

# Электронный архив наблюдательных данных астрофизической обсерватории

О.П. Желенкова, В.В. Витковский, Т.А. Пляскина

## Аннотация

Архивная система обсерватории включает в себя хранилище цифровых данных и информационно-поисковую систему (ИПС) с динамическим веб-интерфейсом и http-доступом. На текущий момент в систему включены 16 цифровых коллекций наблюдательных данных (локальных архивов), полученных на разных приборах, работающих или работавших на телескопах обсерватории. Самые ранние данные относятся к концу 1994 года. В настоящее время активно пополняются 6 локальных архивов. Хранилище данных включает область временного хранения, расположенную на файл-сервере БТА, и область постоянного хранения. Область постоянного хранения это - CD/DVD-диски, жесткий диск выделенного архивного сервера и USB-диск большой емкости. Для сохранности данных при аварийных ситуациях и при возникающих дефектах носителей мы обеспечиваем хранение 2 полных копий на CD/DVD-дисках и двух копий архива на винчестере. Одна копия (A0) повторяет оптические диски, другая (A1), несколько по структуре каталогов модифицированная, используется ИПС. Цифровые носители и устройства чтения-записи данных нельзя отнести к устройствам длительного хранения, поэтому при долговременном хранении цифровых данных необходимо обеспечивать перенос информации на новый тип носителей каждые 5-10 лет. Копии архива A0 и A1 поддерживаются и для этой процедуры переписывания. Вариант архивных данных (A1) повторен на USB-диске с добавлением дампа таблиц и программ. Это резервная копия для восстановления информационной системы при аварийной ситуации на сервере. Для обеспечения модернизации ИПС мы поддерживаем две схемы базы данных — тестовую и рабочую. Разработки ведутся на тестовой схеме. При модификациях схемы после проведенного тестирования ИПС переключается на обновленную версию БД. Оригинальная копия A0 и наличие тестовой схемы позволяют выполнять модернизацию ИПС даже на уровне схемы таблиц. В настоящее время ИПС реализована на СУБД PostgreSQL 8.3.7.

Ключевые слова: цифровые коллекции экспериментальных данных, веб-доступ к архивам наблюдений, виртуальная обсерватория, предметно-ориентированные базы данных

## Архивная система обсерватории

Работа по стандартизации и хранению цифровых наблюдательных данных в САО РАН ведется с конца 80-тых. FITS-формат является стандартом для передачи и хранения данных в астрономии [1]. В обсерватории он впервые был применен для данных ПЗС-камеры [2]. Затем FITS-формат стал использоваться для описания и хранения наблюдательных данных, получаемых на других приборах.

На телескопах обсерватории работают разные наблюдательные приборы, каждый из которых управляется компьютерно-аппаратным комплексом (системой сбора). Локальный архив - это цифровая коллекция данных, получаемых одной системой сбора. Общий архив состоит из локальных архивов, которые отличаются структурой данных и набором параметров, описывающих наблюдения.

В конце 90-х отдел информатики приступил к созданию архивной системы для накопления, постоянного хранения и веб-доступа к цифровым наблюдательным данным. Для архивной системы была предложена архитектура, состоящая из трех взаимосвязанных уровней:

- накопление - каскадная схема архивизации, где поток данных от любой системы сбора направляется на общий файл-сервер, а затем производится запись на носители для постоянного хранения;
- хранение и обеспечение сохранности — технические средства для постоянного хранения наблюдений;
- поиск и доступ - информационно-поисковая система (ИПС) на основе системы управления реляционными базами данных с пользовательским веб-интерфейсом.

При разработке архивной системы обсерватории в основу были положены следующими правилами:

- в архиве не меняются форматы хранимых данных; в каком формате они поступили на вход архива, в таком они выдаются при запросе;
- в архиве хранятся необработанные данные, записанные на оптических дисках;
- логической единицей хранения в архиве является наблюдение;
- данные имеют двухлетний период исключительного авторского права заявителей наблюдательной программы на основании Положения об архиве наблюдательных данных САО РАН, затем они открываются для свободного копирования.

С 1994 года цифровые данные, получаемые на инструментах обсерватории, хранятся на оптических дисках. На рисунке 1 представлены объем и темпы прироста цифровых наблюдательных данных, накапливаемых в архивной системе.

