

ИНТЕГРИРОВАНИЕ ИНФОРМАЦИОННЫХ АСТРОНОМИЧЕСКИХ РЕСУРСОВ

А.Е.Авраменко², А.Н. Беликов¹, М.Л. Белоус¹,
О.Б. Длужневская¹, Ю.П.Ильясов², Е.Ю.Кильпио¹,
Д.А.Ковалева¹, О.Ю. Малков¹, А.Э.Пискунов¹, Л.А.Сат¹

¹ Институт Астрономии РАН (ИНАСАН), Москва, Россия

² Пушчинская Радиоастрономическая Обсерватория АКЦ ФИАН, Москва, Россия

1 Введение в астрономическую информационную среду(АИС).

Астрономия -одна из самых древних наук на Земле. Предметом ее исследования, в отличие от других наук, являются самые многочисленные, удаленные и недоступные для человека небесные объекты: звезды, галактики и т.п. Основным методом их изучения являются наблюдения, в результате которых определяются точные положения объектов на небе, их физические, кинематические и др. характеристики. Эти наблюдения осуществляются как в различных диапазонах длин волн, так и различными наблюдателями. Эти наблюдения оформляются, как правило, в каталоги, которые первоначально издавались в печатном виде. При появлении возможности их компьютерной обработки стали создавать их в машиночитаемой версии. Астрономы были одни из первых, кто отказался от бумажных носителей и перешел на магнитные.

В конце 70-х годов нашего века была создана сеть международных Центров Астрономических данных с целью создания, хранения и распространения астрономических каталогов.

Один из последних типичных каталогов, полученный в результате выполнения космического проекта Hipparcos международным консорциумом, размещается на 6 CDROM'ах и имеет объем порядка 3.5 Гбт вместе с описанием эксперимента и вспомогательными программами для его обработки. Но существуют каталоги и большего объема, например, USNO A2 содержит более 500 млн. объектов и имеет объем 11 CDROM-ов (более 6 Гбт).

Каталоги и по сей день есть основа нашей астрономической коллекций данных. Со временем появляются все более свежие их версии. Ныне из-за резко увеличившегося объема новых данных, их неоднородности и постоянного обновления, а также других неудобств, лучшим хранилищем подобных данных стали базы данных(БД). Они отличаются от каталогов тем, что предоставляют ряд сервисных услуг по обработке, выборке, поиску, пополнению содержащейся в них структурированной информации. Тем не менее основной единицей хранения информации в астрономии остаются каталоги, содержащие от нескольких до тысяч файлов.

Первая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
19 - 21 октября 1999 г., Санкт-Петербург

В астрономии, как и в других областях, существует как предметная, так и региональная специализация в хранении и поддержании АИ. Например, один из главных архивов фотопластинок хранится в Англии. Фонд практически всех известных каталогов располагается в Страсбургском центре звездных данных(CDS) во Франции

<http://cdsweb.u-strasbg.fr/>

2 Архивы ЦАД, ИНАСАН,библиотеки и их интегрирование в АИС.

Центр Астрономических Данных (ЦАД)

<http://www.inasan.rssi.ru/~lsat/index.html>

был организован в 1980 г. при Астросовете АН СССР, ныне (ИНАСАН)

<http://inasan.rssi.ru/> ,

как филиал CDS для обмена астрономическими каталогами между Европой и СССР [1]. Двадцать лет назад наш фонд каталогов размещался на громоздких магнитных лентах, занимавших несколько шкафов, теперь архив вырос в сотни раз, а занимаемый им физический объем резко сократился, превратившись в основном в две сотни CDROM-ов и магнитооптических дисков(МОД).

Попрежнему нашей основной задачей остается создание, поиск, учет, хранение, классификация, развитие и распространение наших как национальных, так и региональных астрономических ресурсов, их интеграция в мировые информационные системы, а также поиск и организация доступа к мировым астрономическим ресурсам в сети как для российских и зарубежных специалистов, так и для любителей и всех интересующихся астрономией.

