

Национальный виртуальный центр геофизических данных*

© В.Н. Ишков
ИЗМИРАН
ishkov@izmiran.ru

© И.А. Кузьмин
МЦД ГК РАН
kuzmin@wdecb.ru

© Е.П. Харин
МЦД ГК РАН
kharin@wdecb.ru

Аннотация

Мировой центр данных по солнечно-земной физике в Москве совместно с институтами геофизического профиля РАН ведет работу по созданию Национального виртуального центра геофизических данных на основе исторических и текущих баз данных и архивов геофизических и солнечных обсерваторий, сведенных в единую систему распределенной сети с единым координирующим центром. Центр должен обеспечить пользователей возможностью выбирать, визуализировать и моделировать исторические и текущие данные по физическим условиям на Земле, в околоземном космическом и межпланетном пространстве и на Солнце. Виртуальный центр должен допускать интеграцию с вычислительными геофизическими моделями состояния околоземного космического пространства, как в реальном времени, так и в любой требуемый промежуток времени. Структура Виртуального центра по сути является распределенной базой данных и является Web-узлом, связанным с серверами организаций и держателями баз данных по протоколам http и ftp.

1 Введение

Для удовлетворения информационных и научных потребностей российских специалистов по солнечно-земной физике и широкой научной общественности назрела необходимость создания общедоступного через ИНТЕРНЕТ Национального виртуального центра геофизических данных (НВЦГД). В настоящее время понятие "геофизические данные", кроме данных о планете Земля, включает в себя и все характеристики космической погоды в околоземном космическом пространстве, а следовательно наблюдательные данные о всей цепочке солнечно-земных связей, включая явления на Солнце, в межпланетной среде и в околоземном космическом пространстве в различных диапазонах длин волн электромагнитного излучения и потоков энергичных частиц. Поэтому нет никакой

возможности свести и хранить эти данные в одном центре. С другой стороны данные получают многочисленные исследователи и организации и информация об этих данных, а также поиск их и доступ к ним не так прост даже с использованием современных информационных технологий. Современные тенденции развития геофизических наук приводят к тому, что

- архивы и базы данных наблюдательного материала приобрели распределенный характер;
- архивы и базы данных становятся общедоступными, что резко расширяет круг их пользователей;
- возникают глобальные международные проекты [1] с более эффективным разделением труда внутри научных коллективов и геофизические исследования начинают терять национальный характер;

- создана сеть Мировых центров геофизических данных, например [2, 3, 4], как посредников для объединения и доступа к региональным научным исследованиям с последующим созданием единого информационного пространства.

Именно на основе сложившейся системы Мировых центров данных по солнечно-земной физике предлагается создание виртуального центра геофизических данных – распределенной сети синхронных баз данных и серверов приложений, позволяющих хранить, получать, выбирать, визуализировать и моделировать исторические и текущие данные по физическим условиям на Земле, в околоземном космическом и межпланетном пространстве и на Солнце. Виртуальный центр должен допускать интеграцию с вычислительными геофизическими моделями состояния околоземного космического пространства, как в реальном времени, так и в любой требуемый промежуток времени. Распределенная архитектура центра является ключевым фактором для сокращения времени ожидания при визуализации мультимедийных данных и быстрой доставки данных на компьютеры пользователей для последующей обработки и анализа.

При создании и разработке концепции НВЦГД активно использовался опыт Российской виртуальной обсерватории (РВО) [5], которая обеспечивает предоставление пользователям доступ в реальном времени ко всем ресурсам актуальной

астрономической информации. Тем более, что базы данных, коллекции и архивы наблюдений Солнца, ближайшей к нам звезды, будут интегрированы как составная часть РВО.

2 Архитектура и компоненты распределенной системы виртуального центра

2.1 Структурно-алгоритмическая организация.

