NSSTC PRELIMINARY REVIEW OF REPLACING THE POLAR, WIND, AND GEOTAIL SPACECRAFT OPERATIONS AND GROUND SYSTEM DATA HANDLING

December 19, 2001

Introduction and Overview

The following discussion is the result of a time-limited analysis of the existing ISTP spacecraft operations and ground system data handling facilities located at NASA's Goddard Space Flight Center. The analysis has been performed in the context of seeking a more cost effective configuration and operation of the functions performed by the current facility, while ensuring successful operations and minimizing any impacts to science productivity. Detailed material and labor costs and a detailed timeline for facility development and transition can be provided following a site visit and close inspection of the current ISTP facilities and their operation.

Overall, we believe the NSSTC team has the skills and experience necessary to design, implement, and operate a ground data system capable of handling the Polar, Wind, and Geotail Mission requirements as discussed in the document "REQUIREMENTS FOR POLAR, WIND AND GEOTAIL SPACECRAFT OPERATIONS AND GROUND SYSTEM DATA HANDLING", dated September 20, 2001. Any change from the existing system will not be easy and we suggest will take longer than one year to fully implement. Although we are not yet prepared to provide a detailed description of a new facility or transition to it, we can provide our conceptual approach to the problem based on what we have learned so far.

Relative to a cost estimate, we suggest approximately 3 programmer man-years (some full time and some part-time) and system support would be required over a minimum of one year to develop a new data system facility. It would take a minimum of two Unix machines and two VMS machines. One Oracle license for the Unix machines could be used and one Oracle license for the VMS machines. The hope is that it may be possible to work with just two Oracle licenses and not four. The hardware also needs to include the data storage units and hardware necessary to produce data products on CDROM. We may only need the Oracle license on the Unix machines for the first year. There is a chance that the Unix level-zero processing software could be rewritten to run without Oracle, as described below. There is less chance of eliminating Oracle on the VMS machines, although that would be studied. This is also discussed further below. This machine and manpower description is based on the possible necessity to execute all existing code here at the NSSTC in order to sufficiently understand the code and its operation in order to make the changes suggested below. This development environment would be transitioned into an operational environment as part of the initial phase of a transition plan to be developed later.

Once developed, operation of the system would be similar to other efforts carried out by our team with a staff of 2-3 full time people. A staff this small would be dependent on successfully automating nearly all of the functions performed in the new facility. We believe that can be accomplished, but we cannot at this time make the necessary detailed trade-offs between programming effort, operational staff, new coding and scripting, and reuse of existing code or hardware. It looks like all of the CDHF software is front-ended with user / operator interfaces written in TAE. A lot of that would need to be replaced or eliminated.

IT security is a major issue at NASA and a detailed proposal would require particular attention to system, network, and firewall configurations. The NSSTC building is currently wired with 100base-T network service, which is sectioned into open Internet, UAH, and NASA networks. Firewalls provide interfaces between each internal network. A flexible wiring closet architecture makes it possible to selectively connect any room outlet to any of these three networks. Our current concept involves the placement of all processing and storage facilities on the internal NASA network, behind a strictly controlled firewall. External access would be made possible by use of one or two Unix machines placed on the open network, which are granted selective read-only access to archived data products, located on machines within the protected NASA network. Our current expectation is that redundant storage of Internet-served data will not be necessary. Instead, the Internet-based machines will serve as portals for retrieving files. Web-folders are an alternative to FTP access already in use on machines with public access at NASA/MSFC. This approach is currently considered the most secure way to serve directories and files of public data to the Internet. Physical access to the NSSTC is controlled through a reception desk and key-card accessed external doors. Additional physical security is provided on a room or area basis through additional key-card access doors. Access to the building and to individually secured areas can be customized to project requirements.

The National Space Science & Technology (NSSTC) team includes the University of Alabama in Huntsville (UAH) Information Technology Research Center (ITRC), NASA Marshall Space Flight Center (MSFC), and professional programmers under contract with Computer Sciences Corporation (CSC). Each organization has particular strengths applicable to the development and operation of a new data handling facility for the Polar, Wind, and Geotail spacecraft. The ITRC is well experienced in the development, management and operation of a high capacity data archive and distribution facility for NASA missions. CSC personnel have extensive experience with the existing ISTP facilities, through their work with the Polar TIDE and UVI instruments. This includes the development of key parameter (KP) data products and the user interfaces to near real-time (NRT) and quick look (QL) data. Their experience includes extensive efforts involving NASA satellite telemetry data from past orbital and suborbital missions. In the context of the development and operation of a new ISTP IT facility at the NSSTC, MSFC researchers in solar and magnetospheric physics would be responsible for providing science oversight and user validation of a newly developed facility and its operation. Should a formal proposal be requested from the NSSTC team, it will be lead by UAH/ITRC in the same close partnership that characterizes the NSSTC. The NSSTC

