Senior Review 2006 of the National Space Science Data Center and the Heliophysics Data and Modeling Centers

June 21, 2006

Submitted to:
Richard R. Fisher, Director
Heliophysics Division
Science Mission Directorate

Submitted by:
Frederick Bruhweiler, Peter Fox, Joe Giacalone, Eric Kihn,
Paulett Liewer (Chair), Joachim Roeder, Haimin Wang

Table of Contents

1. Overview

1.1 Introduction

NASA's Science Mission Directorate (SMD) periodically conducts comparative peer reviews of Mission Operations and Data Analysis (MO&DA) programs to maximize the scientific return from these programs within finite resources. The acronym "MO&DA" encompasses operating missions, data analysis from current and past missions, and supporting science data processing and archive centers. NASA uses the findings from senior reviews to define an implementation strategy and give programmatic direction to the missions and projects concerned for the next two to four fiscal years.

This year, the National Space Science Data Center (NSSDC[1]) and three heliophysics data and modeling centers were reviewed separately from the operating Heliophysics missions; the three heliophysics centers were the Community Coordinated Modeling Center (CCMC[2]), the Solar Data Analysis Center (SDAC[3]), and the Space Physics Data Facility (SPDF[4]). This is the report from that Data and Modeling Center (D&MC) Senior Review, held in May 2006 at GSFC. The review focused on the relevance and importance of the functions and services of the centers to NASA and SMD's strategic research objectives. Performance factors included technical status, effectiveness and efficiency in conducting activities, community participation, and budget.

SDAC and NSSDC were previously reviewed with the Sun-Earth-Connection (SEC) missions in the 2001 SEC Senior Review. Both CCMC and SPDF have been created since 2001 and this is their first senior review; SPDF was essentially "split off" from the NSSDC in 2002.

1.2 Instructions to the Panel

All centers submitted proposals to SMD in response to a call issued by NASA HQ/SMD in March 2006; the call is available upon request. The four centers were asked to give an overview of their activities and services and also to define the "Level-1" requirements for the project, which previously did not exist in written form. They were asked to describe how these level-1 requirements are relevant to NASA's and SMD strategic goals and objectives. A work break down structure (WBS) and budget, organized around the level requirements was also requested.

The panel was instructed to evaluate the proposals of the four centers taking into account (1) the relevance and importance of the Level 1 requirements to NASA's strategic research objectives; (2) the proposed plans and performance measures for meeting those Level 1 requirements; and (3) the performance to date in conducting the project activities,

[1] http://nssdc.gsfc.nasa.gov/
[2] http://ccmc.gsfc.nasa.gov/
[3] http://umbra.nascom.nasa.gov/
[4] http://spdf.gsfc.nasa.gov/

including effectiveness and efficiency, community participation and feedback, education and public outreach, etc.

Based on these, the panel was asked to provide comments on the implementation strategy for the D&MC program for 2007 to 2010, which could include a mix of:

- continuation of the project activities "as currently baselined";
- continuation of the project activities with either enhancements ("optimal budgets") or reductions to the current baseline; or
- termination of project activities

1.3 Methodology

The senior review panel members had received the proposals prior to the meeting. Each proposal was assigned one primary and three secondary reviewers from the panel members. At the meeting, each center was allotted ~2 hours to make an oral presentation and lead a laboratory/web demonstration session with the senior review panel to show web sites and services offered. A question period followed the oral presentation.

The panel recognizes the rapidly evolving data environment, which includes a proliferation of both data-serving centers and virtual observatories (VxO's) that provide services and links to help obtain data, but do not serve data themselves. Many questions from the panel concerned the relationship between these centers and the other data providers and virtual observatories and whether there was duplication of some services. Three types of data-serving centers are now recognized: (1) active archives (AAs) which serve data from active missions, (2) "deep" permanent archives (such as NSSDC) which are responsible for long term preservation of data, and (3) "resident archives" (RAs) which serve data from terminated missions whose data is still actively sought. This later concept is relatively new and the relationships between the existing centers and the "resident archives" are evolving.

Following the presentations, panel member comments were collected by the primary reviewer for each center. The next morning, the primary reviewers presented draft assessments and findings to the panel for discussion. For each center, discussion of the assessment continued until a consensus was reached on the major findings.

