующими регистрами и управлять описаниями своих ресурсов с помощью сайта Caltech или NCSA. Когда у провайдера имеется несколько десятков наборов данных, которые меняются со временем и необходим контроль над описаниями, тогда на его сайте создается свой регистр. Этот способ регистрации больше подходит для архива наблюдений обсерватории, поскольку позволяет по своему усмотрению работать с включаемыми ресурсами. Подходит он также для публикации цифровых коллекций, включаемых в ИБД, которые отличаются по структуре данных от локальных архивов, а именно для каталогов, обзоров.

Пока еще нет готовых программных решений для создания группового локального регистра на сайте любого участника IVOA, но в скором времени они появятся. Поэтому для локальных архивов, входящих в состав общего архива CAO РАН, сначала предполагается выполнить регистрацию на одном из вышеназванных сайтов, а затем перенести их описания в локальный регистр CAO РАН.

Для интеграции информационных ресурсов важно найти методы добавления новых коллекций в среду ВО, так чтобы можно было через унифицированный интерфейс извлекать запрашиваемые данные. Когда коллекция внесена в регистр, появляется информация о том, что ресурс имеется в наличии, но работать с самими данными еще нельзя. Для поиска и извлечения данных из этой коллекции необходимо реализовать web-сервис. Такая служба создается на основе соответствующего протокола. Разработаны спецификации следующих протоколов: ConeSearch — извлечение данных из каталога для объектов, попадающих в указанную область неба, SIA — получение изображений области неба, для спектров — Simple Spectrum Access (SSA) (Доленски, Тоди, 2004). В регистр заносится информация о том, какого типа сервисы обеспечиваются для ресурса, чтобы web-сервисы, которые выполняют поиск по всем опубликованным коллекциям, могли по описаниям ресурсов без посредничества человека работать с данными. К

примеру, на сайтах Caltech и NCSA, где находится регистр ВО, имеется список цифровых коллекций, из которых можно таким способом извлекать запрашиваемые изображения. Поэтому, кроме регистрации архива обсерватории, еще необходимо реализовать для него аналогичные web-сервисы. Это позволит стать архиву частью интегрированных астрономических ресурсов.

В локальных архивах хранятся изображения участков неба и спектры небесных объектов. Для извлечения наблюдений из общего архива требуются web-сервисы, позволяющие получать из локальных архивов изображения указанного участка неба или спектры объектов, туда попадающих. Для создания этих сервисов используются протоколы SIA и SSA. Протокол — это набор правил для клиента и сервера. Остановимся более подробно на SIA-протоколе. Сервис SIA должен по требованию клиента передать изображение участка неба заданного размера. В идеальном случае он предоставляет астроному некий участок виртуального неба, который реально может состоять из нескольких цифровых изображений, покрывающих эту область, и пользователь не беспокоится о стыковке границ отдельных кадров.

При получении данных должны выполняться следующие действия: запрос изображения, подготовка изображения для передачи и собственно передача. Этот сервис подразделяется еще на четыре категории по типу изображений и операций с ними. Коротко о каждом типе сервиса.

1. *Image Cutout* — сервис предназначен для вырезания областей из больших по размеру цифровых изображений, обычно из цифровых обзоров неба. Пользователю возвращается требуемый участок, который может быть составлен из мозаики фреймов. С полученным изображением не проводится никаких дополнительных операций, к примеру, изменения масштаба, пересчета размеров пикселя и проекции. Исходные данные не изменяются.

2. *Image Mosaicing* — сервис похож на предыдущий, но дополнительно проводится заданная пользователем обработка изображения по изменению масштаба, размера пикселя, приведению к другому типу проекции. Имеется возможность генерировать пиксели изображения из данных, получаемых в других диапазонах, и выполнять наложение изображений. При этом изменяются первоначальные данные, и сервис работает медленнее *Image Cutout*.

3. *Atlas Image Archive* — осуществляется доступ к уже обработанным изображениям разного типа обзоров. Не проводится извлечение участков из изображений. Пользователю возвращается под-

ходящее по условиям запроса изображение из обзора.

4. *Pointed Image Archive* — сервис обеспечивает доступ к коллекциям изображений небольших областей неба, в которых целью чаще всего является некоторый конкретный объект. Этот тип сервиса используется для архивов наблюдений.

Из перечисленного выше для архива CAO под-ходит *Pointed Image Archive*. Далее при рассмотрении спецификации более подробно остановимся на параметрах и правилах для этого типа сервиса, а также на последовательности выполняемых им действий.