ЦАД является одним из четырех ведущих мировых центров данных и хранит практически зеркальную копию мирового фонда астрономических каталогов. Помимо этого ЦАД интегрирует каталоги и другую астрономическую информацию на пространстве бывшего СССР и передает в мировой центр данных в Страсбурге.

ЦАД участвует в кооперативных усилиях по предоставлению пользователям разнообразной АИ. Как пример, можно привести совместный проект по интегрированию новых данных астрометрических наблюдений пульсаров в АИС, выполняемый ЦАДом совместно с Пушчинской радиоастрономической обсерваторией (ПРАО) ФИАН.

Необходимость присоединения наблюдательных радиоданных (десятки и сотни мегабайт за сеанс наблюдения) к уже имеющимся данным для совместной обработки, требуют систематизации, долговременного хранения и извлечения данных и результатов, связывания со средствами обработки и использования их в прикладных целях [2]. Для этих наблюдений в ПРАО разработаны соответствующие форматы данных, представляющие собой иерархическую структуру с последовательным обобщением и сокращением первоначальной измерительной информации. Все данные хранятся в электронной форме в виде архивов. Каждый иерархический уровень данных поддерживается набором аппаратно-программных средств, образующих специализированную кластерную структуру, которая обеспечивает:

- целостность и защиту данных;
- связь данных с программой обработки;
- авторизованный доступ и межуровневый обмен.

Кластерная структура поддерживает необходимый набор ресурсов, который определяется спецификой уровня. В частности, такая структура на радиотелескопе включает в себя станции сбора данных, которые управляют приемно-регистрирующей аппаратурой и формирует в стандартном виде первичные данные наблюдений. Локальная сеть, объединяющая станции, агрегирует первичные данные, поступающие во время наблюдения [3]. Программная оболочка для взаимодействия с ресурсами в процессе извлечения и обработки данных предусматривает набор средств, обеспечивающий независимый от платформы пользователя доступ к локальным и удаленным данным, связывание их с программными пакетами обработки и представления результатов. По мере усложнения моделей поведения исследуемых объектов представляется перспективным разработка новых типов объектов, интегрирующих как данные, так и функции их поведения. Такой подход может оказаться полезным также для достижения интероперабельности перемещаемых объектов, например, данных совместно с программами, когда они используются на специализированных физических установках. Получаемые каталоги представляют собой верхний уровень обобщения наблюдательных данных, интегрирующий все существенные результаты предшествующих исследований. Каталоги содержат упорядоченную совокупность признаков, представляющих интерес для любых приложений. Наблюдательные данные аккумулируются и хранятся в унифицированном формате Европейской пульсарной сети (EPN-формат), согласованном и принятом ведущими обсерваториями в качестве единого стандарта обмена [4].

Работа с каталогами осуществляется по схеме клиент-сервер. Используется браузер, интерпретирующий HTML-формы запросов, динамические HTML-страницы с документами, извлекаемыми и пересылаемыми по запросам, графические данные. Исходные файлы каталогов представляют собой базы данных с упорядоченной структурой параметров и фиксированным форматом записей.

Базы данных поддерживаются SQL-сервером, который связан с пользователями через CGI-шлюз. Нами была принята двухступенчатая организация каталогов с разделением функций подготовки, поддержания в актуальном состоянии исходных баз данных, с одной стороны, и удаленного интерактивного доступа к ним через Internet

с другой стороны. Подготовку, пополнение баз наблюдательных данных проводит ПРАО по мере обработки вновь получаемых данных. SQL-серверы с сетевой поддержкой каталогов, программным обеспечением и ресурсами реализуются в рамках ЦАД. Кооперация с ЦАД позволяет использовать развитые средства связи и массивы современных электронных библиотек. В рамках выполняемого проекта разработаны программные средства для удаленного обращения пользователя по http протоколу для Принстонского каталога пульсаров
<http://seth.inasan.rssi.ru/~pucshino>
и журнала наблюдений
<http://seth.inasan.rssi.ru/pucshino/eprn.cgi>
(в настоящее время тестовая версия). Исходная база наблюдательных данных размещается на сервере ПРАО ФИАН:

<http://psun32.prao.psn.ru/data/kalyazin/index.html>
(авторизованный доступ).

