

КОНЦЕПЦИЯ И ОПЫТ ПОСТРОЕНИЯ СОВРЕМЕННОГО АРХИВА ДАННЫХ КОСМИЧЕСКОГО ЭКСПЕРИМЕНТА

Ирина Белова, А. Петрукович, О. Батанов

Институт космических исследований – Российская академия наук
e-mail: ibelova2004@mail.ru

Ключевые слова: база данных, Web-интерфейс, формат CDF

Абстракт: В статье перечислены основные требования к современным системам хранения и отображения данных космического эксперимента и изложен опыт построения такой системы в ИКИ-РАН.

Введение

В современных условиях количество и объем данных космических экспериментов резко возросли, что, в условиях развития компьютерных и телекоммуникационных технологий по-новому поставило вопрос хранения данных космического эксперимента. Вместе с тем, многие проблемы, такие как скорость доставки данных потребителю, емкость средств хранения данных на современном уровне развития практически устранены, по крайней мере для объемов информации в проектах по солнечно-земной физике (за исключением изображений Солнца). Доступны также стандартные средства визуализации данных, что уменьшает объем работ по программированию. С другой стороны, количество исследователей не увеличилось пропорционально увеличению объемов данных, что привело к фактической конкуренции проектов за рабочую силу в части обработки, анализа и интерпретации данных. Большинство современных исследований являются комплексными и используют данные нескольких проектов одновременно.

В этой связи, современные архивы должны соответствовать следующим критериям, помимо стандартных требований к емкости и пр.:

1. Интерфейс должен быть достаточно простым для удобства выполнения относительно несложных запросов, так как они составляют абсолютное большинство обращений. Операции по доступу к данным должны поддаваться определенной автоматизации в случае формулирования повторяющихся запросов. В то же время должна быть возможна интеграция системы в более сложные сервисы, либо активация дополнительных возможностей для выполнения более сложных операций, таких как сравнение, интерполяция и предварительный отбор данных.

2. Формат и описание данных должны быть в максимальной степени стандартными или совместимыми для последующей интеграции архива в системы обмена данными (виртуальные обсерватории).

3. Производственные затраты на поддержание архива и его пополнение должны быть минимальными. Желательно максимально избегать необходимости модификации программного обеспечения при изменении состава данных.

В документах международного сообщества SPASE [1] перечислены основные направления будущего развития современных систем хранения данных. Предполагается, что такие системы должны иметь возможности:

- выполнять поиск с использованием физических единиц, то есть применять поиск по значению параметра к набору данных;
- иметь средства объединения сравнимых переменных из разных экспериментов в единый набор для дальнейшей передачи инструментам визуализации и анализа;
- данные должны быть зарегистрированы в общедоступных каталогах и, таким образом, входить в глобальную сеть данных физических экспериментов;
- для того чтобы в перспективе быть включенными в такой архив, данные должны быть снабжены метаданными.

Система хранения данных космического эксперимента ИКИ РАН

В настоящее время в ИКИ РАН ведутся работы по созданию системы хранения и отображения результатов космического эксперимента. Итогом этих работ должно стать получение единого

электронного архива результатов космического эксперимента, отвечающего современным требованиям к системам такого характера.

На данном этапе архив, созданный в ИКИ РАН включает в себя несколько подсистем:

1. Стандартный интерфейс к архивам обзорных графиков, заранее построенных в едином формате по всей продолжительности эксперимента. Такие наборы графиков являются стандартным продуктом и предназначены для быстрого просмотра данных.

2. Система доступа к данным с возможностью варьирования списка просматриваемых параметров и интервалов времени, а также выдачи данных в графическом и цифровом формате.

3. Система DD, разработанная ранее и позволяющая зарегистрированным пользователям большие возможности в определении форматов графиков и проведении различных операций с данными.

Подсистемы 1 и 2 находятся в свободном доступе, просматривать и получать данные может любой заинтересованный пользователь.

Базовое программное обеспечение

В результате предварительной проработки за основу системы доступа к данным (2) был выбран интерфейс CDAWeb, разработанный НАСА, так как он обеспечивает наиболее простую и удобную в эксплуатации форму пользовательского интерфейса, а программы и утилиты, обслуживающие этот интерфейс, распространяются свободно. Система основывается на формате CDF[2], широко используемом для хранения данных в солнечно-земной физике. Это автоматически обеспечивает совместимость архива с интерфейсами зарубежных центров данных.

CDF является самодокументированным форматом. Файлы содержат внутри себя всю необходимую справочную информацию, в том числе данные для построения графика. Он также обеспечивает сжатие данных. Утилиты CDAWeb распознают справочную информацию и автоматически определяют формат запросов на входе и графиков и текстовых файлов на выходе. Таким образом, добавление новых наборов данных в архив сводится только к перекодированию в формат CDF и исправлению в каталоге, что кардинально снижает затраты на архивацию данных.

CDAWeb включает полный набор утилит интерфейса пользователя, позволяющих представлять информацию в графической и текстовой форме по запросу, включающему время и тип параметра. Утилиты основаны на языке IDL.