В процессе создания НВЦГД используются уже известные принципы организации структуры, которая должна включать следующие основные компоненты: ИНТЕРНЕТ-портал, центр данных, в котором хранятся каталоги, библиотеки программ, архивы – доставка данных, стандартные процедуры обработки и визуализации данных. Доступ к данным организуется по иерархическому принципу. На верхнем уровне находится информация о всех доступных базах, каталогах, архивах и коллекциях данных со ссылкой на авторов и организациях хранителей данных, подробным описанием носителей данных. На следующем уровне располагаются сами каталоги, коллекции и базы данных. На третьем уровне располагаются сами данные. Первый уровень располагается на главном сервере виртуального центра, а второй и третий уровни представляют собой распределенную сеть баз данных и серверов приложений позволяющих выбирать, визуализировать и моделировать данные. Сами данные хранятся частично на сервере НВЦГД, а, в основном, на серверах организаций держателей данных. Распределенная архитектура системы является ключевым фактором для сокращения времени ожидания при визуализации мультимедийных данных и быстрой доставки данных на компьютеры пользователей для последующей обработки и анализа. Узел системы обычно реализуется на компьютерном кластере и обеспечивает параллельный поиск, обработку и визуализацию больших объемов данных с применением методов искусственного интеллекта и нечеткой логики. Серверы, входящие в систему виртуального центра будут иметь интерфейс управления базами данных, который позволяет выполнять модификацию данных удаленным пользователям с правами администратора по сети ИНТЕРНЕТ. Узлы системы НВЦГД должны автоматически обмениваться обновлениями баз данных и программного обеспечения. Каждый узел будет образован web-сервером, связанным с сервером приложений, в свою очередь связан с несколькими локальными или распределенными серверами баз данных. Основные задачи сервера приложений – доступ к метаданным, интеллектуальный поиск данных, визуализация, моделирование и доставка данных пользователям. Web-сервер осуществляет интерактивную связь с сервером приложений по протоколу HTTP. Для программного обеспечения предполагается использование языка Java, что сделает систему

масштабируемой и независимой от операционной системы конкретного сервера приложений.

Заинтересованный пользователь входит в Виртуальный центр через ИНТЕРНЕТ и попадает в главное меню. Здесь он получает возможность познакомиться со структурой НВЦГД, организациями входящими в него, типами доступных баз данных, каталогов и коллекций, описанием интерфейса пользователя, описанием проблемно-ориентированных баз данных под национальные и международные проекты (например, планируемые Международный Геофизический Год и Международный Полярный Год в 2007 году).

Поиск данных может вестись как по типам данных, по заданным временным интервалам, так и по регионам. При задании типа данных пользователю выдается информация о наличии каталогов или коллекций с указанием держателей данных. При запросе самих данных система подключает пользователя к держателю данных, который через соответствующий интерфейс начинает общаться с пользователем. При задании временного интервала выдается информация о наличии каталогов тех видов наблюдений и держателей информации, которые доступны в настоящее время. После соответствующего выбора пользователю выдаются требуемые данные из соответствующих каталогов или доступ в сами каталоги. Далее система подключает пользователя к держателю данных и т.д. При задании географического региона сначала производится выборка всех держателей данных в данном регионе, затем поиск требуемых баз данных и каталогов. Полученная от держателей данных информация по желанию пользователя может транспортироваться непосредственно пользователю, либо в соответствующий раздел интерфейса пользователя для интерактивной обработки данных стандартным набором программ.

Каждый зарегистрированный пользователь автоматически получает специальный объект контейнер, который содержит список всех наборов данных, отобранных пользователем для графического или модельного анализа и экспорта из системы. Эта информация сохраняется от обращения к обращению так, чтобы всякий раз, когда пользователь вновь входит в систему, он уже имел список наборов данных и интервалы дат, с которыми он работал в предыдущий сеанс.

Такой контейнер позволяет иметь одновременную выборку и визуализацию многодисциплинарных данных, существенно расширяя возможности анализа и доступа к данным в целом. Например, запросив данные по геоэффективным солнечным событиям в определенный временной интервал и, кроме того, данные по солнечному ветру и состоянию геомагнитного поля – различные формы запросов и различные базы данных – пользователь получает возможность построить все графики на одной

странице в едином масштабе времени и загрузить все запрошенные данные с сервера.

2.2 Компоненты создаваемого НВЦГД

В качестве базовых принципов построения НВЦГД предполагаются:

- объединение в одной распределенной системе архивов геомагнитных и солнечных обсерваторий и центров данных со стандартизованными средствами сетевого доступа и интерфейсами;
- организация для запросов на получение информации единого входного портала, реализующего удовлетворение запроса в существующей информационной базе и постановку в очередь на проведение геофизических экспериментов на необходимом инструментальном комплексе (например, при наблюдениях Солнца);
- реализация процедуры транспортировки пользователя к источникам данных, а не данных к исследователю, подразумевая под этим расширенный доступ пользователя к информационным и вычислительным ресурсам, включающий возможность использования на базе Виртуального центра собственных программных средств.