Астрономические ресурсы хранятся в ЦАД в форматированном и структурированном видах в различных каталогах и базах данных. В отличие от коллекции, называемой "Электронной библиотекой-[5] (от греч. *biblion* - книга), электронные астрономические архивы правильнее было бы называть в таком случае файлоотекой.

В последнее время на базе современных компьютерных и телекоммуникационных технологий создаются объединенные предметно-ориентированные информационные ресурсы, содержащие как каталоги, базы данных, архивы, библиографию и т.п., так и ссылки на другие коллекции. В астрономии это могут быть сведения о других центрах, институтах, наземных обсерваториях, их Web-сайтах, наблюдательных и др. ресурсах, персоналии и пр. Подобную коллекцию можно было бы назвать более общим понятием, чем Электронная библиотека, а именно АИС. В дальнейшем условимся упоминать Электронная библиотека и АИС в одинаковом смысле.

Основную часть коллекции ЦАД занимает архив астрономических каталогов. Он содержит более 500 единиц, размещенных на 120 CDROM-ах и МОД-ах (общим объемом более 70 Гбт), остальные каталоги (более 4 Гбт) располагаются на других носителях. Фонд каталогов постоянно пополняется. Существуют краткий гид по фонду каталогов. Планируется разработать локальную поисковую систему для облегчения доступа к фонду по ключевым словам, номерам каталогов, авторам и пр. Будет создана автоматизированная система для получения выборок из каталогов по координатам и др. параметрам объектов.

В on-line доступе у нас находится небольшая часть фонда, другая часть может быть предоставлена пользователям по предварительному запросу через электронную почту. Предоставляется также возможность записать необходимые данные на CDROM-ы. Реализован пакетный режим обслуживания с помощью электронной почты для пользователей, не имеющих полных Internet-возможностей. В архиве ЦАД хранятся и будут в свободном доступе постоянно пополняемые электронные таблицы данных из статей, опубликованных в российских астрономических журналах. В стадии разработки находится:

- База данных маломассивных компонентов двойных и кратных систем (LOMBI).
- База данных Областей звездообразования в Галактике.

- Web интерфейс к упакованной версии GSC (один из самых больших каталогов).
- Модель межзвездного поглощения в Галактике, данные которой планируются в открытый доступ.
- Программа визуализации электронных атласов звездного неба.
- Пилотная версия "Автоматизированной библиотечной системы", позволяющая наряду с другими возможностями быть в курсе "Выставки новых поступлений" в библиотеку ИНАСАН (СУБД Informix).

К создаваемой в настоящее время Электронной библиотеке, (см. рис. 1 и Web сайт ЦАД) или АИС будут подключены также следующие ресурсы ИНАСАН и его библиотеки <http://www.inasan.rssi.ru> :

- Постоянно поддерживаемая и администрируемая база данных INES (IUE Newly Extracted Spectra) <http://ulda.inasan.rssi.ru>, содержащая результаты космического эксперимента IUE.
- Библиографическая база данных ASTRODATA (120 тыс. тыс. записей, 40Мбт). OLIVETREE Associates (Sierra Madre, California) (1991 - 1995 г.).
- Архив циркуляров малых планет и комет (MPEC) (объем 10Мбт). САО, Кембридж, США, с 1993 года
- Архив циркуляров Международного Астрономического Союза (IAUC) . (САО, Кембридж, США), с 1992 года (общ.объем 4 Мбт).

Создается общедоступный электронный архив авторефератов, диссертаций сотрудников института; институтских сборников, книг, материалов конференций, научных отчетов (см.раздел Electronic Archives на нашей Web странице).

3 Использование региональных и мировых астрономических ресурсов.

Постоянно осуществляется поиск интересной астрономической информации, доступной через Internet, главным образом, в нашем регионе, т.к. в западных информационных ресурсах она отражена слабо. Эта информация анализируется и, как правило, предлагается пользователям в более удобном систематизированном виде, по возможности, сопровождаемая краткими характеристиками ресурса. В результате появляются новые ссылки в разделе БД, список астрономов нашего региона пополняется новыми email адресами и т.п. Учитывая тот факт, что ЦАД является элементом международной сети Центров, наш регион будет в полной мере интегрирован в мировые астрономические ресурсы.

