P1.25 ANATOMY OF TRMM SCIENCE DATA SUPPORT AT THE GES DAAC

William Teng *, George Serafino, Long Chiu, Hualan Rui, John Bonk, Pat Hrubiak,
Runhua Yang, Suraiya Ahmad, Zhong Liu, Bing Zhang, Nathan Pollack,
Gilberto Vicente, A.K. Sharma, Li Lu

GSFC Earth Sciences (GES) Data and Information Services Center (DISC)
Distributed Active Archive Center (DAAC)
NASA/Goddard Space Flight Center, Greenbelt Maryland, USA 20771

1. INTRODUCTION

Successful science data support for Earth observing satellite missions and their data systems requires a large, well-coordinated set of activities that spans more than the lifetime of the missions. The Distributed Active Archive Center (DAAC) at the GSFC Earth Sciences Data and Information Services Center has been providing such science data support for numerous satellite missions since 1993 (e.g., TOMS, SeaWiFS, TRMM, MODIS). These cradle-to-grave data support activities include user surveys, profiling, and requirements gathering; data ingest, validation, and archive; documentation and informational Web sites; search and order and visualization; read software and other tools; user support; value-added or customized products and services; and outreach. Support activities for MODIS and other Earth Observing System (EOS) data sets also include science software integration and testing (SSI&T support) and troubleshooting user problems related to data production systems. The goal is to enable users to fully realize the scientific, educational, and application potential of DAAC data.

The DAAC's Hydrology Data Support Team (DST) has been helping users of Tropical Rainfall Measuring Mission (TRMM) data in achieving this goal since before the launch of TRMM on November 28, 1997. TRMM is a joint mission of the National Aeronautics and Space Administration (NASA) and the National Space Development Agency (NASDA) of Japan to monitor and study tropical and subtropical rainfall systems (-----, 2000). TRMM data and information are accessible via http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/hydrology/hd_main.html. This paper summarizes the main components—the anatomy—of TRMM science data support provided by the DAAC.

* *Corresponding author address:* William Teng, NASA/GSFC Earth Sciences DAAC (SSAI), Code 902.2, Greenbelt, MD 20771; e-mail: teng@daac.gsfc.nasa.gov.

2. PRE-LAUNCH PREPARATION

For over a year before the launch of TRMM, the DAAC had been laying the groundwork for supporting the mission. The two main components of the groundwork were (1) the TRMM Support System (TSS), an information system that manages the large volume of data received from the TRMM data producer, TRMM Science Data and Information System (TSDIS), and distributes the data to users and (2) an aggressive outreach program. Fig. 1 shows the DAAC's TRMM Data Archive and Distribution system architecture, which includes automated transfer, ingest, archive, and distribution functions.

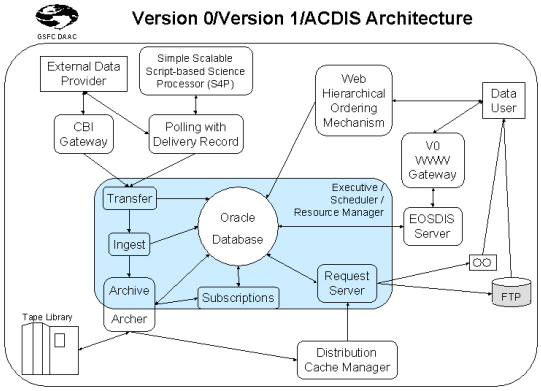

Fig. 1. Overview of GES DAAC Architecture (courtesy of C. Lynnes)

During the development of the TSS, the DAAC's Hydrology DST interacted extensively with the then potential user community to derive requirements (e.g., for alternative data formats, product preferences). These requirements were input to the TSS development process; in effect, the users' interests were represented in the process, thus ensuring that the TSS would be built to accommodate a variety of user needs.

The requirements were gathered, as part of the pre-launch outreach program, using various means, including a TRMM User Survey that was mailed out to potential users, distributed at conferences, as well as accessible via the Web,

http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/hydrology/hd_survey_intro.html. The AGU Fall 1996 and Spring 1997, and the AMS Annual 1997 meetings were the major pre-launch meetings at which the DAAC publicized TRMM and its data products. Discussions with users at these meetings were invaluable to the DAAC's preparation for TRMM support.

3. ENSURING DATA FLOW

The basis of any science data support is the unimpeded (as much as possible) inflow of data from the data producer and outflow to the users. The TRMM value-added user services, described in Sec. 4 and 5, would be largely ineffectual without a reliable data flow. During the TRMM pre-launch and immediate post-launch periods, the focus was on data inflow (i.e., ingest software development, metadata validation, database management). During the post-launch checkout of instruments and data (by the TRMM Science Team) and after the data inflow has stabilized, the DAAC's focus gradually shifted to ensuring data outflow. The Initial public release of TRMM standard products was on June 1, 1998, via the TRMM Web Search and Order System ("Web Hierarchical Ordering Mechanism" in Fig. 1, http://lake.nascom.nasa.gov/data/dataset/TRMM/index.html). This system allowed users to search for data hierarchically or by specifying space, time, and attribute criteria. Some TRMM data were also available via anonymous ftp, ftp://lake.nascom.nasa.gov/data/TRMM/, and the Earth Observing System Data Gateway, http://redhook.gsfc.nasa.gov/~imswww/pub/imswelcome/.

