

К проблеме организации разнородных данных при многочастотных исследованиях радиоисточников

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

Аннотация

Программные средства виртуальной обсерватории обеспечивают удобный доступ к астрономическим данным и повышают эффективность научных исследований. На базе уже существующих веб-технологий ведутся разработки по реализации большей семантической связности данных, что должно перевести сервисы виртуальной обсерватории на новый уровень, обеспечивающий обмен не только данными, но и знаниями. Пока еще отсутствует развитый инструментарий для актуализации собранной из веб-ресурсов информации, а также средства организации разнородных данных, подготовленных пользователем для научного анализа, которые могли бы повысить эффективность дальнейших исследований. Рассматривается подход к организации разнородной информации, относящейся к исследуемому объекту, который основан на развитии существующих форматов астрономических данных и средств поддержки семантической связности данных.

Ключевые слова: виртуальная обсерватория, астрономические данные, стандарт хранения, формат обмена данными.

Введение

Виртуальная обсерватория является средством удобного и эффективного доступа к цифровым астрономическим данным. Первые шаги к организации этой распределенной инфраструктуры были сделаны в 90-х годах прошлого века, когда в США была создана сеть центров данных по поддержке информации, полученной космическими миссиями НАСА. Достижения в области информационных технологий обеспечили основу, на которой распределенные коллекции данных стали рассматриваться как интегрированная информационная система. Виртуальная обсерватория открыла новые направления научных исследований, опирающиеся на статистический анализ, поиск новых закономерностей и объединение данных разных диапазонов.

Обычной практикой в астрономии были отдельные и/или повторные наблюдения индивидуальных объектов. Но по мере того как понимание разных астрономических феноменов и закономерностей становилось более точным, число вопросов, на которые можно ответить с помощью одного наблюдения, становилось все меньшим. Отметим, что наблюдательное время больших телескопов остается и будет оставаться весьма ограниченным, поэтому многие

астрофизические вопросы, для решения которых требуется большое количество наблюдений, часто просто не могут рассматриваться.

Методы наблюдений в настоящее время все больше смещаются в сторону накопления больших объемов данных в разных диапазонах электромагнитного спектра, а подход к решению астрофизических задач – к статистическим методам. Такие исследования предъявляют определенные требования как к постановке задачи, так и к способам их решения, которые все больше опираются на информационные технологии и, в первую очередь, на веб-сервисы, системы управления базами данных, грид и облачные вычисления.

Для эффективной работы с разнородными данными требуются также решение проблем организации информации, которая позволила бы отслеживать изменения и новые релизы обзоров, а также поддерживать смысловую связь между объектами разных каталогов. Мы столкнулись с этими проблемами при многочастотных исследованиях радиоисточников с применением цифровых обзоров и каталогов радио, инфракрасного и оптического диапазонов.

2. Многочастотные исследования радиоисточников

Исследования радиогалактик относятся к широкому кругу задач, связанному с изучением проявлений активности галактических ядер. Феномен мощного радиоизлучения ядра галактики является кратковременной по космологическим масштабам (до $\sim 10^8$ лет) эволюционной фазой самых массивных звездных систем. Хотя мощные радиогалактики – редкие объекты, их наблюдения в радиодиапазоне возможно практически на любых космологических расстояниях, что используется при изучении крупномасштабной структуры Вселенной, проверке гипотез формирования самых первых звездных систем и их эволюции.

Отождествление радиоисточников с объектами в других диапазонах – обязательная процедура при подобных исследованиях, и не такая простая, как это кажется на первый взгляд. Кросс-идентификация по координатам (около объекта из заданного списка в области с заданным радиусом ищется объект сравнимого каталога) оптических и радиокаталогов при разном угловом разрешении, предельной чувствительности и координатной точности последних, а также морфологической структуре самих источников дает небольшой процент совпадений (5-30%). Только по спектру можно оценить расстояние до родительских галактик радиоисточников, которые, по большей части, весьма слабые оптические объекты, особенно на больших красных смещениях. Из-за этого наблюдения требуют больших затрат времени на крупнейших оптических телескопах, причем при наилучших погодных условиях.