Below, the findings are summarized. Some are the findings are directed not specifically to the reviewed centers, but more broadly, to NASA HQ and the scientific community.

1.4 Summary Findings

1.4.1 Community Coordinated Modeling Center (CCMC)

- The panel was impressed with what the CCMC has accomplished in a relatively short period of time: 24 models are available at present to be run on request (ROR) by outside users. The system seems to be highly used (~20 new users per month) and the panel was satisfied that CCMC will be able to continue to provide this service.

- The CCMC makes good use of its steering committee to decide which models are mature and useful enough to incorporate into the center. The center is proactive in going out to community, e.g. via sponsored workshops. The panel notes that presently, there are few solar physics models. The center might consider going to solar physics meetings such as SHINE to find out what solar models this community would like to see supported by the CCMC.

- There is good synergy between CCMC staff and model providers. The model provider benefits by having his/her models at the center because frequent use of a model over wide parameter ranges, as well as comparisons with other models, generally leads to improvements to the provider's model.

- The center nicely emphasizes model validation and robustness. This is highly valuable because similar codes should give similar answers and it is important that these comparative studies are performed. This is important for transitioning the models to operational space weather forecasting.

- The panel finds a lack of models in several areas such as solar physics and ionosphere-thermosphere-mesosphere (ITM) and believes that the community would benefit if CCMC received their optimal budget to expand their expertise and import more models in these areas.

- The project would benefit from more time spent on gauging the value of its science output, e.g., as reflected in peer-reviewed publications. Regular science nuggets supplied by users and posted on the CCMC website may also be a useful in gauging the center's science output.

- The panel expresses concern about future demands for data storage relating to data assimilation and ever-larger simulations. Discussions by the CCMC with the community on future plans for data storage may provide guidance on this issue. Presently, the plan is to only store the code and the input because the output can then be recreated.

1.4.2 National Space Science Data Center (NSSDC)

- The panel finds that NSSDC has very strong relevance and value, significance and importance to SMD by providing the deep archive for NASA space science data.

However, the panel is concerned about the lengthy process to move data into NSSDC for permanent archiving. We believe that the establishment of standards (by NASA and/or the community) for permanent data archive and transfer to the archive would minimize the need for extensive negotiations in the memorandum-of-understanding (MOU) process.

- The panel believes that the NSSDC could place stronger emphasis on data ingest through more negotiation of MOU's. An unusual amount of time seems to be devoted to the MOU process, however. NSSDC could explore ways, with the help of NASA and the community to streamline this process, such as involvement by NSSDC on issues relating to data products, documentation and archive resource planning and by working extensively to understand and define the emerging distributed archiving paradigm and the role of NSSDC with respect to Resident Archives (RAs) (proactivity).

- The panel finds strong compliance with data preservation standards (i.e. those of the National Archives and Records Administration - NARA) and notes the NSSDC's expanding role into photoproducts. However, the documentation generation and preservation packaging seems to be handled on a case-by-case basis. There is a need for a more comprehensive plan. Developing and distributing the MPGA (Multifile Package Generator and Analyzer) software that creates the metadata needed by the NSSDC archive information package (AIP) is a step in the right direction.

- The panel finds that NSSDC could take a leading role in the development of standards (performance metrics), best practices (e.g., security) and guidance for resident archives and would like to have the RA work funded in the base program.

- The panel endorses NSSDC's plan to export AIP generation software, including the MPGA to active and resident archives and other data providers to facilitate the delivery of useable data sets to NSSDC.

- One task area that is missing or under emphasized is that of data stewardship to support NSSDC's level 1 requirements (insuring long-term data usability). This includes activities such as data quality control, making of data products for datasets not available elsewhere, any necessary repackaging, etc. in order to maximize the return-on-investment for archived data.

- The panel is concerned about the rate of technology infusion (e.g. extensible Markup Language (XML) work seems to be lagging peers). Working more with others in the community, e.g. those developing resident archives should help expose staff to newer technologies.

- NSSDC's use of the Space Physics Archive Search & Exchange (SPASE) dictionary of data keywords, which was developed by the space physics community, is commendable. The panel is concerned, however, that they may be relying too heavily

on this metadata model. It has not yet reached "critical mass," and other metadata models may be adopted by the community in the future.