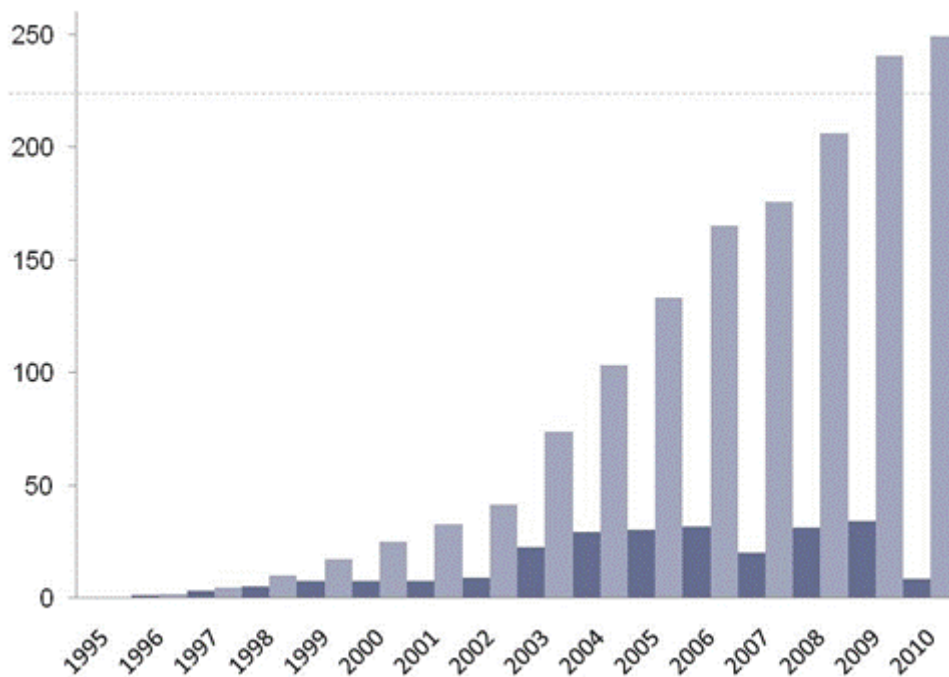


Рис. 1. Объем и темпы прироста (GB) данных общего архива наблюдений обсерватории.

## Накопление и хранение архивных данных

Схема накопления данных в архивной системе определяется информационными потоками, получаемыми на инструментах обсерватории, и структурой постоянного хранения. На оптических телескопах САО РАН в настоящее время используются 8 приборов, один из которых, SCORPIO [3], имеет 5 режимов наблюдений. Для каждого из режимов ведется отдельный локальный архив. Два метода наблюдений «спекл» и «МАНИЯ КЧД» имеют собственные архивные системы. Из-за большого объема данных, получаемых в течение ночи, и особенностей обработки наблюдений локальные архивы этих инструментов не поддерживаются общим архивом обсерватории. В ИПС архива включены 16 локальных архивов. Шесть из них активно пополняются, а оставшиеся находятся на постоянном хранении.

Информация, накопленная на компьютере системы сбора в течение сета, по его завершении копируется на выделенный файл-сервер 6-м телескопа в соответствующий методу наблюдений каталог. Затем из сетов наблюдений, полученных на одном приборе, формируется образ архивного диска, который по мере заполнения записывается на CD/DVD-носитель. Из этого сложились простые правила, которые определяют внутреннюю структуру архивного CD/DVD-диска:

- метка тома (файл нулевой длины);
- каталоги с наблюдениями (иногда имеются вложенные каталоги);
- в одном каталоге записываются данные одной наблюдательной ночи;
- название каталога должно содержать дату наблюдения;
- семантическая единица архива - один файл с наблюдательными данными.

В этих правилах нет ограничений на форматы файлов, что дает возможность добавлять данные новых приборов в архив и информационно-поисковую систему. Тем самым мы обеспечиваем масштабируемость (т.е. возможность добавления новых функций или данных без изменения структуры самой программной системы) архивной системы в смысле форматов файлов. Отметим, что можно добавлять в архивную систему наблюдения, полученные в других обсерваториях.

Вышеперечисленные правила определяют алгоритм верификации CD-диска перед помещением его в область хранения и в ИПС, в котором проверяется:

- сколько локальных архивов записано на диске;
- хранятся ли данные одной ночи в одном каталоге;
- имя каталога с данными одной ночи должно содержать дату наблюдений;
- соблюдается ли в каталоге правило записи наблюдения в один файл.

Помещение диска в информационно-поисковую систему начинается с копирования в буфер, где проверяется, соответствует ли его содержимое принятой структуре. При записи диска в хранилище выполняются следующие преобразования:

- если на диске записано несколько локальных архивов, то обработка диска проводится в несколько проходов - столько, сколько записано локальных архивов на диске. Символьная ссылка на один и тот же диск устанавливается тогда в нескольких соответствующих каталогах логической структуры области хранения;
- данные одной ночи переписываются в один каталог, если на копируемом диске это не соблюдается;
- если имя каталога не содержит дату наблюдений, то он переименовывается так, чтобы дата присутствовала в названии;
- наблюдения извлекаются из tag-архивов, если таковые имеются;
- файлы заново сжимаются, если метод компрессии отличается от bzip2;
- выполняется преобразование файлов во внутреннем формате системы MIDAS (файлы с расширением bdf) в FITS-формат;
- производится анализ FITS-заголовков, если в заголовках указан другой инструмент, чем тот, к которому относится диск, то принадлежность файла локальному архиву изменяется.