Не только поиск далеких радиогалактик, но и статистические свойства радиоисточников в разных диапазонах электромагнитного спектра важны для понимания природы активных галактических ядер. Массовое исследование радиоисточников позволяет уточнять существующие и открывать новые селекционные критерии, которые можно использовать при классификации этих объектов. В настоящее время подобные исследования проводятся с привлечением современных цифровых обзоров в разных диапазонах.

2.1. Исследования радиоисточников обзоров, проведенных на радиотелескопе РАТАН-600

На крупнейшем российском радиотелескопе РАТАН 600 была проведена серия глубоких обзоров полосы неба шириной около 40 угловых минут. По данным этих обзоров получен каталог RC (RATAN Cold) [1], а затем RCR (RATAN Cold Revised) [2].

Были проведены исследования выборки источников с крутыми спектрами (SS, Steep Spectra) каталога RC [3]. Для кандидатов в выборку учитывались угловые размеры, морфологическая структура, а также яркость объекта в радиодиапазоне. Так из ~1000 источников каталога RC в выборку с крутыми спектрами вошли ~100 объектов. Потребовалось 15 лет фотометрических и спектральных наблюдений на 6 м оптическом телескопе БТА (Большой телескоп азимутальный), чтобы отождествить и получить спектры для объектов выборки [4]. Из этих объектов у четырех источников оказалось 1 ? Z

С появлением глубоких цифровых обзоров в оптическом и инфракрасном диапазоне, таких как SDSS (Sloan Digital Sky Survey) [6] и UKIDSS (United Kingdom Infra-red Deep Sky Survey) [7], а затем WISE (Wide-Field Infrared Survey Explorer) [8] появилась возможность провести отождествление каталогов RC и RCR. Для выполнения этой задачи были максимально использованы все имеющиеся в открытом доступе данные. Радиоисточники каталога RC [9,10], а затем RCR [11,12] были отождествлены с данными этих обзоров. Радиоисточники, у которых не обнаруживались кандидаты в оптических и/или инфракрасных каталогах, дополнительно отождествлялись с суммарными изображениями обзора SDSS в трех фильтрах (g, r, i) и/или инфракрасного обзора UKIDSS в фильтрах (J, H, K) для достижения более глубокого предела кадров.

При исследованиях источников каталога RC, а затем каталога RCR разработана методика детального отождествления радиоисточников, включающая подбор информационных ресурсов, автоматическую подготовку данных из выбранных ресурсов для каждого источника, морфологическую классификацию, визуальную инспекцию подготовленных данных для принятия решения об отождествлении. Средствами интерактивного атласа неба Aladin [13] (программного интерфейса Perl для командного интерфейса и макроконтроллера), а также с помощью программного интерфейса Python к SAOImage DS9 [14] реализованы потоки работ по списку радиоисточников для подготовки данных и визуализации результатов. Для полученного компилятивного каталога была разработана информационно-поисковая система, которая использовалась при отождествлении радиоисточников [15].

2.2. К эффективности работы с разнородными данными

В результате отождествления радиоисточников каталога RCR у нас получен компилятивный каталог, который предполагается использовать для дальнейших исследований, а именно: определение физических характеристик родительских галактик радиоисточников, их классификация, подготовка выборок источников со схожими свойствами, поиск далеких объектов, поиск переменности в оптическом и радиодиапазоне и пр. Этот материал включает табличные данные из 12

каталогов (VLSS, TXS, FIRST, NVSS, RCR, GB6, SDSS, USNO-B1, GSC 2.3.2, 2MASS, LAS UKIDSS, WISE), цифровые изображения из 5 обзоров (FIRST, NVSS, SDSS, LAS UKIDSS, WISE), составные рисунки радио-оптика, а также параметры и характеристики источников, полученные авторами на основе анализа компилятивных данных. Большая часть атрибутов используемых каталогов вошла в компилятивный каталог. Таким образом, в компилятивном каталоге у одного источника имеется порядка нескольких сотен параметров. Отметим еще, у небесного объекта связи с строками каталогов могут быть не только «один-к-одному», но и «один-ко-многим».