Главным центром в Европе, как уже было упомянуто является CDS. Он располагает самой крупной звездной БД SIMBAD, содержащей на август 1999г. 2 717 295 объектов с различными параметрами и 107397 библиографических ссылок. Их фонд каталогов и таблиц содержит около 3200 единиц. Большая часть находится в on-line, остальная часть предоставляется в on-line по запросу. В совместном международном проекте AstroWeb собрана разнообразная астрономическая информация, подобная

той, что мы обсуждали в АИС. ЦАД осуществляет тесное сотрудничество с CDS и активно использует накопленный там материал. Мы взаимодействуем также с другими Центрами, среди которых самым крупным является Astronomical Data Centre в США. Он располагает более 670 каталогами и другой астрономической информацией. Ссылки на Web- сайты ADC и других Центров доступны из нашей Электронной библиотеки.

4 Структура Электронной библиотеки (АИС).

На рис.1 приведено содержание нашей Электронной библиотеки. Как видно из рисунка, она содержит следующие основные разделы.

1. Справочный, куда включены иностранные словари, справочники, библиографические каталоги, различные средства навигации в массивах данных как в ЦАД (гид по фонду каталогов), так и выходы на известные поисковые системы.
2. В отдельный раздел "Фонды" выделены каталоги и БД.
3. Периодика. Здесь установлен Web-доступ к электронным версиям 24 астрономических и физических журналов русских и иностранных издательств <http://www.inasan.rssi.ru/~lsat/eLib-Journals.html> Сюда же относится и раздел "Препринты".
4. Уже упоминавшийся раздел "Электронные архивы" будет включать подобные материалы из других коллекций.
5. Раздел "Проекты" знакомит с планируемыми наземными и космическими экспериментами.
6. Включены также материалы по разработке Электронных библиотек и ссылки на другие астрономические и близкие по темам электронные коллекции.
7. Раздел Новостей будет держать специалистов в курсе последних событий в мире астрономии.

5 Заключение.

Для реализации наших планов, нами было запланировано при достаточной финансовой поддержке грантом РФФИ, приобретение CDROM jukebox на 100 дисков (устройство хранения, записи и доступа к оптическим дискам). Сервер устройства является Web- и ftp-сервером. К сожалению, необходимо отметить, что многие наши планы и будущие проекты выполнимы исключительно при постоянной финансовой поддержке.

Как и многие в России, мы часто испытываем проблемы в устойчивой работе нашей АИС. Однако, активное использование информации, накопленной в Центрах данных, позволяет даже удаленному пользователю, не имеющему хорошей библиотеки, тем не менее держать руку на пульсе астрономической жизни, осуществлять обмен информацией, повышать качество и оперативность научных исследований и публикаций.