Первое действие — это запрос изображения. Размер, масштаб, тип проекции могут использоваться в качестве дополнительных параметров для уточнения запроса. Сервис должен вернуть URL-ссылки на изображения, наиболее подходящие условиям запроса. Ввод данных выполняется как GET-запрос HTTP-протокола. Выглядит это следующим образом:

```
http:// <server-address> / <path-to-service-
program>?[extra-GET-arguments]&[...]
```

Сервис должен передавать два параметра, которые определяют координаты центра и размеры прямоугольного участка неба. Координаты центра задаются в градусной мере в системе координат ICRS (International Celestial Reference System), что соответствует каталогу FK5 на эпоху J2000.0. Размеры области задаются также в градусах. В запросе может передаваться параметр INTERSECT, который определяет, каким образом выбранные изображения должны совпадать с запрашиваемой областью неба. Для архива CAO этот параметр опускается, поскольку по умолчанию полагается, что INTERSECT=OVERLAY, то есть изображение удовлетворяет условию запроса, если частично перекрывается с запрашиваемой областью.

В строке запроса для второго типа сервиса используются дополнительные параметры, которые определяют желаемый размер, масштаб, ориентацию и проекцию изображения. Другие типы SIA не должны выдавать сообщение об ошибке при наличии этих параметров в строке запроса. Сервис должен поддерживать параметр FORMAT, который отмечает формат или форматы полученных изображений. Они могут быть следующие: fits, html, jpeg, png.

Ответ на запрос должен выдаваться в виде таблицы в формате VOTable. В ней передаются все изображения, удовлетворяющие условиям запроса. В спецификации определяется, какие элементы VOTable-формата являются обязательными. Для описания каждого изображения в таблице используется одна строчка с набором параметров

в виде UCD-дескрипторов. Передается информация, идентифицирующая изображение, координаты и размеры, спектральный диапазон, действия, произведенные с изображением. Координаты изображения представляются в упрощенном варианте FITS WCS (World Coordinate System) (Грейсен, Калабретта, 2002). Сервис должен выдавать сообщения об ошибках в параметрах, в процессе передачи.

Следующий шаг связан с подготовкой изображений. Сервер извлекает данные из хранилища, помещает их во временную память, расположенную на другой машине, или производит динамическую генерацию изображения (масштабирование, поворот и тому подобное). Этот процесс является существенным при передаче больших по объему файлов или извлечении данных из больших хранилищ. При передаче данных из архива CAO этот шаг можно не рассматривать. Он будет выполняться стандартными средствами операционной системы.

Последний этап — это передача данных. В этой версии спецификации клиенту передается URL-ссылка на каждое изображение, по которой он может получить данные, используя механизмы Интернета.

Этих правил необходимо придерживаться при реализации web-сервиса для архива обсерватории.

Для выполнения SIA-запроса в структуру таблиц ИПС необходимо включить следующие параметры наблюдений: дата наблюдения, тип файла, телескоп, прибор, детектор, система сбора, координаты объекта, размер фрейма. На рис. 3 представлена схема таблиц для извлечения из архива изображения участка неба. Координатный запрос применим к части коллекций, поскольку не в каждом локальном архиве данные имеют набор координатных параметров. Для локальных архивов имеются варианты в формате и названии параметров, которые представляют одну и ту же физическую величину. К примеру, дату наблюдения в разных цифровых коллекциях можно получить из значений следующих ключевых слов: "DATE", "DATE-OBS", "Date of observation", "OBS-DATE". Такое положение с параметрами усложняет процедуру наполнения таблиц ИПС-архива.

Список изображений, удовлетворяющих запросу, выводится в виде таблицы, в столбцах которой записываются параметры каждого файла. Эти параметры являются UCD-дескрипторами. Поэтому для локальных архивов необходимо устанавливать соответствие параметров наблюдений с элементами UCD-словаря. Это выполняется через локальный словарь, где сопоставлены внутренние названия параметров наблюдений и соответствующие UCD-элементы.