1.4.3 Solar Data Analysis Center (SDAC)

- The operation of SDAC is highly cost effective, and provides an extremely critical service to solar community. FY07-10 is a critical period for SDAC to be well supported, because of the new missions Solar-B, STEREO and SDO.

- SDAC's combination of data archiving with data analysis is an excellent concept. The solar physics community strong interacts with this facility. SDAC's support of the solar soft tree is highly valued.

- Linking in ground-based data sources via SDAC's portion of the Virtual Solar Observatory (VSO) to support mission data analysis is unique and needs to be encouraged.

- The panel believes SDAC could be more proactive in educating the solar physics community about data analysis tools available through the center via the solar soft tree. SDAC could also consider working with CCMC on supporting solar physics models such as coronal magnetic field extrapolations from vector magnetograms.

- The panel finds that SDACs expanded efforts in VSO development are very valuable to the community and that SDAC, and the community would greatly benefit from the proposed optimal funding to perform advanced VSO work.

1.4.4 Space Physics Data Facility (SPDF)

- The panel was generally impressed with the Virtual Space-Physics Observatory (VSPO) and its capabilities and accepts VSPOs incorporation into SPDF. VSPO, in fact, seems to make it easier to find and to access data that are otherwise served and accessed by the Coordinated Data Analysis Web (CDAWeb) interface. The panel believes that SPDF should drop the development of a CDAWeb follow-on and instead extend the VSPO interface to provide the full CDAWeb functionality. Changes to the CDAWeb would be better focused on those services supporting discipline specific virtual observatories (VxO's) in general.

- SPDF would benefit from more community involvement. SPDF provides middleware, and it is important to be connected to the users as well to the data. The community, as the user of their services, should drive the development.

- The panel finds that there is no need whatsoever for a "VxO center" and that such an endeavor would be contrary to the very idea of how VxOs work.

- The panel feels that there should be an "oversight committee" to guide the SPDF in critical choices, such as to which data sets to serve or to drop, or which services to

develop and offer. A review every 3-4 years is not enough and a "user group" may not have a broad enough perspective.

- The panel noted that some of the data sets served by CDAWeb are duplicates of data sets that are served from other sites. A prime example is the Cluster data, which are now becoming available from the European Space Agency (ESA) Cluster active archive, but there are other data sets as well. In addition, CDAWeb holds data that are known to be deficient (by admission of their creators and as labeled by SPDF.) The panel believes that the SPDF should make efforts to drop data sets that are otherwise available and accessible (and higher quality) via their originators.

- The OMNI database is widely used by the community and it is important to continue its support and development. Likewise, the orbit and query service is unique. Both are core services of the SPDF, which the panel believes should continue to be supported under any circumstances.

- In several places in the proposals, the panel found the work not to be part of SPDF's core mission. The panel believes that such work should be competed (see discussions of WBS 1.3 and 2.7).

- The panel is concerned that there is no clear data policy that delineates the role of the SPDF from that of NSSDC to help avoid duplication of efforts. In general, NSSDC should not have to duplicate the machinery to catalog (for VxOs) and serve data in the same fashion as the SPDF. We believe that there needs to be a policy that outlines the rules as to which data sets should be both permanently archived by NSSDC and at the same time be offered by SPDF. Perhaps a HQ-sponsored oversight committee would be able to delineate the roles.

- The fact that only 15% of the SPDF staff are scientists appears to be having a negative impact on SPDF's choices and its utility for the community. Many of the proposed projects seem to be driven by computer science (CS) desires for the newest software technologies rather than the actual need of the science community. This is clearly reflected by the semantics of the proposal, which is abuzz in CS words. The panel also notes that the centers that have a larger fraction of scientists on their staff (SDAC, CCMC) seem to achieve higher "user satisfaction" levels.

2. Evaluation of Centers

2.1 Community Coordinated Modeling Center (CCMC)

Project Summary:

The CCMC resulted from a cooperative effort of several federal agencies. The goal was to provide an unbiased center that can house a variety of models relevant to space weather, and that outside users can run. The CCMC provides a unique and valuable service to the scientific community by establishing the infrastructure and expertise to

allow users to perform state-of-the-art numerical simulations without having to work directly with the code developers. The CCMC is located at GSFC but also receives support from other agency stakeholders, such as the NSF and the U.S. Air Force. The CCMC provides the necessary hardware and expertise to run the models. Users request model runs with parameters tailored to their specifications, and CCMC personnel perform the runs. Web-based, user-friendly interfaces enable users to tailor specific initial and boundary conditions and simulation parameters to individual science problems of interest. A variety of plotting options are provided to help visualize and analyze the simulation results. The simulations are performed on a 200-CPU Beowulf cluster, housed at GSFC, and are run through a unique, easy-to-use, automated Run on Request (RoR) system with good turnaround time.