There have been periodic events that have affected the flow of TRMM data. Improvements to the TRMM algorithms were addressed at scheduled reprocessing intervals by the TRMM Science Team and implemented by TSDIS in major reprocessings of TRMM data, starting from launch. The DAAC has supported two such major reprocessings, each several months long, beginning on October 1998 and November 1999, respectively. The average operating altitude for the TRMM satellite was moved from 350 km to 403 km during the period from August 7 to 24, 2001, which will significantly extend the mission lifetime for TRMM. For all these events, the DAAC had to make corresponding modifications to its database, ingest software, other software tools, and documentation. In all cases, the DAAC sought to minimize any effects on users by careful and thorough preparations and timely notifications to users.

4. VALUE-ADDED PRODUCTS AND SERVICES

Although an unimpeded data flow (Sec. 3) formed the basis of the DAAC's TRMM data support, it would have been largely unused by the general users, were the flow not directed at actual user needs, helping to solve user problems. Towards this latter goal, the DAAC has committed significant resources to developing value-added products and services, aimed at these needs.

4.1 *Products*

Through extensive and continual interactions with the user community, the DAAC has developed a suite of value-added products (subset, resampled, reformatted), derived from the TRMM standard products. The latter are in Hierarchical Data Format (HDF). These value-added products (Table 1) support TRMM research and facilitate analyses and processing by users. The main purposes of these products are to (1) reduce the file sizes that users have to handle (from hundreds of Megabytes to a few Megabytes), which also facilitates access via FTP; (2) in some cases, convert from HDF to formats more familiar to users; and (3) satisfy the needs of interdisciplinary and applications users. These products will evolve as users' requirements change. They can be accessed from the URL's given in Sec. 3.

Table 1. Value-added data products derived from TRMM standard products.

Value-Added Product	Description	Format
Parameter subsets	Surface rain and related parameters	HDF
Gridded subsets	Re-sampled from orbital data onto Earth grids	Binary
Geographical region subsets	Spatial subsets of Gridded data	Binary
Satellite coincidence subsets	Instrument scans corresponding to coincidence[1]	HDF
Ground validation coincidence subsets	Radar volume scans corresponding to coincidence[1]	HDF
GIS-compatible subsets	Converted from selected TRMM data products	ARC/INFO interchange, shapefile

[1]Coincidence is between sub-satellite point and ground validation radar site.

4.2 *Services*

To facilitate and promote the use of TRMM data, the DAAC has developed or made available a number of tools and other value-added services for users. All readmes accompanying distributed data contain simple read programs that allow users to quickly open and read DAAC data. The DAAC has developed a generic HDF read software that can read all TRMM standard products and output individual data layers in binary. This and other software tools can be accessed via http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/hydrology/hd_software.html.

As data volumes increase, the proportion of data that can be distributed to users decreases. One mitigating service the DAAC has provided is to migrate more data mining and data reduction activities from the user to the DAAC, thus reducing the data volume that needs to be distributed. The user receives, as a result, more useful, manageable, and information-rich products, thus saving significant amount of time that would otherwise have been spent in data preprocessing and management activities (Lynnes and Mack, 2001).

Another method to increase the information content of distributed data is the DAAC's Remote Sensing Information Partners (RSIP) program. It provides participating institutions (RSIPs) (1) easy and inexpensive access to DAAC science data holdings and value-added products and services and (2) the ability to routinely acquire these data for local storage, use and distribution (Qu et al., 2001). The RSIPs serve as secondary data distribution sites that 'own' the data for their own research and application use and serve their local user communities. As a result, the end users obtain data in a more useful form for their purposes; at the same time, the distribution load of the DAAC is reduced.

To broaden the TRMM user base and to expand the potential types of applications of TRMM data, the DAAC has been engaged in several efforts to bridge the gap between TRMM data and Geographical Information System (GIS) tools (Pollack et al., 2001). The first is the automated conversion of selected TRMM data into GIS-compatible formats. The second is the creation of a WebGIS prototype, which allows users to combine TRMM data with ancillary data to perform mapping and analysis online. The third effort is making TRMM data available in formats that are compliant with current, developing interoperability standards.

4.3 *Field Experiments Support*

A special category of value-added data support that the DAAC has provided for TRMM was that for field experiments carried out for TRMM algorithm validation. The DAAC is the central archive for three of these experiments, TEFLUN-A and -B, TRMM-LBA, and KWAJEX; and the ancillary data archive for a fourth, SCSMEX. To expedite the availability of field experiment data to the research community, the data are archived (size permitting) in their native formats on anonymous ftp and front-ended by web sites, and accompanied by documentation. Field experiment data access is via http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/TRMM_FE/index.shtml.

