Value added services with the SwRI data system

Joey MUKHERJEE(1), Sandee JEFFERS(1), Carrie GONZALEZ(1) and Andrew GALUS(1)

(1)Southwest Research Institute; 6220 Culebra Rd., Bldg. 178; San Antonio, TX 78238; USA

joey@swri.edu, sjeffers@swri.edu, cgonzalez@swri.edu, agalus@swri.edu

Abstract

Southwest Research Institute (SwRI) is constantly improving the services we offer to the general public in terms of data archiving and data visualization. First, we have standardized on the IDFS data format for our data which is modular, extensible, and offers many advantages in the area of value added services. Second, we offer a suite of utilities to analyze all data received from us. These utilities offer the ability to plot the data in a series of spectrograms, energy angle plots, and line plots as well as a multitude of three-dimensional plots. The data also need not be immediately available. All data will be transferred from the archive site as needed. Furthermore, once the data is in IDFS format, it can be exported into the Common Data Format (CDF), netCDF, as well as straight ASCII text files. Finally, through the World Wide Web, we are offering quick looks at our data as well as the ability to find data via keyword stored in one of our archives.

1. Introduction

The Data Analysis and Science Support Section of Southwest Research Institute (SwRI) have been developing data formats and systems for ten years. We have developed our own data format currently being used with thirteen satellites and plans to use it with more. Also, we have developed an analysis system to visualize the data. This analysis system is available either on or off the World Wide Web. Furthermore, we are developing more searching and cataloging tools to make it easier for the end user to find the data they need. Each of these capabilities add value to the data making our system extremely useful for the long term preservation of data.

2. IDFS

IDFS stands for the Instrument Data File Set and is a binary representation of the raw telemetry of a particular instrument or experiment. IDFS is primarily geared toward space physics data and can be used for both science and engineering data. An advantage of IDFS is the fact that it is self documenting in that the documentation for the instrument is contained within the meta data. Since both the data format and the instrument are documented, future engineers can either use the data as is or reconstruct the data for their needs. Since IDFS stores raw data, no precision is lost due to the conversion of units. IDFS uses tables defined in the meta data to convert the raw data into physical units.

Another advantage of IDFS is its modular format. Every instrument is divided into smaller related fields which are referred to as virtual instruments. Virtual instruments are generally divided by sensor groups which keep related sensors together and can make it easier on the plotting or analysis software to plot all the sensors as a group. Physically, each virtual instrument consists of a minimum of three files. The first file is a VIDF, or virtual instrument description file. This file contains all the documentation for a particular virtual instrument as well as the tables and constants necessary to convert the data into science units. The other two files occur as a pair and are referred to as the header and data file. The header file contains certain information that is slowly varying in time while the data file contains all other data. Since all files are time tagged with a start and end time they may be duplicated as needed when files get too large for efficient data management.

IDFS has another advantage in that it is extremely extensible. As the instruments get more complex, the scientists' needs change. Thus, IDFS must be able to adapt with the times. The VIDF file which initially was a fixed format file, and still may be, can now be added to as needs change without sacrificing backward compatibility. Programs and data sets which can utilize the new fields are free to do so, while programs that cannot will simply ignore them. This idea is very similar to HTML in which tags and values can be added carefully without causing older programs any harm.

3. SDDAS

SDDAS, or the Southwest Data Display and Analysis System, is an open source software distribution used for analyzing IDFS data and offers many methods for analysis. These include spectrograms, line plots, three dimensional orbit plots, contour maps in both two and three dimensions, XY plots, and energy angle plots. All three dimensional plots we create offer the scientist the ability to rotate and zoom the visualization in order to see all of the data from all look directions. Using specially made 3D glasses from Crystal Eyes, the visualization is enhanced even further. The plots will appear to be floating out of the screen in true 3D perspective. In addition to offering the visualizations, SDDAS has a framework by which other groups can add software to SDDAS.

For cataloging data, SDDAS uses either one of two databases. On the client side, SwRI has developed a database very similar to the old dBase III format. This particular database is tied strictly to time series data. However, on the server side, we use a relational database called MySQL. Having a relational database allows us to make relational queries on the data. A relational query allows one to combine the results of multiple queries and do either a union, intersection, or difference of the queries. One application of a relational query is to show the intersection of two distinct satellite datasets (for instance: UARS and DMSP) and show regions where the two satellites are both functional. We plan to move the MySQL database to the client side as MySQL becomes easier to install.

Data is kept on remote servers around the world and through a concept known as *promotion*, the data is retrieved. However, only the data for the certain period that has been requested is brought over. This allows an institution to archive data for the entire mission, and have it accessible at any time by a requestor. Data promotion is built over the common internet protocol, HTTP, or Hypertext Transfer Protocol. Using standard mechanisms allows us and others to easily script interfaces to our archive sites. Furthermore, data is not the only information available via promotion. All of SDDAS can be updated using our promotion concept.