Все эти проверки выполняются программами-фильтрами перед созданием текстовых файлов, отражающих структуру диска. Эти текстовые

файла получают при выполнении команды системной команды ls. Они являются частью архивной системы, сохраняются и используются при наполнении таблиц базы данных. При верификации сообщения об ошибках и действиях с дисками сохраняются в протокольных файлах. Дополнительные операции с каждым диском фиксируются в программе на языке bash, который используется при полном или частичном восстановлении хранилища и ИПС.

Хранилище данных включает область временного хранения, расположенную на файл-сервере БТА, и область постоянного хранения. Область постоянного хранения это - CD/DVD-диски, жесткий диск выделенного архивного сервера и флэш-диск большой емкости. Для сохранности данных при аварийных ситуациях и при возникающих дефектах носителей мы обеспечиваем хранение двух полных копий на CD/DVD-дисках и двух копий архива на винчестере. Одна копия (A0) на жестком диске повторяет оптические диски, другая (A1), несколько по структуре каталогов модифицированная, собственно используется ИПС. Вариант архивных данных (A1) повторен на USB-диске с добавлением дампа таблиц и программ - это резервная копия для восстановления информационной системы при аварийной ситуации на сервере.

Область хранения наблюдательных данных на жестком диске архивного сервера имеет физическую и логическую организацию. Каждый диск помещается в каталог с названием диска, соответствующим его архивному номеру - CDxxx. Ниже уровнем находятся каталоги, именованные по дате наблюдений и содержащие наблюдения на соответствующую дату. На физическую структуру архива накладывается логическая, в которой отображается распределение дисков по локальным архивам (методам наблюдений).

Астрономические явления часто носят переменный характер на разных временных интервалах, совершенствуются и меняются методы обработки и анализа данных и возникает потребность провести повторную обработку наблюдений, поэтому долговременное хранение наблюдений входит в компетенцию обсерваторий.

Появление цифровых приемников излучения обеспечило астрономов большим количеством данных, а цифровые записывающие устройства - компактными средствами хранения. Однако информацию уже невозможно рассматривать человеческим глазом, как это можно было делать с фотографическими пластинками.

При считывании, записи и хранении данных могут возникнуть ошибки, потеря информации. Конечно, надежность устройств растет, а ошибки, возникающие при чтении и дублировании носителей информации, можно контролировать программно. Но компьютерное оборудование меняется с такой скоростью, что время физического разрушения носители информации оказывается больше, чем время жизни устройства считывания.

В настоящее время все цифровые носители и устройства чтения-записи данных нельзя отнести к устройствам длительного хранения. Это требует постоянного отслеживания состояния систем хранения и переписывания данных на новые носители. Поэтому для архивов астрономических данных необходимо предусматривать такую технологию хранения, которая позволяет раз в несколько лет (от 5 до 10) переносить содержимое с устаревших носителей на более новые.

Переписывание архивных данных производилось у нас частично в конце 90-х, когда данные со стримерных лент, EXABYTE и DAT-кассет были скопированы на оптические диски. Копии архива A0 и A1 поддерживаются теперь у нас и для этой процедуры переписывания.

Всего в архиве 340509 файлов, наблюдательные данные, включая и радиодиапазон, составляют 92%, журналы наблюдений — 0.5%, вспомогательные — 5%, без категории — 1%. Из 264706 наблюдательных файлов, полученных на оптических телескопах, 96% записаны в FITS-формате. Из них:

- файлы с наблюдениями объектов (OBJ) составляют - 57%,
- байесы (BS) - 19%,
- темновые кадры (DK) - 2%,
- плоские поля (FF) - 11%,
- стандартные лампы (ST) - 10%,
- а также те файлы, для которых тип данных не определяется программным алгоритмом (undf) - 1%.

Разделение по типу для файлов с наблюдениями в архиве выглядит так:

- прямые снимки - 42%,
- эшелле-спектры - 9%,
- спектры, полученные с длинной щелью - 24%,
- наблюдения интерферометром Фабри-Перо - 12%,
- мультиобъектная спектроскопия — 7%,
- поляриметрия — 0.2% и др.

## Поиск и доступ к наблюдательным данным

Прототип информационно-поисковой системы для общего архива обсерватории разрабатывался отделом информатики обсерватории совместно с коллегами из ЮгИнфо РГУ [4]. Система имела двухуровневую организацию, включающую ИПС и хранилище файлов. Информационно-поисковая система базировалась на системе управления базами данных Oracle. Сервер базы данных располагался в Ростове. Запросы и копирование данных выполнялись с помощью веб-интерфейса. Тестирование ИПС проводилось по данным, полученным на РАТАН-600. В ИПС для каждого файла из параметров наблюдений формировался сервисный информационный блок, который включал также информацию для идентификации файла. В сервисный информационный блок, как оказалось, мы заложили избыточное количество параметров. Из-за этого возникли сложности с наполнением таблиц ИПС для других локальных архивов, поскольку значительная часть параметров отсутствовала в описательной части файлов.