was formed just over a year ago. As initially formed it is a partnership between NASA and the six PhD granting universities in the State of Alabama. The NSSTC continues to grow, however, with requests to join from other universities in the southeastern United States. The NSSTC cooperative partnership is bringing together a wide range of research and technology skills, with world-class facilities located in Huntsville, Alabama on the UAH campus.

The following discussion is organized in a similar manner as the Mission data system requirements document, referenced above. In the spirit of an informal document, we offer our observations and opinions in addition to describing the capabilities of our team.

1. **Science Operations Planning**
    a. Event Identification – To a large extent, events are readily identified through geophysical indices, such as Dst, and through observations of earthward directed coronal mass ejections. Events can be of particular importance when multiple satellite observations are available. If centralized event identification and archival continues, we envision it following the historical method, with the exception that our team has developed "data mining" tools that may facilitate significant automated recognition of events. At this time we do not recommend continuing event identification and archive of event data.
    b. Predictive Orbit Information - FDF already produces orbit and attitude data. We would receive, process, and archive the orbit, attitude, and platform data. There is a web site that is sorely out-of-date that is used to keep track of events. Should that be continued and become our responsibility, we would open a new site at the NSSTC to replace the old one. The commanding will remain with operations. They currently generate files and are now looking at which ones, if any, need to be kept in the archive.
    c. Polar Despun Platform Pointing Planning – FDF and SPOF create pointing data, which we consider necessary to continue as currently performed. We would archive and make this available, along with other data system products.
    d. Submission of Commanding Sequences by Instrument Teams – We suggest moving this activity from the SPOF to the MOC. The command sequences can flow through and be archived in a new ground data system.

2. **Flight Operations Planning and Flight Operations**

    We believe that the existing spacecraft engineers and key operators have gained such a degree of training and experience in the operations of these spacecraft that it would be unnecessarily risky to make any change in basic operations. Consequently, it makes no sense for us to attempt to reproduce this function at the NSSTC and little sense that those functions be performed elsewhere. However, the contractual execution of flight operations planning and flight operations is not currently well managed

relative to cost. We are prepared to offer project management of those functions by directly contracting with the current personnel parent companies or to contract through a new organization selected to perform these functions.

We believe the current spacecraft engineers and at least some key operators should continue to operate the spacecraft with this function remaining at GSFC. The management of those personnel and the viability of existing hardware, applications, and procedures should be reviewed and changed as found necessary.

DSN schedules and spacecraft status reports are currently on a web site and are made available to users on the CDHF via instrumentor user accounts. We would most likely get rid of the user accounts and provide these files via the data archive.

3. Data Processing, Archiving, and Distribution

a. **Level Zero Processing –** NASA/MSFC would be responsible, through contract to Computer Science Corporation, to modify and re-apply all aspects of LZ processing. This includes QL and LZ data product production. It can also include NRT data. We suggest that the handling of telemetry data in all three forms can be greatly simplified over the current processing. The same basic computer code is used to synchronize and parse the TDM telemetry stream into LZ component elements by major frame. This code can be reutilized in a new system architecture.

The differences in operating on NRT or tape dump data sources relate to the technical method of acquiring and delivering the data. The telemetry for NRT data is obtained via NISN RPC connections between running programs. The LZ NRT products are saved to disk and again passed on in real-time through NISN and the Internet to multiple destinations. Tape dump files from JPL require the same processing and disk destination for LZ elements. Regardless of the source, LZ elements on disk are essentially QL data, requiring little or no additional processing. The ultimate LZ daily product files are based on selecting from the LZ elements accumulated for each day. That selection process can be conducted, as LZ elements are accumulated or just prior to assembling LZ daily product files.

While the core telemetry parsing software can be reutilized in a streamlined architecture, the handling of LZ elements must be replaced, if significant cost savings are to be realized. We believe that the presentation of NRT, QL, and LZ data products to the research community can be entirely automated. Instrument teams need access to the NRT data whenever they are doing real-time commanding and whenever there is

something wrong with the spacecraft. Someone will have to be on call to get NRT data running if for some reason it is not being created when it should be or needs to be.