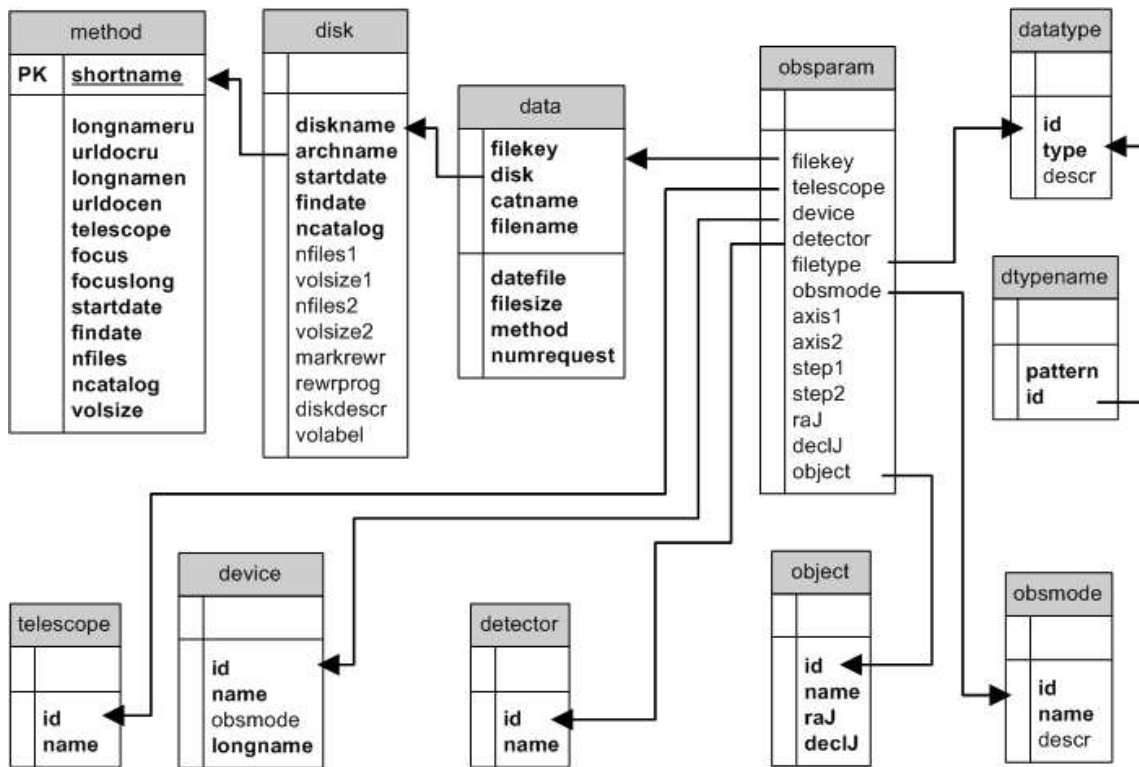


Рис. 3: Схема таблиц ИПС для запроса по координатам.

Для включения новой цифровой коллекции в общий архив обсерватории необходимо выполнение следующей процедуры. Коллекция публикуется в регистре обсерватории. Описание коллекции выполняется по правилам регистрации информационных ресурсов IVOA. Затем находится соответствие новых параметров, которые используются в данной коллекции для описания наблюдений, ключевым словам локального словаря и, если необходимо, производятся добавления. На основе выбранной спецификации реализуется web-сервис.

5. Заключение

В состав общего архива САО РАН входят 16 локальных архивов. В нем хранятся на оптических дисках необработанные наблюдательные данные. Общий объем архива (на 15.12.2004 г.) — 170 гигабайт (сжатые данные) при среднесуточном потоке данных 90 мегабайт. Информация о текущем состоянии архива имеется на сайте обсерватории (<http://www.sao.ru/oasis/archive/docs/boacs.html>). Наиболее активно увеличиваются объемы локальных архивов следующих инструментов 6-метрового телескопа: NES, Scorpio, MPFS, 1-метрового телескопа — ZMCCD.

В отделе информатики ведутся поэтапные ра-

боты по объединению разнородных цифровых коллекций для выполнения запросов к ним. Для этого разработана структура таблиц ИПС. Хранение, обновление, запросы к ним обеспечиваются с помощью СУБД. Созданы программы для наполнения таблиц и унифицированный web-интерфейс доступа к данным. На первом этапе реализован запрос к локальным архивам по дате наблюдения с предварительным просмотром данных и заголовков файлов. Идет разработка структуры и наполнения таблиц, программного обеспечения для запроса по координатам.

Астрономическое сообщество активно использует Интернет и новые информационные технологии для извлечения информации из имеющихся астрономических ресурсов. Международный альянс IVOA направляет и координирует работы сообщества по интеграции разнородных данных, разработке стандартов и спецификаций. Эти работы объединены общим понятием “виртуальная обсерватория”. Через некоторое время процедура публикации любого астрономического ресурса (каталога, обзора, архива, web-сервиса, сайта обсерватории и т.п.) в регистре ВО будет стандартным действием, так же как и применение специальных сервисов для работы с ним.