Идеи, заложенные в этой разработке, послужили основой для создания существующей архивной системы. Правда, мы отказались от занесения в таблицы ИПС всех параметров, которые имеются в заголовках файлов. Чтобы обеспечить доступ ко всем локальным архивам, потребовалось выделить набор параметров наблюдательных файлов, которые присутствовали бы во всех локальных архивах, а также были устойчивы к пропускам и ошибкам со стороны наблюдателей. Самым надежным параметром оказалась дата наблюдений.

Первая версия ИПС обеспечивала запросы по дате наблюдений и выбранному локальному архиву наблюдений, копирование данных по http-протоколу, просмотр заголовков файлов и содержимого файлов. Для архива наблюдений использовался специализированный сервер,

на котором размещалось хранилище файлов. База данных поддерживалась СУБД Oracle 9i (trail-версия). Веб-интерфейс был реализован на основе спецификаций CGI-интерфейса (Common Gateway Interface), DBD (DataBase Driver) и DBI (DataBase Interface) интерфейсов к СУБД.

Вторая версия ИПС [5] разрабатывалась для обеспечения наиболее часто используемых типов запросов для выборки архивных данных, а именно, - по дате наблюдения, прибору, типам файлов, координатам наблюдаемого поля/объекта, имени астрономического объекта, программе наблюдений, заявителю программы и наблюдателям, принимавшим участие в наблюдениях.

#### Web-интерфейс общего архива наблюдений

Вид пользовательского web-интерфейса общего архива наблюдений обсерватории представлен на рисунке 2. Имеется англоязычный вариант интерфейса. Интерфейс администратора архива отличается выводом дополнительной информации для локализации файлов в области хранения.

Веб-интерфейс генерируется скриптовой программой на Perl и динамически отображает диапазон дат помещенных в архив наблюдений и включенные в поисковую систему локальные архивы. Результаты запроса для отмеченного локального архива выводятся в новое окно браузера.

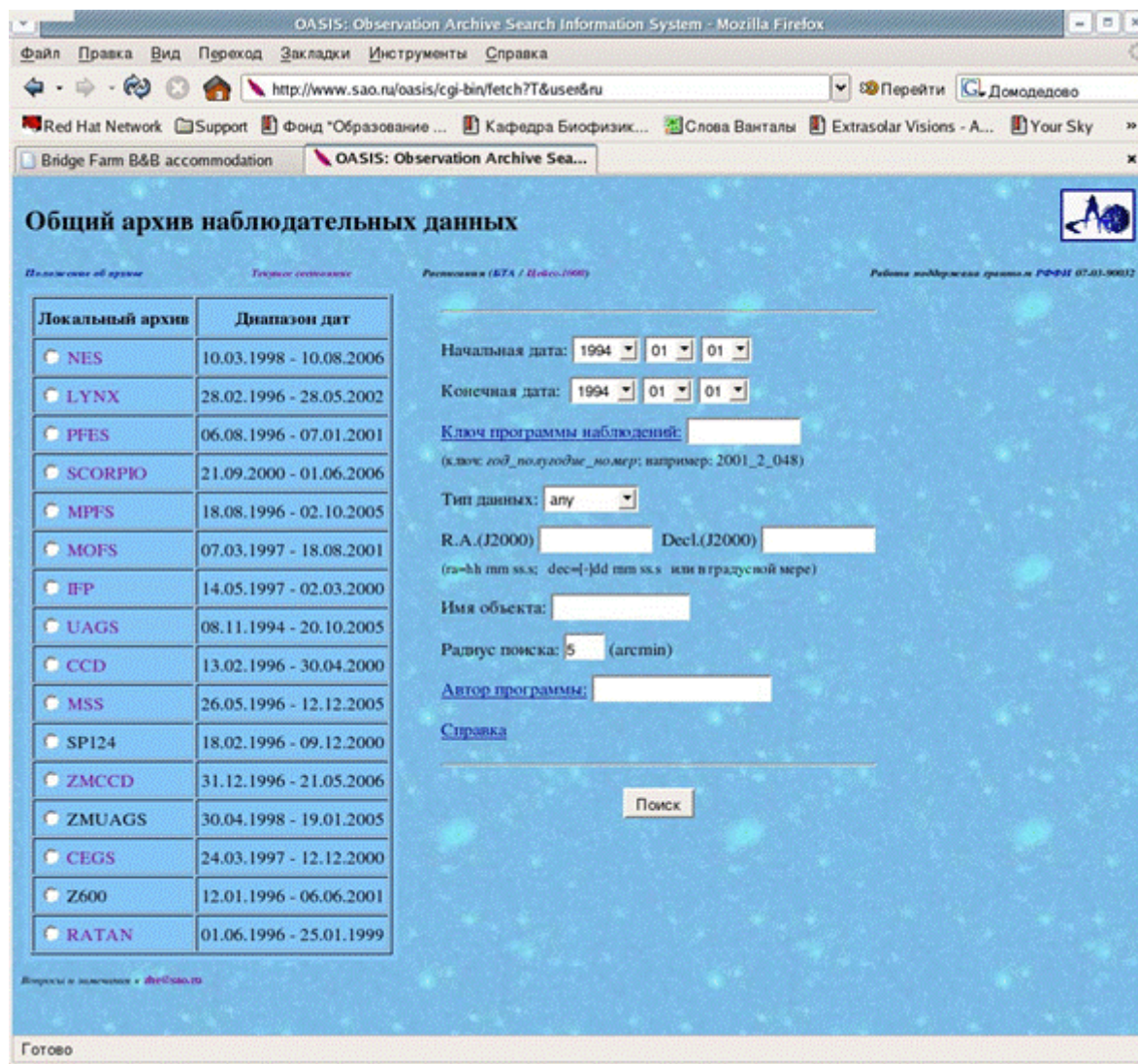


Рис. 2. Веб-интерфейс общего архива наблюдательных данных САО РАН.