We realize that the past level of fidelity provided by the ISTP data system is not required, however an automated assembly of LZ daily product files can provide a relatively high level of completeness so long as "markers" are available or can be identified, which quantify the quality of each LZ data element.

It is undoubtedly necessary to describe what we mean by "LZ data element". Instrument and satellite telemetry data are acquired in the form of minor frames that compose major telemetry frames. Each LZ daily product file consists of a series of major frames assembled for a given day. A LZ data element is the smallest component of a LZ daily product file, which is to be handled by the ground data system as a unit. Information for any single time may be represented in several LZ data elements, originating from NRT data or spacecraft tape dump download. A given time period may be reproduced in more than one file provided by JPL for download to the ground data system. The accumulation of LZ data elements for a given day, therefore, may provide the opportunity to selectively assemble a day's worth of data, based on the quality of each element. We suggest that this process be automated, based on the best available quantities that can be used to identify the quality of LZ data elements. We anticipate that the association of data quality may be best done on major telemetry frame basis.

SOHO, Fast, and other sources of ancillary data may or may not be practical to integrate into a new data system for distribution to mission research teams. This aspect of the former system has not been studied and appears to be of a lower priority.

ITRC personnel would perform operation of a new data system. As stated previously, it is anticipated that 2-3 people will be needed for routine operations. One of these people would be on-call at all hours, although the daily routine would be during normal business hours.

b. **Spacecraft Health and Safety Data Processing –** As discussed above, MSFC will be responsible for developing or adopting code to create this LZ data. The ITRC will be responsible for developing routines to perform any necessary database operations and to distribute LZ products to the NSSDC. Actually, this is true for all completed LZ data products, which would be immediately delivered to the NSSDC and archived for 90-days. It is recognized that health and safety are the highest priority products and must be processed and distributed quickly.

We would be responsible for making sure that the NRT are available for socket links and the QL data are processed and put on the archive as soon as possible after receipt. The NRT data needs to be available with in ~10 seconds of data capture, which includes links from the spacecraft to JPL to NSSTC to the Unix LZ software to the CDHF software to the instrument ground operations team. The QL data would need to be processed as soon as the playback is complete and the LZ instrument data products made available on the archive as soon as possible. Both of these processes can occur anytime of the day or night – seven days a week. The processes need to be automated but someone needs to be on call if the processing should stop for any reason.

c. **KP Generation –** We find that some KP data products are useful and some are not. Those products, which are not useful, should not be carried over into the new data system architecture. Those products that are useful to the research community should continue to be produced. This is another area that requires significant overhaul. A site visit is necessary for us to properly evaluate the steps taken in the existing system to produce KP data. Our initial impression is that it may be necessary to start the redevelopment process by purchasing two DEC/Alpha/VMS computers and run the same Oracle-based software previously used at GSFC. The automation of this code and/or the removal of dependence on Oracle is anticipated to be a time consuming process. Replacement of Oracle with a much-simplified approach may be possible, but would require the replacement of existing utility subroutines with alternative routines that perform the same function, but without Oracle. We have just learned that some form of NASA-wide license for Oracle has recently been negotiated, which may mean that the cost for retaining an Oracle application in a new data system would not be too costly. We see no reason why the generation of KP data products cannot proceed in an automated manner. It may even be possible to release the KP code from execution only on VAX/VMS platforms, however, that can only be determined during the course of closely examining the user and CDHF provided utility computer code. At a minimum, we believe considerable cost savings would be possible through procedure automation by scripting in VMS. The NSSTC team retains extensive VMS experience.

This is a point for discussion, but if an instrument team creates their own KP data, then they should be responsible for getting it to the NSSDC. We should not act as the conduit. The KP process should be automated as much as possible. There is no need for multiple shifts. The KP processing should start automatically as soon as the LZ processing finishes. Operators need to monitor and restart it if something goes wrong.