The CCMC does not perform any model development by itself, except for separate, grant funded efforts. However, it serves a number of other functions besides being a RoR facility. Most importantly the CCMC participates in, and facilitates verification and validation (V&V) efforts. Such independent (from the model developers) V&V is necessary if models are to be considered for operational use. Furthermore, the CCMC facilitates the transition of models into operations. In addition, the CCMC participates in educational efforts, such as providing modeling capabilities to summer schools.

The procedure for adding new models or numerical simulations is to submit a request through the CCMC web page. The CCMC steering committee, with input from the CCMC then decides if a model is mature enough and if there is sufficient community interest to warrant the model's ingestion by the CCMC. If the model is adopted by the CCMC, center personnel work with the model developer to have the model installed at the CCMC and set up for ROR. One of the duties of CCMC's steering committee (composed of inter-agency personnel and scientists) is to charter working groups to help the CCMC with its operations and to decide on the integration of new models in the CCMC. The CCMC holds bi-annual workshops at which model providers and users exchange their experiences with CCMC personnel.

CCMC's Level-1 requirements (from the proposal):

- Provide access to state-of-the-art space science models for the international science community.
- Assist in the development of advanced space science models.
- Perform impendent, unbiased, model evaluations for space weather forecasters and decision makers
- Support the transition to operations of space-weather models.
- Support education and public outreach.

Panel Assessment (strengths, weaknesses and other findings):

The panel was impressed with what the CCMC has accomplished in a relatively short period of time. Currently CCMC houses 24 separate simulation models covering a wide range of space-physics applications from the solar corona to the inner magnetosphere and upper ionosphere. They provide the framework for space-weather modeling (a very useful tool to perform rapid simulations of new solar phenomena that may affect Earth). The panel agreed that their activities are important for transitioning useful space-weather models to operations, one of their level 1 requirements.

The system seems to be highly used, as evidenced by recent RoR requests that indicate a large number of different users (~20 new users per month). The panel was satisfied that CCMC will be able to continue to provide this service even though the number of requested simulation runs and incoming new models is rapidly increasing.

The panel finds that there is good synergy between modelers and the CCMC staff. They have a few scientists with expertise in each of the models to provide assistance to the users and feedback to the modelers. The center nicely emphasizes model validation and robustness. This is highly valuable because similar codes should give similar answers and it is important that these comparative steps are performed. Also, the panel agreed that CCMC is proactive in advertising the operations of CCMC to the community and soliciting input. CCMC is a community-driven enterprise. The panel believes that the center could strengthen its relations with grassroots organizations, such as SHINE, that holds an annual workshop on topical science topics, including campaign events. These activities motivate the use of the many models housed at CCMC. The center may also serve as a repository for models that are developed (as promised) by LWS TR&T proposals (and others).

The allocation of the staff's time and their list of priorities seem reasonable. Much time is spent on the important steps of code integration, creating and maintaining user interfaces and visualization routines, and performing code validation and metrics.

The optimal scenario budget requests support of a staff member in the area of thermosphere/ionosphere, as a first choice, and a staff member in the area of solar and heliospheric physics as a second choice. The panel finds that the center would greatly benefit from this optimal-funding scenario.

One concern of the panel is that a policy regarding data storage has not yet been established. CCMC presenters indicated that the simulations are fast enough to simply be performed on demand and that storage is not an issue. However, the panel agreed that the data policy should be clearly stated. The panel finds that CCMC should continue planning for data assimilative models and how to handle the associated data streams.

The panel finds that the models housed at CCMC lack some important physical regimes and applications (e.g. new solar models, high-energy particle transport, kinetic and hybrid simulations, etc.). It is not clear whether this is because the code developers are not

sending their codes to CCMC, whether CCMC lacks the expertise to incorporate these models, or something else.