В интерфейсе реализованы следующие типы запросов к архивным данным:

- по интервалу дат. Если в запросе указаны только даты, то выбираются наблюдения, а также журнальные и вспомогательные файлы, попавшие в диапазон дат.
- по координатам. При задании координат поиск FITS-файлов производится в квадрате со стороной, равной удвоенному радиусу поиска. Центр области задается введенными координатами. В результате запроса выдаются файлы, у которых координаты, заданные в параметрах FITS-заголовка и приведенные на 2000.0, попадают в указанную область. Если задается только координата прямого восхождения, то поиск файлов выполняется в полосе склонений  $[-90^\circ - +90^\circ]$ , а если же вводится только склонение, то поиск файлов выполняется в полосе прямых восхождений  $[0^\circ - 360^\circ]$ .
- по имени объекта. При запросе по имени объекта его координаты извлекаются с помощью веб-сервиса Sesame (`http://cdswww.strasbourg.fr/axis/services/Sesame`) из базы данных Simbad, при этом, если имя не обнаруживается, то выполняется поиск в Vizier, а затем в NED. Следует отметить, что если заданы координаты и имя объекта, то поиск в архиве будет выполняться по координатам, имя объекта в этом случае игнорируется.
- по фамилии наблюдателя/заявителя программы. При поиске наблюдений по фамилии наблюдателя или заявителя программы допускается неполный ввод фамилии. Если при этом возникает неоднозначность, то поиск в архиве осуществляется по первой фамилии, совпавшей с введенным шаблоном. Например, введено «mon», этому шаблону соответствуют: «Monin», «Montmerle».
- по программе наблюдений. Поиск наблюдений по программе наблюдений, производится по ключу программы, состоящему из года,

полугодия и порядкового номера программы в полугодии. Для справки в отдельное окно браузера выводится список программ с соответствующими ключами.

#### *Анализ параметров, используемых в стандартных запросах*

Для описания наблюдений на приборах БТА используются разные наборы параметров. Эти наборы имеют общую часть, которая включает информацию об объекте, программе и т.п., а также технические характеристики, присущие конкретному прибору. В описание могут входить до сотни параметров (на радиотелескопе РАТАН-600 - до 300). Значения этих параметров формируются в системах управления телескопом, инструментом, а также в системе сбора данных. Часть параметров поступает в заголовок файла автоматически, часть заносится наблюдателем.

Сложность заполнения таблиц ИПС необходимыми для стандартных запросов параметрами состоит в том, что:

- при модернизации систем сбора и приборов меняются форматы данных и, как правило, в локальном архиве имеется несколько версий формата, которые отличаются по набору ключевых слов, а также формату записи величин;
- разные системы сбора формируют заголовок файла с отличающимися по названию ключевыми словами, но обозначающими 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».

По этим причинам программный фильтр для синтаксического разбора, анализа и извлечения значений параметров из заголовков файлов реализован с использованием дополнительной таблицы в схеме базы данных, которая связывает названия ключевых слов и атрибуты ИПС. Эта таблица-словарь содержит все ключевые слова локальных архивов, обнаруженные при разборе FITS-заголовков файлов, их семантические значения, определяющие физический смысл величины, UCD [7], а также связь величины и параметра ИПС.

Если параметры, по которым производится обращение к поисковой системе, отсутствуют в заголовке файла, то он не попадает в результат запроса. Такая ситуация не является редкой для архивных данных, особенно для ранних по датам файлов архива.

Пропущенную, потерянную при перезаписях информацию, а также ошибки, вносимые в данные наблюдателем, можно в ряде случаев дополнить или исправить, используя другие источники, описывающие наблюдения. Мы старались в таких случаях, где возможно, использовать дополнительную информацию, которая имеется в имени файла. По нему определялась дата наблюдения, метод компрессии, тип изображения (темновой кадр, байес, плоское поле, объект), тип фильтра при широкополосной фотометрии, формат записи данных и т.п. Значения параметров ИПС, извлеченные из имени файла, имеют самый высокий приоритет по сравнению с информацией из заголовка файла. Когда имеются две величины для параметра, в таблицу заносится значение, полученное из имени файла.

Дополнительная информация о наблюдениях имеется также в расписаниях и журнальных файлах. Она пока не используется при добавлениях в таблицы величин параметров, которые отсутствуют в заголовках наблюдательных файлах.