Some consideration has been given to running instrument team high-resolution software applications. In some cases this might be done as a replacement for KP data. For example, the TIDE KP data is not considered valuable by the team, however the high-resolution data is widely used. Although doubtful, it is not currently possible to judge whether it makes sense to take on this user software operation. High time resolution data in ISTP compliant CDF files is most suited for archive and access at the NSSDC through their CDAWeb Internet facility. Whether KP or high-resolution data, the NSSTC team has extensive experience producing ISTP compliant CDF format files.

d. **Definitive Orbit and Attitude Files –**Definitive orbit and attitude files are no longer produced. Once a month we would expect to process new 70-day orbit/attitude files and from them create a months worth of daily orbit and attitude files. We would create daily definitive platform pointing files. We believe the spin phase files are daily definitive files. These two files would need to be made available as soon as their input is received. We assume the software needed to perform the creation of all these files exists and is available.

e. **Ancillary Data Ingestion –** Given the lower priority currently assigned to ancillary data by the Polar, Wind, and Geotail missions, we have not yet evaluated whether it would be affordable to continue this function in a new data handling facility. A decision whether to perform this function would likely be made late in the process of developing a new data system.

f. **Data Archiving and Distribution –** ITRC can provide the development required to implement the data archive and distribution functions associated with the efforts described in sections 4a through 4e above.

ITRC is experienced with ingest, management, archive and distribution of large Earth Science data sets. Through the Global Hydrology Resource Center (GHRC), ITRC has provided a full compliment of data management functionality and independent IT research that has served Earth scientists and users of Earth science data since 1994. This experience directly translates to the Space Science environment, since the structure and handling of the data is similar. The GHRC does not directly deal with level zero (or lower) data streams but it does accept over a dozen level-1 (or higher) data products ingested automatically on a daily or even hourly basis with a volume of approximately 15 GB/day. Therefore, the size and complexity of data associated with the ingest of level zero products from Polar, Wind, and Geotail would not be a major risk factor.

The current GHRC data streams take different paths as determined by their source and level of service. For example, data from the Lightning Imaging Sensor (LIS) aboard the TRMM satellite is passed on to the

product generation system where quick look products are created and posted to a website. This involves software produced by the LIS science team and integrated into the GHRC. The process here appears to be very similar to that of the key parameter processing of the CDHF. In each case, science products are created automatically and posted to a website. In each case, software from the instrument team was integrated into the data processing system.

The GHRC contains several product generation systems that automatically generate data products, post quick look images to the web, and then archive the raw, and higher level data products to a DLT archive system. We propose to archive ISTP products for a 90-day period (first-in, first-out), implemented as on-line storage with FTP access in a directory structure familiar to ISTP science teams. In addition, special science events can be maintained on-line through the period of sustained operations or as constrained by costs. All generated products would be copied to the NSSDC at NASA/GSFC for permanent archive as those products become available.

The data distribution capabilities of the GHRC are also similar to those employed by the CDHF and IDDS. The GHRC distributes data in a variety of ways. Each month, on average, approximately 500 Mbytes of data are mailed to customers on CD, 8mm, or 4mm tape media. There are also three methods to get data electronically. On-line data may be directly downloaded from the FTP server. A user may request data from the archive. In that case, an operator will stage data to an FTP server, then the user will be notified to download. The third method is probably most applicable to ISTP. In this case a user may set up a subscription for data. This can take place in two ways. If the data volume is small, files can be pushed by FTP to the user as soon as the data appear. (This might be an option for the NRT level zero data.) Or, for larger volumes of data, a tape or CD can be created and mailed on a periodic basis – usually monthly.

The GHRC has on-line systems that allow users to search and order data products. The user may search by date and time and even specify a subset region for certain products. Automated scripts enter the inventory of data products into an Oracle database, which also holds the user information. Similar interfaces exist for the ISTP user community. Most data maintained by the GHRC is in HDF-EOS format. However there are many similarities to CDF. In fact HDF was modified to incorporate CDF features. In another similarity, the GHRC uses the Data Interchange Format (DIF) to register products with the Global Change Master Directory. ISTP uses DIFs to express the metadata in the SFDU.

Thus, there appear to be many similarities between existing GHRC data management systems and those of the CDHF. The experience of ITRC in

this area would be employed to analyze the current CDHF system in order to ascertain areas where efficiency can be improved. There are potentially several GHRC capabilities that could be substituted for current CDHF capabilities. This however, will require a close inspection of the current CDHF system. Regardless, ITRC has the database and systems experience to implement a system that will yield quality science data for the Space science community.

g. **Data Archiving –** ITRC would keep Level-0, key parameter, and special events data online for 90 days. All generated products would be copied to the NSSDC at NASA/GSFC as soon as they become available. Long-term archive and retrieval are considered by us to be best provided by the NSSDC at NASA/GSFC.