Логическая организация данных

Самой крупной логической единицей деления архива является проект. В текущей версии используются данные проекта ИНТЕРБОЛ, в дальнейшем возможны подключения данных других проектов.

В директории каждого проекта содержатся данные его продуктов. Под продуктом понимается совокупность переменных, логически связанных между собой, описывающих одно физическое явление, имеющих одинаковое временное разрешение и, возможно, относящихся к одному прибору. Итоги работы одного прибора могут быть представлены в виде нескольких продуктов, но и один продукт может быть составлен из данных нескольких приборов.

К каждому продукту прилагаются метаданные, т.е. стандартизированная справочная информация. Сейчас метаданные составлены в соответствии со стандартом ISTP, в дальнейшем планируется провести адаптацию под разрабатываемые стандарты SPASE .

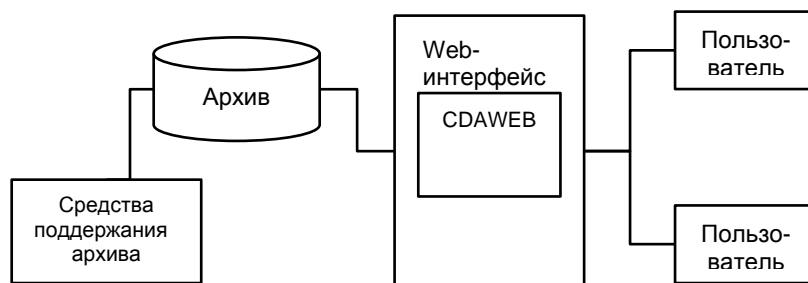


Рис. 1. Логическая структура архива.

Физическая организация данных

Особенностью системы хранения данных является отсутствие в ней средств традиционной СУБД. Это решение часто применяется в таких системах. Причины такого выбора следующие:

- ключи данных в классическом понимании - это время и название продукта, при этом вся информация о них содержится в структуре каталогов и именах файлов.

- к данным нет необходимости применять операции сортировки;
- в системе не используются сложные реляционные запросы;

То есть для функционирования системы достаточно, чтобы физическая структура архива соответствовала его логической структуре. Например, корневом каталоге архива должны лежать каталоги данных (DATA) и справки. Внутри каждого из них – подкаталоги, названные по именам проектов. В DATA находится дерево каталогов, каждая ветвь которого состоит из стартового каталога названного по имени продукта и внутри нее набор директорий, названных по годам измерений.

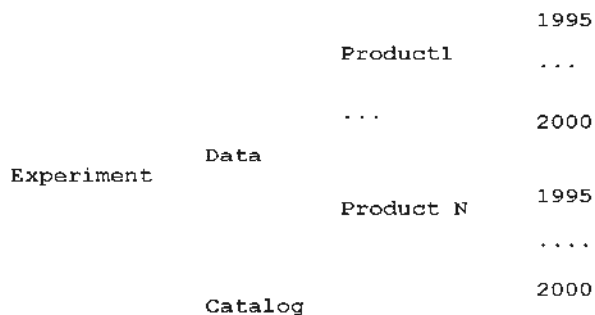


Рис. 2. Структура дерева директорий

В директории Catalog находятся справочные файлы для продуктов. Эти файлы в дальнейшем используются для динамической генерации WEB-страницы.

WEB-интерфейс

Для доступа к данным разработан WEB-интерфейс, доступный по адресу:
<http://stdad.iki.rssi.ru:8180/DataBag/>.

При входе на первую страницу пользователю предлагается выбрать, данные какого из научных космических проектов он предполагает использовать, а так же выбрать интересующий тип научных измерений (частицы, волны и т.п.).

Затем появляется предложение выбрать приборы, участвующие в выбранном проекте. Здесь же пользователь имеет возможность получить краткое описание приборов, узнать их научных руководителей и перейти на страницу с более подробными описаниями приборов.

И наконец пользователь должен выбрать интересующий его интервал измерений, вид измерений и желаемый способ получения результатов. Предусмотрены четыре варианта вывода результатов :

- график измерений;
- текстовый файл с результатами (или архив текстовых файлов). Предусмотрена возможность передачи данных для дальнейшей обработки на компьютер пользователя.
- получение исходных CDF-файлов с результатами измерений за выбранный период времени;
- создание файлов в формате CDF для выбранных режимов работы приборов и выбранного временного интервала.

Также на этом этапе автоматически запоминаются выбранные пользователем параметры, которые можно восстановить при следующем сеансе работы с системой на начальной странице.

Выводы

В целом, установка подобной комбинации систем хранения данных позволяет решить проблему архивирования и доступа к данным космического эксперимента по солнечно-земной тематике на современном уровне в условиях ограниченных ресурсов. В дальнейшем предполагается провести заполнение системы всеми данными прошлых экспериментов, находящимися сейчас в разрозненном виде. Так же будет необходимо провести работы по адаптации элементов системы к обмену данными в рамках виртуальных обсерваторий в части определения метаданных.

Литература:

1. A Space and Solar Physics Data Model from the Space Consortium. <http://www.spase-group.org>
2. The Common Data Format (CDF). <http://cdf.gsfc.nasa.gov>