Работа поддержана грантами РФФИ € 97-07-90069 и N 98-07-91154

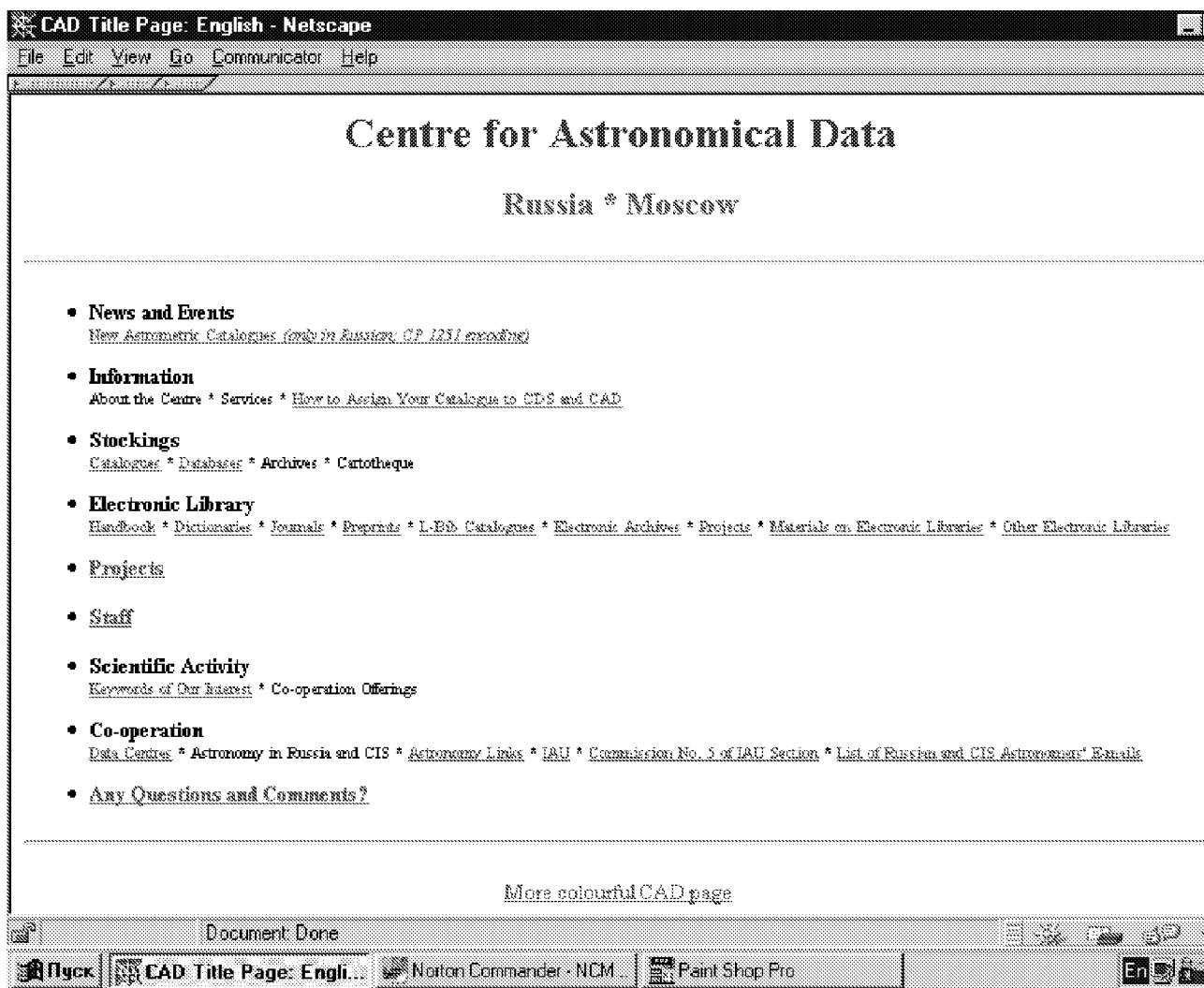


Рис. 1: Титульная страница Web-сервера ЦАД

Библиография

- [1] Proceedings of the international colloquium "International cooperation in dissemination of the astronomical data". St.Petersburg. July 2-9,1996. "Baltic Astronomy", vol.6,N2,1997.
- [2] Avramenko A.E., Akhmetov R.R., Doroshenko O.V., Pyasov Yu.P., Kitaev V.V., Potapov V.A. "Pulsar Astrometry Databases Management in a Cooperative Astronomical Time and Coordinate Research", Proc. of the First East-European Symposium on "Advances in Databases and Information Systems"(ADBIS'97), 2, St.Petersburg, 82 (1997).
- [3] Авраменко А.Е., Гипшиус А.А., Горшенков Ю.Н., Дедюхин А.В., Илясов Ю.П., Лычкин В.П., Попереченко Б.А., Хечинашвили Г.Н., "Система автоматизации научных исследований на радиотелескопе ТНА- 1500Ц", Препринт ФИАН, е 53 (1993).
- [4] "TMR Research Network Proposal "European Timing of Pulsars"(EuroToP)", Bonn, Germany (1997).
- [5] Журнал "Электронные библиотеки", <http://www.iis.ry/el-bib/>