Опыт работы показал, что функционирование таких виртуальных образований (ВО) должно основываться на быстро развивающихся технологиях в телекоммуникациях и информатике. Чтобы обеспечить их постоянную жизнеспособность должны выполняться определенные условия:

- ВО должно эволюционировать, то есть быстро реагировать на изменение технических и научных возможностей и потребностей научных исследований;
- ВО в принципе распределенная структура, поскольку значительное количество научного опыта по работе с архивами данных уже имеется в существующих центрах данных. Создание новых центров и развитие новых направлений в уже существующих происходит, и будет происходить в географически разнесенных местах;

- все компоненты ВО должны быть максимально интегрированными между собой и поддерживать необходимые перекрестные связи между ресурсами разных серверов в соответствии с общей открытой схемой баз данных, каталогов и архивов, а также обеспечивать в интерфейсах возможность обмена и объединения между ресурсами из разных компонент;

- ВО должно иметь развитые интерфейсы для поиска и навигации в пространстве ресурсов на русском и английском языках, а также средства администрирования и поддержки всех баз данных, и архивов входящих в НВЦГД;

- ВО должно поддерживать общепринятые средства тематической классификации ресурсов (рубрикаторы, тезаурусы, другие системы классификации) и обеспечивать поиск ресурсов по этим средствам классификации;

- для большей эффективности в научных исследованиях ВО должно действовать объединяющее для всех возможных данных наблюдений в различных диапазонах электромагнитного и всего спектра энергий корпускулярного излучений;

- ВО должно быть открытым для образовательных и познавательных целей;
- ВО должно быть ориентировано на глобальное использование, сотрудничая с другими странами в подобных проектах.

Важным компонентом виртуального центра, вероятно, будет информационная система Центра изучения геофизических данных ЦИГЕД ИФЗ РАН [6]. Базы данных и программное обеспечение системы размещены на вычислительном кластере из 12 параллельных компьютеров под управлением операционной системы Linux, соединенных между собой по протоколу Fast-Ethernet 100 Мбит с высокоскоростным оптоволоконным 100Мбит выходом в ИНТЕРНЕТ. Управление и внешний доступ к кластеру осуществляются через сервер приложений с двумя процессорами Intel Pentium III 1 ГГц с оперативной памятью 1 Гбайт. Доступ к системе осуществляется через web-сервер Apache 1.3.19 и JSP-сервер 1.3.2, которые установлены на сервере приложений. Запросы на выборку данных из различных баз данных, входящих в систему, рассылаются сервером приложений по протоколу JDBC и обрабатываются параллельно СУБД (MySQL 3.23.31) на нескольких узлах кластера, на которых размещены запрашиваемые базы данных. Это позволяет осуществлять многим пользователям одновременную выборку, визуализацию и анализ больших наборов данных из различных разделов солнечно-земной физики в интерактивном режиме.

2.3 Базы данных и архивы - будущие составляющие НВЦГД

Система Виртуального центра объединит распределенные тематические базы данных, коллекций и архивы по различным направлениям солнечно-земной физики. Кроме того, в нее войдут дополнительные базы данных для административных целей. Адреса всех распределенных серверов баз данных объединит общий конфигурационный файл. Административная база данных пользователей регионального узла связывается со встроенной базой данных отзывать и предложений по работе системы. База метаданных должна содержать имена Java-классов для загрузки и экспорта данных из тематических баз, интервалы времени наличия этих данных, физические единицы измерений, заголовки и способы размещения графиков временных рядов, имена и координаты солнечных, геомагнитных, ионосферных обсерваторий, станций космических лучей, а также параметры орбит, времен жизни и характеристики приборов научных спутников, данные от которых войдут в систему. Предусматривается создание ядра системы – базы данных основных солнечных,

геомагнитных и ионосферных индексов, параметров межпланетного магнитного поля и солнечного ветра, которое должно быть установлено на каждый узел НВЦГД, так как к ним постоянно обращаются подавляющее число пользователей.