Публикация данных современных обзоров в настоящее время чаще выполняется серией релизов, а не одной окончательной версией при полном завершении работ по обзору. Это добавляет авторам компилятивных каталогов дополнительный объем работ по актуализации данных и, как следствие, сравнение и проверку имеющихся результатов.

Астрономом для исследований нужно небо физических объектов, то есть, в первом приближении требуются механизмы поддержки связей между строками каталогов. В этом случае атрибуты таблицы становятся параметрами, характеризующими физический объект. Это относится и к информации, которую может добавлять пользователь, включая и актуализацию данных. И такие механизмы должны поддерживаться на уровне центров данных.

Остается еще информация, которая не представляется в табличном виде. В астрономии второй после таблиц формой представления данных являются изображения. Они являются потенциально более богатым источником информации, если считать, что их можно обработать повторно новым способом и получить дополнительные характеристики исследуемого объекта. Современные системы обработки астрономических данных еще не обладают встроенными средствами распределенных вычислений, и входные файлы для которых не включены в гриды данных. В таких условиях добычу знаний из цифровых изображений для одного объекта можно было бы реализовать путем передачи и хранения разнородной информации в виде «контейнера» или «стека файлов».

3. Астрономические стандарты хранения данных

В астрономии для представления данных используется несколько форматов, часть из которых являются внутренними форматами стандартных систем обработки астрономических данных. Объединяет их структура, которая включает описательную информацию и собственно табличные или цифровые данные. Описательная информация может включать параметры наблюдения (небесный объект, его координаты, телескоп, прибор, метеоусловия и пр.), цифрового обзора (область покрытия, частотный диапазон и др.) или каталога (количество объектов, точность измерений и т.д.).

Самыми широко распространенными форматами являются FITS (Flexible Image Transpost System) [16], принятый Международным астрономическим союзом в качестве стандартного формата для астрономических данных, и формат VOTable [17], являющийся форматом ввода-вывода для всех приложений, совместимых со

стандартами виртуальной обсерватории, которые разработаны IVOA (International Virtual Observatory Alliance).

3.1. Формат астрономических данных FITS

Формат FITS используется научными организациями и правительственными агентствами для хранения и передачи астрономических изображений и таблиц. Формат поддерживается практически всеми средствами обработки астрономических данных и архивными системами.

Первоначально FITS-формат (basic FITS) разрабатывался для транспорта между различными программными системами и компьютерными платформами цифровых массивов, полученных при наблюдениях на астрономических телескопах. Позднее были добавлены другие структуры данных, поскольку одно-, двух-, трехмерные цифровые матрицы не могут отобразить разнообразие структур астрономической информации. Правила описания и представления структур данных, отличных от матриц, называются расширениями FITS-формата. К ним относится расширение «*random groups*», которое используется для представления данных, состоящих из серии массивов, каждый предваряемый своим набором параметров; «*tables*» - представление ASCII-таблиц; «*binary tables*» - для таблиц, в которых каждое значение ячейки таблицы может быть вектором.

FITS-файл состоит из последовательности блоков, включающих *заголовок* и *группу данных*. Заголовок и данные записываются как логические записи длиной по 2880 байт. HDU (Head and Data Unit) включает целое число таких записей. Заголовок содержит описание данных, данные идут сразу после заголовка. HDU может повторяться несколько раз, а может содержать только заголовок. Каждая логическая запись заголовка содержит 36 80-символьных ASCII-строк. В строке записывается ключевое слово, его значение и комментарий.

Ключевые слова *SIMPLE*, *BITPIX*, *NAXIS*, *NAXIS1*, ..., *NAXISn*, *END* являются обязательными и должны появляться в заголовке в фиксированном порядке. Значениями ключевого слова *SIMPLE* могут быть T (формат файла согласуется полностью со стандартом) и F (не согласуется со стандартом).