The panel also agreed that the center should gauge the value of its own science output. At present, this is not clear. Regular science nuggets supplied by users and posted on the CCMC website may be a useful for accomplishing this.

CCMC was strong on E&PO support and the budget for this seems reasonable. This program naturally lends itself to public outreach. Activities at CCMC provide a natural entry point for young researchers to get into the field of space physics: Most young people are very knowledgeable of computers and are eager to apply their skills to scientific problems.

2.2 National Space Science Data Center (NSSDC)

Summary of Project:

NSSDC provides long-term data archive and preservation for NASA space-science data (Heliophysics, Planetary Science, and Astrophysics). NSSDC undertakes to assure that the archived data are independently understandable (usable) through a comprehensive set of documentation activities. Where NSSDC is the sole provider for a given archive, they provide access via the web and on-line services. For some data sets NSSDC provides a preservation function and provides "back-up" while the data is being served to the community by an Active Archive (AA) or Resident Archive (RA). For NASA, NSSDC is the permanent archive for SPDF, SDAC, the Planetary Data System (PDS) and Astrophysics' Science Archive Research Centers (SARCs).

In addition to its role in specific data set preservation NSSDC has an emerging role in defining data preservations standards and best practices for NASA. NSSDC is developing or contributing to community metadata standards such as Space-Physics Archive Search and Extract (SPASE) and data archiving reference models such as the Open Archival Information System (OAIS) reference model. In addition, the staff from the NSSDC participates in defining the role of NASA RAs and their role in the over all data management plan of NASA. Finally, NSSDC performs a data management education function through the collection of various documents, plus archiving and serving as a conduit for active amateur radio astronomy projects served via the NSSDC web site.

An important component of the NSSDC mission is to assist the mission scientists in planning for the preservation of their collected data. NSSDC does this through MOU's and project data management plans developed with the project scientist. The NSSDC can assist but not direct projects in the best way to preserve their mission data for future scientific use.

NSSDC Level-1 requirements (from the proposal):

- Data from missions must be captured, held safely and securely, and made available to the research community for as long as it has significant value.
- Data made available to the research community need to be well documented in order to support independent usability via virtual observatory access and in recognition of the reality that a large fraction of NASA and support contractor staff are likely to retire in the near future.
- Repositories, as the source of the desired data/metadata, need to participate in Virtual Observatory development efforts to assist in the practical evolution of those concepts.
- The NSSDC should pursue Education and Public Outreach efforts by repositories when they can be done cost-effectively, leverage their research support, and also encourage an adequate Science, Technology, Engineering and Math (STEM) workforce to meet NASA's needs.

Panel Assessment (strengths, weaknesses and other findings):

The panel finds that NSSDC has very strong relevance and value, as well as significance and importance to SMD. NSSDC proposes FY (07-10) work in 6 areas: Ingest Evolution, Administrative Evolution, Archival Storage Evolution, Data Management Evolution, Access and Preservation Planning. The panel finds that the activities outlined are all germane to the NSSDC mission. The panel also finds that NSSDC has developed a strong formalism of ingest procedures and has well defined agreements with data providers through developed MOU's. However, the panel believes that the NSSDC should place stronger emphasis on data ingest through more negotiation of MOU's. It should work with both NASA and the community to find ways to streamline the time-consuming MOU process. This can be facilitated by working extensively to understand and define the emerging distributed archiving paradigm and the role of NSSDC with respect to RAs. Perhaps a more proactive approach can be followed where there can be more involvement earlier in the mission on data products. Here, documentation and archive resource planning can improve the transfer process. The panel believes this can be achieved with the "in guidance" budget via stream-lining elsewhere in the project.

The panel finds that one task area, that is missing or under emphasized, is that of data stewardship, which is necessary to meet the first two NSSDC Level 1 requirements. This includes activities such as data quality control, productization of data, repackaging, etc. All of these are essential to assure the long-term usability and to maximize the return-on-investment for archived data. We believe that NSSDC should undertake and active program in this area using the "in-guidance" funding level.