Дополнительные базы данных могут быть установлены локально на узлах системы, но обычно должны запрашиваться с удаленных узлов сервером приложений, в зависимости от конкретной конфигурации. В них должны войти базы данных изображений Солнца и отдельных активных солнечных явлений с различных спутников и наземных обсерваторий, базу данных параметров околоземного космического пространства со всех спутников, временные ряды наблюдений ионосферных и геомагнитных станций, временные ряды наблюдений станций космических лучей и каталоги событий космической погоды. Любая новая база данных может быть установлена в систему посредством объединенного интерфейса без остановки работы сервера приложений. Чтобы добавить новую базу данных в систему, нужно указать сетевой адрес новой базы данных в конфигурационном файле, подготовить и разместить на web-сервере системы HTML форму для запроса данных и добавить метаданные для нового типа данных (имя Java-класса для доступа к данным, размер окна, логарифмические/линейные масштабы, метки осей на графиках и т.д.).

2.4 Представление изображений в НВЦГД

Проблема визуализации временных рядов в системе Виртуального центра будет решаться главным образом на технологиях Java-апплетов, динамически встраиваемых в web-страницу. Если на странице необходимо разместить больше, чем один график, Java-апплеты взаимодействуют между собой и синхронно масштабируют временные оси всех графиков. Если потребитель захочет изменить масштаб изображения в интересный ему период времени на одном из графиков, все остальные графики на этой странице автоматически масштабируются к тому же временному окну. Для распечатки графиков на принтере и для включения их в статистические электронные документы сервер приложений может создавать графики временных рядов в виде статических изображений в формате GIF. Однако в этом режиме масштабирование графиков будет недоступно.

В системе Виртуального центра должна быть решена еще одна задача визуализации – построение карт с расположением геомагнитных, ионосферных и солнечных обсерваторий и станций космических лучей станций. Ее решение также возможно с применением специального картографического Java-апплета, встроенного в HTML форму, которая получает с сервера приложений системы список с именами, координатами и временами функционирования обсерваторий и станций. Пользователь может по своему желанию вызвать нужные станции и в любой момент может послать

запрос на добавление или удаление любой станции из списка выбранных станций.

Визуализация изображений Солнца и любых активных явлений на Солнце в разных диапазонах, полученных со спутников и наземных станций предполагается с помощью периодически обновляемого HTML слайд-шоу. Подобный доступ к большой базе данных изображений удобен, когда длина последовательности показываемых изображений неизвестна заранее.

3. Заключение

Организация и функционирование НВЦГД должно отвечать следующим принципам:

1. Все исследователи и организации держатели данных входят в систему на добровольных началах.
2. После включения держателей данных в систему НВЦГД они должны подчиняться требованиям и правилам структурной организации Виртуального центра.
3. Все исследователи и организации держатели данных при этом остаются исключительными собственниками данных с правом разрешения использования данных пользователями, как без ограничения, так и на договорных началах. В свободном доступе будет находиться только информация о наличии данных и каталогах, коллекциях и архивах.
4. Не будет предоставляться информация, связанная с секретными данными либо с данными, связанными с национальной безопасностью.
5. Контроль качества данных возлагается на держателей данных.

Литература

- [1] IMAGE 21 – Международный мониторинг авроральных геомагнитных эффектов Web site 1998, .
<http://www.geo.fmi.fi/image.data.html>
- [2] Мировой центр данных по солнечно-земной физике, Москва.
<http://www.wdcb.ru/stp/index.ru.html>
- [3] Мировой центр данных по индексам солнечных пятен, Брюссель (Бельгия).
<http://www.astro.oma.be/SIDS/index.html>
- [4] Мировой центр данных по солнечно-земной физике, Боулдер (США).
<http://www.ngdc.noaa.gov/stp>
- [5] О.Б. Длужневская, О.Ю. Малков, и др., Российская виртуальная обсерватория. Информационная система центра астрономических данных. Труды четвертой Всероссийской научной конференции RCDL 2002, Дубна, 2002, с. 202-211.

- [6] М. Жижин, А. Бурцев, Е. Харин.
Интерактивный ресурс по солнечно-земной
физике. Солнечно-земная физика, вып. 2. Изд.
СО РАН, Иркутск, 2002 г., с. 119 – 123.

National Virtual Center of Geophysical Data

V. Ishkov, I. Kuzmin, E. Kharin

The National Virtual Center Of Geophysical Data is a system in which the vast geophysical archives and databases around our country and the country of former USSR, together with analysis tools and computational services are linked together into an integrated facility. Solar-Terrestrial Physics play important role in the space exploration as the human society begun to use the outer space for practical purposes. The article deals with the description, main goals and current status of the NVCGD.