Finally, SDDAS is able to export data from IDFS. Some users and other software may prefer a different data format other than IDFS, so SDDAS is able to accommodate their needs by exporting subsets of IDFS into ASCII, CDF (Common Data Format), netCDF, and XML data formats. By going to an XML format, a transformation file (XSLT, extensible style sheet) can be used to further convert the XML data set to another data format specified by the style sheet. As an example, a style sheet to convert the XML to HTML is included with SDDAS.

As one can see, SDDAS offers many options for the user with IDFS data. The downside of this is having to install this software on their computer, although not difficult, some users are reluctant to do this. Furthermore, learning any new software is time consuming. To alleviate this problem, a website for one particular satellite, Cluster, has been developed.

4. Cluster Web Site

The interactive web site for Cluster data analysis was initially brought on by our work with the PEACE (Plasma Electron and Current Experiment) data set. Many users were overwhelmed by the power and complexity of SDDAS so a way was needed to simplify the interfaces such that anyone could quickly get a plot of Cluster data. The easiest way to do that was to make some predefined plots where the only thing the user would have to specify is the start time and how long they wanted to plot. We met and exceeded these goals by offering a substantial portion of SDDAS available from the web. Users are able to plot both spectrograms and three dimensional contour plots of Cluster data; see what data is available; download the data in ASCII, CDF, or netCDF, all from the web page. Since the only parameters the user must define are the time ranges, the interfaces are incredibly simple. Documentation is available and searchable for all the virtual instruments as well so a user can find out exactly which instrument is appropriate for what they might need.

The web site is further enhanced through the use of Java or JavaScript. A spectrogram is plotted and then becomes an interactive widget in which a region can be highlighted such that the highlighted region is expanded out to the full length of the plot. In effect, the plot is zoomed in like a magnifying glass for a specific region. Furthermore, plots can be paged to go either forward or backward in time letting a user go through the entirety of the data in any time-spatial resolution.

Also with the Java enabled plot, users are also able to zoom out similar to the way one zooms in. A proportional amount of time is added back to the time ranges expanding out the amount of time plotted. A full history is kept of all the plots created so one can page between all the plots generated. If a user wishes to save a plot, they may either use their browser to save the plot or download the same plot in postscript format from the site.

Since certain aspects of the Cluster data are sensitive, our web site offers security in the form of usernames and passwords. Administrator users are granted the authority to add and remove other users and give them certain rights over particular areas of the site. For instance, downloading the data is one such aspect which is protected by username and password. Furthermore, to the regular user only two instruments are available from the web page. If a user is granted access and signs on to the site, the full list of instruments and plots are available. Statistics are also generated for the site describing what people do when they visit the site allowing the principal investigator (PI) the ability to see how their data is being used. Finally, all plots are watermarked with a "Not for Publication" banner to ensure that the rights of the PI are not violated.

5. Other Value Added Services

As demonstrated, merely providing data is not enough. Users need a way to look at and analyze the data. However, is that even enough? At SwRI, we feel there is more to the value added services of a data center so we are trying to develop novel and new methods to add value to our archive sites.

The first way to make a better data center is to archive the science as well as the data. Any papers that are published based on Cluster data are requested to be sent to the site where we will archive them for long term storage. These papers, which should be submitted as PDF, are then archived and indexed so that they may be searched. Scientists of the future can see all the papers published on a specific instrument or experiment and reference them with ease. Also, news of the mission is also kept archived on the site. At any time, the PI or other select few are able to add information about the site or mission to keep the site fresh and the content current.

Our long term approach is two fold. First, we standardized on a particular format so people can be assured that any data received from us is going to be usable in the same way. Secondly, any development we might do for one instrument, all other instruments and datasets will benefit. Furthermore, the extensibility of our data systems is one of its greatest strengths. Although complex, we have not run into any dataset which we can not make use of in some way.

6. Future Developments

For the future, we are working on all areas of the system. On the data side, we are redeveloping our imaging software to make better use of coordinate transforms to better render all types of images. SDDAS is always in development and incorporating new databasing and visualization techniques, as well as making it easier to use. The work on the web site continues at a frantic pace as well. In addition to the searching and archiving capabilities, we are adding several administration level tools to track what visitors are doing when they visit to the web site. Exporting XML via the web is another future goal. In addition to adding a straight XML download, we wish to offer the ability to apply a style sheet and have the data converted to even more formats right from our server.

7. References

1. Mukherjee, J., Kalla, L., Chilson, C., Jeffers, S., Gonzalez, C., *Archival of Time-Series Data at SwRI, MSSL, and IRF*, Cospar, 2002.
2. Gonzalez, C., *IDFS Programmers Manual*, Southwest Research Institute, San Antonio, TX, 1985.
3. Mukherjee, J., Gonzalez, C., Jeffers, S., *SDDAS Documentation*, Southwest Research Institute, San Antonio, TX, 1998.
4. TcX, *MySQL Reference Manual*, TcX AB, Detron HB, and Monty Program KB, Sweden, 1997.
5. IDFS Web Site, http://www.idfs.org
6. SDDAS Web Site, http://www.sddas.org
7. Cluster Data Analysis Website, http://cluster2.space.swri.edu