Число битов в одном значении данных определяется *BITPIX*. Оно может быть равным 8 (беззнаковое целое), 16 (целое), 32 (целое), 32 (вещественное), 64 (с двойной точностью), причем число битов в изображении вычисляется по формуле: $NBITS = |BITPIX| * (NAXIS1 * \dots * NAXISn)$, где *NAXIS* - число осей массива данных (от 0 до 999. Если 0, то данных нет, имеется только заголовок) *NAXIS1* - число точек первой оси и т.д.

Ключевое слово *END* является концом заголовка. Если до конца 2880 байтовой записи еще остается место, то оно заполняется пробелами, если это конец заголовка, или нулями, если это конец данных.

Для расширений FITS-формата первым ключевым словом в заголовке является *XTENSION*. Его значение - символьная строка, содержащая название расширения (например для двоичных таблиц, *XTENSION=BINTABLE*). Итак, заголовок

определяет реальный формат и размер последующей группы данных, а также содержит описательную информацию, по которой можно определить, что собственно хранится в файле.

3.2. Формат обмена данными программных приложений виртуальной обсерватории VOTable

VOTable разрабатывался как формат для хранения и обмена данными, представленными в табличной форме, на основе опыта, полученного при разработке формата Astrores [18], и по образцу расширений FITS-формата для таблиц и двоичных таблиц.

VOTable опирается на стандарт XML и, тем самым, использует его механизмы, как в части проверки (validation) входного документа программными приложениями, так и использования XSLT-преобразований. VOTable имеет встроенные механизмы для работы с большими массивами данных, и может использоваться в грид-вычислениях.

VOTable можно использовать как для хранения и передачи данных, так и для передачи только метаданных. Таблица рассматривается как несортированный набор строк одинаковой структуры, которая определяется метаданными таблицы. Строка является набором ячеек, каждая из которых содержит примитивный тип данных или массив таких примитивов. Самая сложная структура, которая может быть представлена в ячейке VOTable, это многомерный массив.

Формат позволяет хранить данные и метаданные отдельно, поддерживая связь между ними. Часть, относящаяся собственно к данным, в VOTable может быть представлена в одном из трех форматов - TABLEDATA, FITS и BINARY. TABLEDATA является встроенной возможностью XML-формата, так что таблицы малых размеров полностью обрабатываются средствами XML. Данные в формат FITS «binary table» либо инкапсулируются в VOTable как файл, либо FITS-заголовок раскодируется в VOTable-метаданные. Режим передачи данных BINARY поддерживается VOTable-форматом для эффективности работы, причем в этом случае не требуются библиотеки для работы FITS-форматом и поддерживается потоковая парадигма передачи данных.

Как и XML, VOTable включает как элементы разметки, так и информационное наполнение, причем элемент имеет два тега, которые ограничивают некоторый контент. Элементы могут содержать другие элементы, а также иметь атрибуты в комбинации ключ-значение.

Он, как считают разработчики формата, совмещает два способа представления структурированных данных - XML и FITS. В нем применяются UCDs (Unified Content Descriptors) [19], чтобы формализовано отобразить смысловое содержание параметра самой таблицы или ее полей. Посредством элементов GROUP VOTable поддерживает иерархическую организацию, что обеспечивает необходимую гибкость в представлении данных. С помощью этого элемента колонки таблицы могут быть сгруппированы в сколь угодно сложные иерархии.

3.3. Сравнение форматов FITS и VOTable

FITS-формат создавался в начале 80-х годов 20-го века и предназначался только для транспорта данных между различными программными системами и платформами. Важным достижением было введение текстовой информации, описывающей смысловое содержимое и структуру передаваемых файлов, чего ранее не применялось к научным данным. По сути это было введение метаданных в форматы представления научных данных. Для формата была проведена формализация ключевых слов, описывающих параметры файлов, а также групп ключевых слов, отвечающих за физическое представление данных. В качестве примера такой группы можно привести параметры World Coordinate System (WCS), отвечающих за астрометрическую привязку цифровых изображений.

Формализация данных позволила реализовать стандартные библиотеки, которые используются во всех астрономических программных приложениях.