The DAAC has also made available ancillary data products, principally temporal and spatial subsets of satellite data, for each TRMM field experiment and made available to users via ftp and web sites. Ancillary data access is via http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/TRMM_FE/trmm_anc.shtml.

5. USER SUPPORT

5.1 *Documentation and Help Desk*

Several forms of documentation are available to help users. All DAAC data products are distributed with an accompanying readme, containing the following information: summary, data description, sample software, references, data access, and points of contact. The DAAC's Hydrology Web site, at http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/hydrology/hd_main.html, contains information about TRMM and other hydrological data. Summaries of all TRMM standard data products, in the form of DIFs (Data Interchange Format), have also been created and submitted to the Global Change Master Directory, accessible via http://gcmd.gsfc.nasa.gov/.

Users have multiple ways to contact the DAAC to ask questions. All questions are stored and tracked in a User Assistance System, which allows the DAAC staff to readily access the history of a particular user query and to coordinate the response to the user.

5.2 *Outreach*

The DAAC conducts TRMM outreach as service; i.e., the purpose is not to just push data to users, but to make available potential solutions to users' problems. This outreach is conducted by making information about TRMM data and their potential uses easily available to users (e.g., publications, informational brochures, newsletters, exhibit

displays at conferences) and by engaging users to solve their problems (e.g., user surveys, site visits, seminars, conferences, workshops, partnerships). Since from about one year before the launch of TRMM, the DAAC has conducted outreach (presenting poster or oral papers and exhibit displays on TRMM) in some 20 conferences, including AGU, AMS, ASPRS, IGARSS, IUGG, and PIERS.

5.3 *Metrics*

Another important aspect of user support is the continual gathering of user requirements, via usage statistics. The DAAC regularly collects and updates TRMM usage statistics (e.g., Fig, 2), which are used not only to better understand user behavior and needs, and refine user support strategy accordingly, but also as useful inputs to the design of future rain measuring satellite data support systems, such as the TRMM follow-on mission, the proposed Global Precipitation Mission (GPM). See Rui et al. (2002) for more details on TRMM statistics.

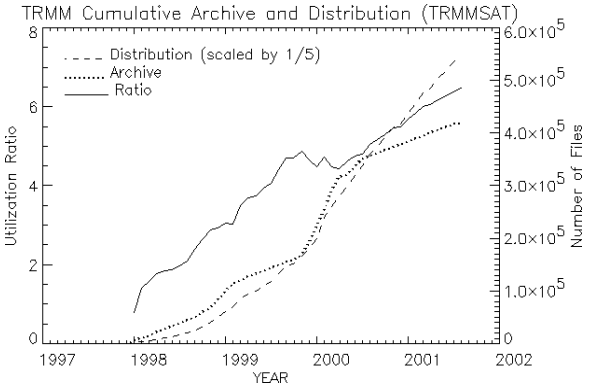

Fig. 2. Cumulative statistics of all TRMM standard satellite products, as of August 2001.

6. CONCLUSIONS

At all times, the DAAC's TRMM support remains responsive to user needs, accommodating to unanticipated demands, and innovative in availing the users of the latest appropriate technology. Through this support, in the four years of TRMM thus far, the DAAC has largely achieved the goal of enabling an increasing number and variety of users to fully benefit from the use of TRMM data in solving problems.

7. REFERENCES

Lynnes, C. and R. Mack, 2001: An approach to data center-based KDD of remote sensing datasets. In First SIAM Int'l. Conf. Data Mining, Apr. 5-7, 2001, Chicago, IL.

Pollack, N., W. Teng, and G. Serafino, 2001: Enhancing GDAAC data access and usage with a WEBGIS. In Proc. 2001 ASPRS Annual Conf., Apr. 23-27, 2001, St. Louis, MO.

Qu, J., G. Serafino, M. Kafatos, R. Lathrop, and J. Trimble, 2001: New application of Earth science remote sensing data at NASA/GES DISC: Remote Sensing Information Partner (RSIP) with Rutgers University. In Proc. IEEE 2001 Int'l. Geosci. Remote Sens. Symp., July 9-13, 2001, Sydney.

Rui, H., B. Teng, L. Chiu, G. Serafino, P. Hrubiak, and J. Bonk, 2002: Statistics of TRMM data archive and distribution at the Goddard DAAC. In Proc. 18th Int'l. Conf. Interactive Info. And Processing Sys. (IIPS) for Meteor., Oceanography, and Hydrology, Jan. 13-17, 2002, Orlando, FL.

-----, 2000: Special issue on the Tropical Rainfall Measuring Mission (TRMM). Combined pub. 1 Dec. 2000 J. Climate, **13**, 4069-4228 and Pt. 1 Dec. 2000 J. Appl. Meteor., **39**, 1961-2301.