The panel finds that NSSDC fulfills a clear and important mission within NASA, namely providing the permanent archive for space sciences data from a very large number of NASA missions. NSSDC has adopted the Open Archival Information System (OAIS) reference model, which provides a very solid basis for assuring adequate data archiving practices. NSSDC has very good tracking of storage technology and evolution of archival storage media. The panel finds strong compliance with data preservation standards (i.e. NARA) and notes the NSSDC's expanding role into photoproducts. However, the documentation generation and preservation packaging seems to be handled on a case-by-case basis, and the panel finds no evolution towards a more comprehensive plan. This procedure is cost-effective, however, in that the digitizing of images is done by an "as the demand arises" basis.

The panel finds that the NSSDC's work and cost estimates are realistic: archiving and preservation is a challenging task and NSSDC's approach appears to be appropriate for an important national archive. The panel notes that a significant number of FTEs going into archival activities (~5/year). It is difficult to ascertain from the proposal whether the cost/output is reasonable because the rate of new data ingest is not described. The panel finds that the in-guidance funding level is quite reasonable for NSSDC to increase its activities in science data stewardship, become better connected with mission data preservation planning and help lead the transition to a more distributed data archiving environment.

The panel finds that NSSDC's role in the VxOs framework is not specified (how/where does it fit the OAIS model?). It would be useful for NSSDC to define its role in a distributed data environment, e.g. could it act as active back up (replica) for mission archives (including SPDF and SDAC, PDS, SARC, active archives)? The panel finds that NSSDC has an apparent lack of influence/leverage with providers, and needs a stronger role in assuring data preservation across all NASA space data holdings. Also, the panel was concerned with NSSDC's reliance on the SPASE metadata model – the panel finds no real backup plan for such a critical function.

The panel noted NSSDC's lack of connection to relevant international programs. For example, what streamlining of the World Data Center (WDC) is planned? In addition, there was no mention of the WDC role on a larger scale (e.g. recent initiatives coming form the Science Data and Information Forum of the International Council of Science Union). The panel also finds a lack of emphasis on NSSDC system compatibility with other major data systems (i.e. CLASS, EOSDIS).

The panel finds that NSSDC's plan for technology infusion (other than storage technology) is not well articulated and some NSSDC efforts (e.g. in proposal section 3.5.d – XML work) seem to be lagging those of its peers.

The panel finds that a focus on data preservation and integrity is in order as is the development of standards, best practices, and guidance for resident archives (RA). Regarding scientific data stewardship activities, the panel finds that the NSSDC, acting as the permanent archive (PA) is performing a very good service. The panel finds that the substitution of RA effort for some other work (~$100K/year) to bring it in-guideline may assist NSSDC to fulfill its mission.

Awareness and importance of E&PO was found to be very good. As the role of NSSDC evolves, will the current mix of E&PO efforts evolve (as the 'public' for NSSDC evolves)?

Overall Summary Assessment:

NSSDC continues its long and robust history as a permanent archive for space science data. The panel believes that the NSSDC should place more attention on the (computer) interfaces between RAs, AAs, and VxOs in addition to the organizational interfaces (e.g. MOUs). The NSSDC should also devote resources to implementation of leading-edge technology to facilitate a number of its goals (e.g. documentation, provenance, metadata services, web service interfaces, etc.).

2.3 Solar Data Analysis Center (SDAC)

Summary of Project:

SDAC provides three critical services to the solar physics community using mission data.

(1) It provides open data service for Yohkoh, SOHO, TRACE, RHESSI, GOES (3s cadence SXR photometer data), and in the future, Solar-B (2006), STEREO (2006) and SDO (2008). Some older mission data are also archived and available, such as SMM, CGRO/BATSE flare, OSO-7.
(2) It provides software access to analyze the above data, through the development of the SolarSoft software tree.
(3) It leads the development of Virtual Solar Observatory.

SDAC's mission is to combine the expertise in scientific research, mission operation and programming to most efficiently serve the solar community and collaborate in scientific research. SDAC is currently serving 4.4 Tbytes of data on its archive. The two largest ones are SOHO (at 2.7 Tb) and TRACE (at1.3 Tb). In the near future, new services of Solar-B will have 10 Tb/year, STEREO@5 Tb/year, and SDO@1Tb/day. SDAC has led the effort of developing the VSO, which has included all the data from solar missions and substantial fraction of ground-based data.