Ко всем архивным файлам можно обратиться по дате наблюдения. Дата наблюдений в архивных файлах оказалась наиболее устойчивой к пропускам и ошибкам со стороны наблюдателей. Мы извлекаем ее программно из названия файла с наблюдениями ночи. Дата наблюдений из заголовков файлов является менее надежной. В них отмечались пропуски в значении ключевого слова, определяющего дату, а при переходе на новый формат записи даты в течение 2000 года некоторые системы сбора записывали ошибочные значения.

По расширению имени файла мы разделяем файлы на следующие типы - наблюдательные, журнальные, вспомогательные и неопределенные или без категории (по содержанию не можем отнести к наблюдательной информации). Отнесение файла к одной из этих категорий производится с помощью справочной таблицы, в которой собраны все встречающиеся в архиве расширения имен файлов. Так на текущий момент расширения .bdf, .mt, .fts, .fits, .tar приписаны наблюдательным файлам, .tbl, .log, .plog, .base, .pro, .dbf, .lst имеют журнальные файлы, а остальное - вспомогательные и без категории. Встречаются журнальные файлы с расширением .mt. Для этого дополнительно анализируются заголовки файлов, чтобы разделить таблицы и изображения. Табличные файлы помечаются затем в ИПС, как журнальные. Отметим, что журнальные, вспомогательные файлы и файлы без категории можно извлечь из архива только по дате наблюдения.

Наблюдательные файлы мы разделяем на научные (с наблюдениями изучаемых небесных объектов) и сервисные (данные, используемые при редукции - байесы, темновые кадры, плоские поля, стандарты). Разделение наблюдательных данных на типы выполняется при анализе имен файлов и ключевых слов в заголовках. При определении типа изображения производится проверка на наличие в значении ключевого слова сочетаний символов, которые совпадают с сочетаниями, извлеченными из справочной таблицы. Если такое совпадение имеется, то файлу устанавливается признак соответствующего типа наблюдательных данных. Такой же сравнение выполняется и для имени файла. Если имеется информация о типе данных, как в имени файла, так и в ключевых словах, то при программном анализе приоритет отдается имени файла.

Приборы, используемые на телескопе, формируют разные по структуре данные (например, прямые снимки, спектры эшелле, спектры с длинной щелью и др.). Сравнивая значения соответствующих ключевых слов со справочной таблицей, в которой устанавливается связь прибора с типом наблюдений, мы разделяем научные данные по структуре. Для приборов с несколькими режимами (например, SCORPIO) тип наблюдения уточняется по ключевому слову и дополнительной таблице, содержащей все встречаемые значения ключевого слова и связь их с типом наблюдений. Это важно для реализации запроса по координатам.

Запрос по координатам выполняется к файлам с научными данными. В запросе участвуют координаты, извлеченные из заголовка файла и приведенные к стандартной эпохе 2000.0. Для спектров - это координаты объекта, а для прямых снимков - центра области. В последнем случае при выборке данных надо учитывать размер поля, который вычисляется по количеству и угловому размеру пикселей ПЗС-матрицы. Эти значения извлекаются из соответствующих ключевых слов заголовка, но часто для прямых снимков угловой размер пикселя равен некоторому фиктивному значению - единице. По этой причине мы решили пока не учитывать для прямых снимков размер области и ограничиться координатами центра.

Для получения координат по имени объекта используются веб-сервисы CDS и NED. Если значения координат в ключевых словах заголовка файла отсутствуют, то можно попытаться получить координаты по имени объекта.

Имена наблюдателей и авторов программ извлекаются из ключевых слов OBSERVER и AUTHOR. Как правило, эта информация заносится в заголовок наблюдателями, при этом часто используются сокращения фамилий. Чтобы разобраться с многочисленными синонимами, нам

пришлось извлечь из FITS-заголовков наблюдательных файлов список всех возможных вариантов записей фамилий и сокращенных обозначений. В файл попало 838 уникальных сочетаний. Некоторые наблюдатели и заявители программ имеют до десятка вариантов записи фамилии. Каждой персоне, оказавшейся в списке, присвоен номер. Всего оказалось 206 персон, из которых только заявителями программ является 154, наблюдателями - 133, авторы и наблюдатели - 81. Список сочетаний и список персон послужили основой для справочной таблицы. Выборка данных по автору программы или наблюдателю выполняется посредством этой таблицы, то есть, по совпадению фамилии определяется идентификатор, а затем уже идет поиск с ним по наблюдательным данным.

Сложность организации запроса по названию наблюдательной программы состоит в том, что названия в заголовки файлов заносятся наблюдателями в произвольной манере, так, например, это может быть смысловое сочетание (Исследование GRB), общепринятое сокращение (GRB) или неполное название, но не то, которое фигурирует в расписании наблюдений. В таких случаях одной наблюдательной программе в архиве может соответствовать несколько названий, которые, однако, никак не совпадают с названием в расписании наблюдений. В списке программ (около 920 названий), которые были извлечены из наблюдательных файлов (ключевое слово PROGRAM в заголовке), не оказалось ни одного совпадения с названиями программ из архива расписаний наблюдений на БТА.