В настоящее время необходимо рассмотреть и опробовать спецификации IVOA для объединения

цифровых коллекций обсерватории. В частности, из имеющихся спецификаций IVOA для общего архива наблюдений отобраны следующие: SIA, SSA-протоколы, VOTable-формат, UCD. Следующим шагом в процессе интеграции локальных архивов является использование этих стандартов для разработки и реализации web-сервисов доступа к данным.

Благодарности. Работы поддержаны грантом РФФИ 07-03-90032.

Список литературы

- Витковский В.В., Желенкова О.П., Рядченко В.П., Шергин В.С., 1988, Сообщения САО, **59**, 60
- Витковский и др. (Vitkovskij V., Zhelenkova O., Kalinina N., Shergin V., Chernenkov V., Kiyashko T., Pischic L. and Kondratenko V.), 2000, *Baltic Astronomy*, **9**, 578
- Грейсен, Калабретта (Greisen E.W. and Calabretta M.R.), 2002, *A&A*, **395**, 1061
- Дерье и др. (Derriere S., Gray N., Mann R., Matrinez A.P., McDowell J., McGlynn T., Ochsenbein F., Osina P., Rixon G., Williams R.), UCD (Unified Content Descriptor) – moving to UCD1+, IVOA Working Draft 2004-04-26, <http://www.ivoa.net/Documents/UCD/WD-UCD-20040426.html>
- Доленски, Тоди (Dolensky M., Tody D.), 2004, *The Simple Spectrum Access Protocol*, SPIE, 5493-47
- Исуда и др. (Ysuda N., Mizumoto Y., Ohishi M., O'Mullane W., Budavari T., Haridas V., Li N., Malik T., Szalay A.), 2004, *Astronomical Data Query Language: Simple Query Protocol for the Virtual Observatory*, ADASS XIII, ASP Conf. Ser., v.XXX
- Кононов В.К., Моносов М.Л., Витковский В.В., Липовецкий В.А., 1990, *Сообщения САО*, **65**, 32
- Ошенбайн и др. (Ochsenbein F., Williams R., Davehall C., Durand D., Fernique P., Giaretta D., Hanish R., McGlynn T., Szalay A., Taylor M.B., Wicenec A.), VOTable Format Definition, ver. 1.094, IVOA Proposed Recommendation 2004-06-04, <http://cdsweb.strasbourg.fr/doc/VOTable/v1.09>
- Плейнт и др. (Plante R., Green G., Hanish, McGlynn T., O'Mullane W., Williams R., Williamson R.), 2004, *Resource Registries for the Virtual Observatory*, ADASS XIII, ASP Conf. Series, **30**
- Тоди, Плейнт (Tody D., Plante R.), 2004, *Simple Image Access Specification ver.1.0*, IVOA WG Working Draft, <http://www.ivoa.net/Documents/WD/SIA/sia-20040524.html>
- Уэлс и др. (Wells, D. C., Greisen, E. W., and Harten, R. H.), 1981, *A&AS*, **44**, 363
- Уильямс и др. (Williams R., Hanisch B., Linde T., McDowell J., Moore R., Ochsenbein F., Ohishi M., Rixon G., Szalay A., Tody D.), *Virtual Observatory Architecture Overview*, IVOA Note 2004-06-14, <http://www.ivoa.net/twiki/bin/view/IVOA/IvoaArchitecture>
- Фостер и др. (Foster I., Kesselman C., Nick J.M., Tuecke S.), 2002, *The Physiology of the Grid. An Open Grid Services Architecture for Distributed System Integration*, Draft 2.9, 31pp
- Четти, Бьюа (Chetty M., Buuya R.), 2002, *CiSE*, **4**, 61 <http://www.gridbus.org/papers/WeavingGrid.pdf>

Спецификации и протоколы

- Bulding the Framework for the National Virtual Observatory. NSF Cooperative Agreement AST0122449, Quarterly Report, April-June 2004, 44 pp.
- Extensible Markup Language (XML) 1.0 (Third Edition) W3C Recommendation 04 February 2004, <http://www.w3.org/TR/2004/REC-xml-20040204>
- IVOA SkyNode Interface, ver.0.7.4, IVOA Working Draft 2004-05-17, <http://www.ivoa.net/internal/IVOA/IvoaVOQL/SkyNodeInterface-0.7.4.pdf>
- The Open Archives Initiative Protocol for Metadata Harvesting . Protocol Version 2.0 of 2002-06-14, <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>