Астрономические данные, получаемые при наблюдениях, имеют разную внутреннюю структуру, и эта структура может меняться в зависимости от инструмента и способа наблюдений. В FITS-формате можно представить данные практически любой структуры, а также поддерживается широкий набор примитивов для отображения числовых форматов.

К минусам формата, который создавался в самом начале 80-х, можно отнести то, что он разрабатывался еще до появления современных веб-стандартов. По этой причине FITS-формат не поддерживает потоковую передачу данных, так как в заголовке необходимо указывать максимальный размер массива переменной длины (ключевые слова NAXISi). Отметим, что FITS не поддерживает символы Unicode.

В отличие от FITS-формата VOTable опирается на индустриальные стандарты, что позволяет применять механизмы веб-доступа и распределенных вычислений. В VOTable используются UCD, которые можно назвать прототипной моделью данных предметной области, связанной с астрономией. Формат позволяет включать существующие модели данных, например модель данных координаты-время [20].

VOTable, в отличие от FITS-формата, поддерживает обособление данных от метаданных с сохранением связи между ними, потоковую передачу данных и другие идеи современных распределенных вычислений.

4. Элементы смысловой связи как поддержка физических объектов

Информация о небесных объектах, собрана, в основном, в каталогах. Объединение больших каталогов, в которых представлена информация из различных диапазонов электромагнитного спектра, требует дорогостоящих операций кросс-идентификации. Строки из двух каталогов, содержащих различную, но при связывании вдвойне полезную информацию. Кросс-идентификацию приходится проводить многократно, поскольку связи между объектами каталогов не

сохраняется.

Если создается новый каталог, полученный из одного или нескольких существующих каталогов, но с добавлением какой-то новой информации, то для него, как правило, не отслеживается связь с родительскими каталогами.

Сложно работать с компилятивными каталогами, полученными на базе объединения нескольких каталогов разных диапазонов, даже если эта информация собирается по небольшому списку объектов. Если появляются новые релизы каталогов или новые каталоги, то надо заново для интересных пользователю объектов выполнять одни и те же запросы. Нет средств для оповещения пользователя о появлении новой информации для исследуемых объектов, а также для обновления данных пользователя.

Поскольку данные в астрономии не связаны по смысловому содержанию, а также не обеспечивается передача знаний о небесном объекте, полученных другими исследователями, то невозможно выполнить запрос типа «найти все источники в каталогах, которые являются квазарами» и т.п. Это существенно ограничивает пользователю эффективную работу с информацией. В последнее время появилось несколько проектов, направленных на развитие сервисов, поддерживающих смысловую связность информации. Далее для примера приводятся несколько проектов.

4.2. AstroDAbis - семантическая связь каталогов

Цель проекта AstroDAbis [21] - создание независимого механизма публикации пояснений (комментариев, аннотаций). Пояснения могут создаваться пользователем для одиночного объекта («объект X есть квазар») или для нескольких объектов («объект с номером 123 в каталоге A есть то же самое, что объект с номером 456 в каталоге B»). Как полагают авторы AstroDAbis, этим решаются проблемы передачи знаний, создания компилятивных каталогов и реализации их связи с родительскими каталогами. Авторы статей, где представлена информация, полученная на основе анализа каталогов, с помощью аннотаций могут передать знания о небесном объекте в форме, которая может быть использована в последующих запросах к каталогу. Когда возникает потребность объединить два каталога и создать компилятивный, например, слияние оптических данных SDSS и инфракрасных данных тех же источников из UKIDSS, то такая связь позволит обойтись без повторной кросс-идентификации ресурсов. С помощью аннотации новые каталоги, полученные таким образом, можно сделать доступными для программного обеспечения, и связи между каталогами будут однозначно зафиксированы.