Общепринятой практикой в других обсерваториях является присвоение программе наблюдений уникального идентификатора. Из архива расписаний составлена таблица (2515 названий). Каждой программе приписан символьный ключ, включающий год, полугодие и порядковый номер программы в полугодии. Выборка данных, относящихся к сету наблюдений, производится по ключу программы.

Нужно учитывать, что стандартные запросы, кроме выборки данных по дате наблюдения, реализуются к части наблюдательных файлов из-за отсутствия параметров запросов в заголовках файлов. Приведем статистику наличия параметров в файлах, которые используются в запросах. Отметим, что в ней не учитываются ошибочные значения параметров, а только пропущенные:

- 17% - нет значений координат в ключевых словах заголовка,
- 4% - нет имени объекта,
- 30% - не определено название программы,
- 29% - не указан заявитель программы,
- 29% - не указаны наблюдатели, участвовавшие в наблюдениях.

#### *Схема таблиц базы данных архива*

Существующая третья версия ИПС архива работает под управлением СУБД PostgreSQL. Эта система является свободно распространяемым некоммерческим программным продуктом, поддерживает стандарт ISO и ANSI языка запросов SQL-92, обеспечивает транзакции, поддерживает представления и сложные структуры данных, создаваемые пользователями, то есть ничем не уступает коммерческим системам [9]. Для нее разработаны встроенные пакеты процедур, обеспечивающие работы с астрономическими координатами, поддерживающие пикселизацию неба. По этим причинам решено поисковую систему архива перевести в СУБД PostgreSQL.

Исходя из накопленного опыта эксплуатации поисковой системы, мы внесли изменения в схему ИПС. Она была дополнена представлениями и ограничениями целостности для обеспечения непротиворечивости данных.

Приведем описание схемы таблиц базы данных поисковой системы (рис. 3). По способу занесения информации таблицы условно можно разбить на три группы по частоте обновления и способу занесения в них новых записей:

- К первой группе относятся справочные таблицы, информация в которых не меняется редко. Они заполняются при создании схемы базы данных. Новые записи могут появиться только, когда в информационно-поисковую систему добавляется новый локальный архив. Такие добавления происходят не часто, поэтому эти таблицы можно считать статическими.
- Ко второй группе относятся справочные таблицы, которые могут пополняться новыми записями при анализе впервые вносимого в поисковую систему CD/DVD-диска, например, новая программа наблюдений, наблюдатель, тип файла и т.п. Добавление записи происходит программно. Информация, которая заносится в справочные таблицы, получена в результате визуального и программного анализа архивных дисков и заголовков файлов.
- Третья группа - это таблицы с данными о каждом архивированном файле. Записи добавляются в них при внесении нового CD/DVD-диска, и при заполнении полей этих записей используются таблицы первых двух групп.
- Особое место занимает таблица, связывающая атрибуты таблиц, хранящих информацию о наблюдательных файлах, с ключевыми словами FITS-заголовков и UCD.
- В схеме ИПС есть представления, которые формируются динамически средствами СУБД.