Результатом проекта AstroDAbis является прототип сервиса, который, в общем-то, является кросс-идентификацией нескольких существующих каталогов. Однако, он вносит новое качество в работу с имеющимися ресурсами. Целевая аудитория этого проекта, в первую очередь, астрономы, но взаимосвязанную информацию в базах данных можно предоставить и более широкой публике, что облегчит будущим разработчикам создание удобной системы поиска информации об астрономических объектах для любой группы пользователей.

p>Аналогичные разработки не являются новыми в науке. Например, имеется аннотирование данных в генетике - Distributed Annotation System, <http://www.biodas.org>), или в Интернете - RDF (Resource Description Framework) и LOD (Linking Open Data)). AstroDAbis также имеет LOD-интерфейс, который обеспечивает создание URI для аннотируемых объектов каталогов, что подготавливает платформу для экспериментов с Semantic Web в астрономии.

4.3. ADSASS - семантическая связь цифровых изображений и библиографии

Разработки проекта ADSASS (The ADS All-Sky Survey) [22] направлены на превращение системы NASA ADS (Astrophysics Data System), широко известной среди астрономов своей полнотой в качестве полнотекстового библиографического ресурса, в карту неба. Система ADS не является источником наблюдательных данных, но является неявным хранилищем ценных астрономических данных из публикаций в форме изображений, таблиц и ссылок на небесные объекты. Необходимо сделать эту ценную информацию доступной для запросов и просмотра. Рассматриваются три категории данных:

1) ссылки на небесные объекты, которые предполагается собрать из внешних баз данных и добавить связь в виде аннотации (astrotag) со статьями в ADS. Так же, как это сделано в geotags для объектов на земной поверхности для ГИС-систем, astrotags являются координатными и временными аннотациями для небесных объектов;

2) изображения, имеющиеся в статьях, также получают astrophotoreference-аннотацию, аналогичную georeferencing, которые ссылаются на карты, имеющие привязку к системе земных координат. Astrophotoreferencing свяжут изображения, которые будут приведены к одной небесной координатной системе, с учетом ориентации, координатной привязки и масштаба пикселей каждого кадра;

3) другого сорта данные, такие как текст или подписи под рисунками, будут привязаны к координатам или имени источника.

В результате выполнения проекта астрономы получают карту неба, которая по нажатию клавиши будет активировать ссылки на статьи, показывая, какая часть неба в них описывается, а также слой исторических данных на базе хранилища astrophotoreference и изображений, извлеченных из статей, которые можно использовать для анализа. Система ADSASS будет опираться на постоянно обновляемую базу данных тэгов, которая предназначена как для обнаружения новой информации о небесных объектах по любой тематике, так и для поиска событий переменного характера по данным исторического слоя.

4.3. VODKA - актуализация данных

Чтобы найти и получить данные пользователь сам инициирует взаимодействие с виртуальной обсерваторией посредством клиентских приложений или веб-интерфейсов к базам данных. Всякий раз, когда пользователь хочет узнать о возможно уже появившихся обновлениях, ему надо повторить первоначальный

запрос, сравнить полученный результат с существующим и скопировать, если это требуется, данные. Постоянно растущие объемы данных, включающие новые релизы существующих обзоров и публикации новых каталогов требуют другого подхода при отслеживании свежей информации о небесных объектах, интересных пользователю. Особенно это полезно при обновлении и актуализации компилятивных каталогов, баз данных. Решение этой задачи предлагается с помощью веб-приложения для поддержки данных пользователя VOdka (VO Data Keeping-up Agent) [23], который ретранслирует запросы пользователей в инфраструктуру ВО и рассылает уведомления об обновлениях. При выбранном пользователем темпе опроса агент асинхронно посылает один и тот же запрос, сформулированный пользователем, и фиксирует результаты, отражающие временной срез информации, выполняет сравнение этих срезов и оповещает пользователя по электронной почте. У пользователя есть возможность просматривать результаты запросов, сохраненные в snapshot-файлах, журналы сравнения этих файлов, копировать снимки и новые появившиеся данные, а также инкрементальные файлы, включающие старые, новые и пропущенные данные.

5. Развитие астрономических форматов представления данных

Выше мы привели перечень проблем при работе с компилятивным каталогом и наборами цифровых изображений для исследуемого списка радиоисточников, которые, как нам кажется, можно решить другим подходом к организации данных.

5.1. Внутренний формат представления данных интерактивного атласа неба ALADIN

Программное приложение ALADIN [14], так называемый интерактивный атлас неба, предоставляет пользователю удобный доступ к распределенным астрономическим данным. ALADIN разработан в Страсбургском центре данных CDS, имеет развитый графический интерфейс для визуализации и анализа как цифровых изображений, так и табличных данных. В него встроен механизм для соединения и обмена сообщениями и файлами по стандартному протоколу IVOA SAMP (Simple Application Messaging Protocol) с другими приложениями виртуальной обсерватории. Дополнительно к графическому интерфейсу ALADINом можно управлять строковыми командами встроенного языка, которые можно вводить интерактивно или загружать текстовый файл с командами, а также вводить команды из другой программы. Последнее дает возможность использовать ALADIN в скриптовом режиме для выполнения повторяющихся команд или управления им в удаленном режиме. Удобным средством отображения загруженных в память приложения данных является стек, который образован из плоскостей или слоев, сложенных стопкой по мере копирования данных. Слой или плоскость сохраняет данные, полученные при единичном запросе из какого-то ресурса или в результате действия команд. Стек сохраняется во внутреннем формате приложения.

Хранение и передача накопленной информации об исследуемом объекте может

производиться в виде такого стека файлов.

5.2. Расширение FITS-формата для стека файлов

Как уже упоминалось выше, FITS-файл может, кроме обычного HDU-блока или блоков данных, которые представляют метаинформацию и 1-, 2-, 3-мерную матрицу чисел, может еще иметь блоки, включающие данные со структурой, отличной от матрицы. Такие данные называются расширениями [16] FITS-формата и разрабатываются для представления новых структур данных. На текущий момент в FITS-формате можно сохранять отдельные слои стек данных ALADIN, но стек целиком не сохраняется в файле FITS-формата. Нам представляется целесообразным подготовить предложения по разработке расширения FITS-формата для представления стека файлов, которые являются коллекцией данных для одного объекта, тем более, что механизм расширений определен в стандарте FITS. Структура расширения стека должна состоять из главного заголовка, который определяет тип расширения, а также параметры, идентифицирующие объект (имя, координаты, размер области и т.п.). Далее следуют HDU (Header-Data Unit) для каждого слоя стека. Заголовок для каждого слоя описывает его структуру и содержимое.

6. Заключение

Выше мы привели наши предложения по подготовке нового типа расширения FITS-формата, которое предназначено для передачи и хранения коллекции разнородных данных, собранных для исследуемого объекта. Как нам кажется, это могло бы повысить эффективность работы с многочастотными данными в астрономии. Можно эту идею развить и для VOTable-формата, включив сюда возможности смысловой связности данных, как в проектах AstroDAbis и ADSASS, так и актуализацию данных, как в агенте VOdka, добавив соответствующие слои в стек коллекции, в которых могла бы сохраняться информация о семантических связях и команды запросов для актуализации данных.

Аналогичные коллекции разнородных данных, которые имеют смысловую связь, значимую при изучении свойств объекта, используются и в других областях научных исследований, а также и в других областях человеческой деятельности. Как нам кажется, хранение и передача таких коллекций в виде стека («в одной коробке»), стандартизация формата для стека на базе XML может послужить развитию механизмов передачи и сохранения знания, повышению эффективности работы с разнородной информацией, объемы которой продолжают расти, простота доступа обеспечена инфраструктурой и технологиями Интернета.

Литература

[1] Parijskij Yu.N., Bursov N.N., Lipovka N.M., Soboleva N.S., Temirova A.V. The RATAN 600 7.6-cm catalog of radio sources from "Experiment Cold-80" // Astronomy and Astrophysics Supplement Series, 1991, 87, 1-32.

[2] Soboleva N.S., Majorova E.K., Zhelenkova O.P., Temirova A.V., Bursov N.N..

RATAN-600 7.6-cm Deep Sky Strip Surveys at the Declination of the SS433 Source
During the 1980-1999 Period. Data Reduction and the Catalog of Radio Sources in the
Right-Ascension Interval 7h