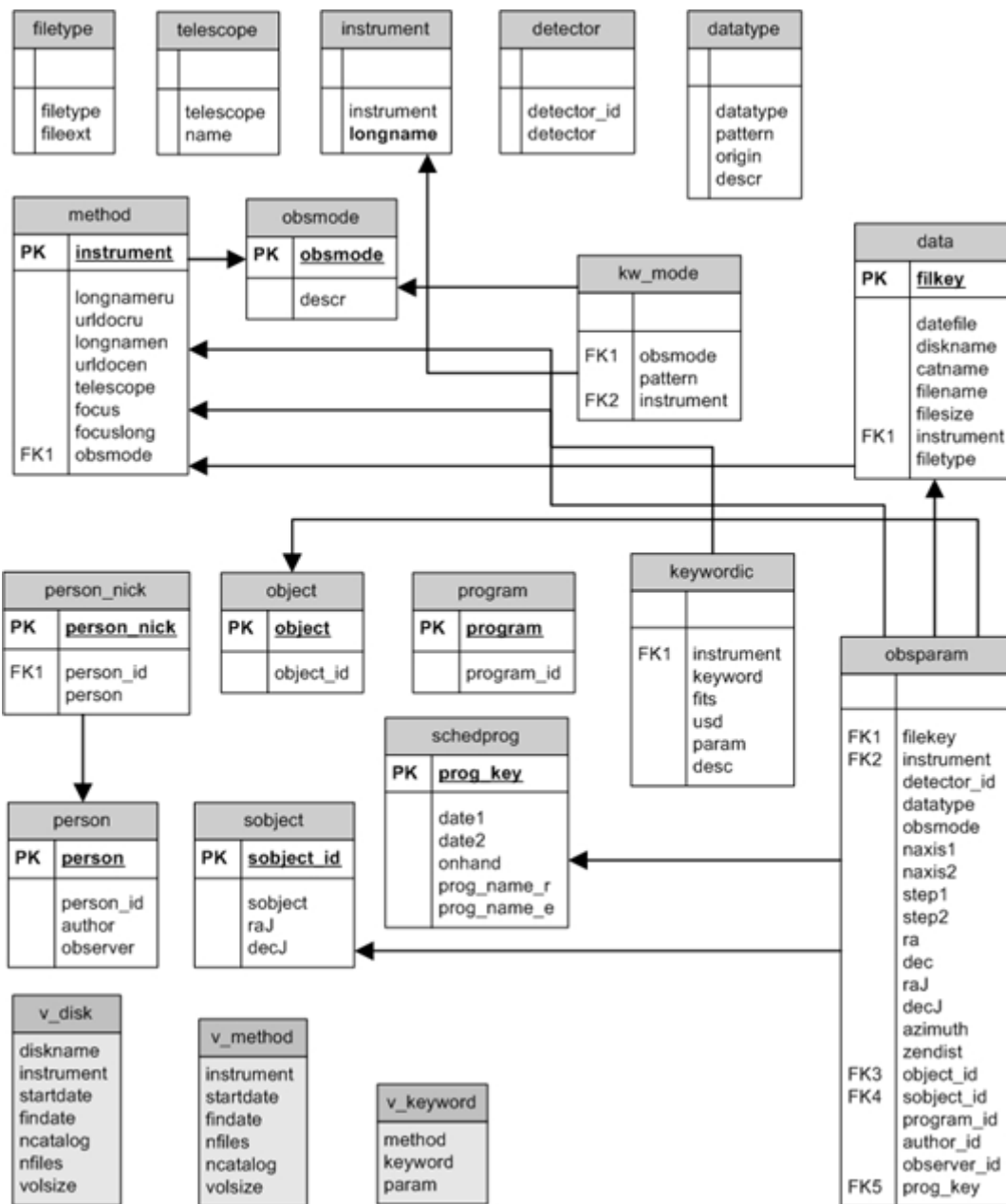


Рис. 3. Схема таблиц информационно-поисковой системы общего архива наблюдательных данных САО РАН.

## Заключение

В 2003 году Генеральная Ассамблея Международного астрономического союза приняла резолюцию [8] об открытом Интернет-доступе к архивированным наблюдениям, полученным в обсерваториях, которые финансируются из государственного бюджета. Методы работы с данными в инфраструктуре виртуальной обсерватории определяются стандартами Международного альянса

«Виртуальная обсерватория» (International Alliance Virtual Observatory, IVOA). Реализация как рекомендации МАС, так и стандартов IVOA применительно к любому цифровому архиву включает существенный объем работ по организации информационного ресурса и стандартизации данных.

В обсерватории поддерживаются, кроме ИПС архива, еще программные средства с веб-интерфейсами — система для on-line заявки на наблюдательное время и подготовки расписания наблюдений [9], подготовка журналов наблюдений. В настоящее время проводится комплекс работ по интеграции этих систем с ИПС архива, а также по развитию функций архива САО РАН не только в соответствии с требованиями резолюции МАС, но и альянса IVOA. Чтобы обеспечить такого рода разработки, для ИПС архива поддерживается две схемы - основная и тестовая.

## Литература

- [1] Wells, D. C., et al., A&AS, **44**, 363 (1981)
- [2] Витковский и др., Сообщения САО, **59**, 60 (1988)
- [3] Afanasiev V.L. & Moiseev A.V., PaZh, **31**, 214 (2005)
- [4] Vitkovskij V., et al., Baltic Astronomy, **9**, 578 (2000)
- [5] Витковский В.В. и др., Состояние и перспективы развития архива наблюдений обсерватории, Известия САО, **58**, 52 (2005)

[6] Derriere S., et al., <http://www.ivoa.net/Documents/UCD/WD-UCD-20040426.html> (2004)

[7] Бартунов О., <http://www.sai.msu.su/~megeera/postgres/talks/what-is-postgresql.html> (2005)

[8] Public Access to Astronomical Archives. The Resolution of 5 Commission of IAU <http://www.atnf.csiro.au/people/rnorris/WGAD/Resolution.htm> (2003)

[9] Витковский В.В., Желенкова О.П., Малхасян С.В., Комплексная web-система подачи заявки на наблюдательное время и работы с расписанием наблюдений на телескопах САО РАН. Научно-технический отчет САО РАН N293, 23с (2003)

#### **Об авторах**

О.П. Желенкова - Специальная астрофизическая обсерватория РАН

[zhe@sao.ru](mailto:zhe@sao.ru)

В.В. Витковский - Специальная астрофизическая обсерватория РАН

[vvv@sao.ru](mailto:vvv@sao.ru)

Т.А. Пляскина, Специальная астрофизическая обсерватория РАН

[taap@sao.ru](mailto:taap@sao.ru)

Работа выполнена при частичной поддержке гранта РФФИ № 10-07-00412.