SDAC Level-I requirements (from the proposal):

- Provide easy access research-quality data from all NASA-funded Solar Physics Missions.
- Provide easy access of data analysis tools through the SolarSoft software tree.
- Provide access to current solar observations to enable observational and operational decision-making through SolarMonitor and Latest Events services.
- Manage SDAC to satisfy government requirements, in particular, the network security rules.
- Investigate Commercial Off-the-shelf (COTS) technology transfer.
- Maintain VSO baseline operation, including new STEREO, Solar-B and SDO data, add more ground-based data, add more catalog/events lists, and enhance the SDAC capacities.

The optimal budget requests additional funding for further development of VSO, mainly to provide the ability to join searches on multiple event lists and data sources, and to be compatible with heliophysics meta-VO and semantic Web.

Panel Assessment (strengths, weaknesses and other findings):

The panel strongly endorses SDAC's plan to continue to maintain and develop Solar Soft—a must for every solar physicist in the world that analyzes imaging data.

The panel finds that SDAC manages large amounts of data and data analysis tools efficiently through their expertise in mission operations, software design and engineering; and peer collaboration. The SDAC also has a strong familiarity with science and mission needs, use of data, software needs and the challenges they represent. The location of this group at GSFC, where there is also a very strong solar physics group very closely tied to RHESSI, SOHO, and other solar space missions has provided tremendous leverage in accomplishing the work they have done. The panel finds that the combination of data

archiving with data analysis is an excellent concept. The solar physics community, in general, strongly interacts with this facility. The panel believes that SDAC directly supports many missions in the concept of the Heliophysics Great Observatory. Their work is critical to understand the Sun and its effects on earth and solar system. The panel finds that the initiation of the use of ground-based data sources to support mission data analysis is unique and should be encouraged.

The panel recognizes that SDAC has led the effort in the successful development of the Virtual Solar Observatory. It sets a good example for other VxOs. The panel finds that SDACs expanded efforts in VSO development are very valuable to the community and that SDAC would greatly benefit from the proposed optimal funding level to cover advanced VSO work. Moreover, the panel believes that this work would benefit the overall NASA VxO initiatives if it worked in a coordinated and collaborative manner with the other VxO projects, selected competitively. The SDAC may be able to perform this via its VSO development and that the specific form of the work would be defined in terms of the mutual benefit to each group.

The panel believes that some community effort in data analysis tools should be organized by SDAC, such as Artificial Intelligence to detect events and routines to resolve the 180 deg magnetic field ambiguity in vector magnetograms and Stokes inversion and coronal magnetic field extrapolation (jointly with CCMC).

Overall Summary Assessment:

The panel finds that the operation of SDAC is highly cost effective, and provides extremely critical services to the solar community. FY07-10 is a critical period for SDAC to be well supported, because of new missions, Solar-B, STEREO and SDO. The panel finds that SDACs expanded efforts in VSO development are very valuable to the community and that SDAC would greatly benefit from the proposed optimal funding level to cover advanced VSO work.

The panel believes that SDAC could be more proactive in educating the solar physics community about data analysis tools available through the center via the SolarSoft software tree. This education could be accomplished in community workshops organized by SDAC. SDAC could also consider working with CCMC on supporting solar physics models such as coronal magnetic field extrapolations from vector magnetograms.

2.4 Space Physics Data Facility (SPDF)

Summary of project:

The SPDF provides a variety of data and services to the space physics community through their web interfaces. The SPDF is also active in the collection of models and the construction of certain data sets. It cites "… ensuring integrated access to the data of the Heliospheric Great Observatory (HGO), to best enable science discoveries and understanding of the heliophysics system across the boundaries of missions and disciplines." as the overarching Level-1 requirement for the SPDF. SPDF also proposed to incorporate VSPO into the SPDF where it would receive direct funding. Previously, it was supported under the AISR program, but has now matured to the point where it makes sense to attach it to a data center.

SPDF Level-1 requirements (from the proposal):

- Build Services to Lead the HDE
- Integrate Access to HGO Data (Develop and operate multi-mission cross-disciplinary data and orbit user services)
- Continue our assigned mission and legacy active archive roles for Cluster and NSSDC
- Leverage our core capabilities to enable data accessibility through sonification

Panel Assessment (strengths, weaknesses and other findings):

In our detailed evaluation, we follow the work breakdown structure given in the proposal (WBS numbers follow parenthetically):

SPDF Web Services, VSPO support, and CDAWeb (WBS 1.1, 1.2 & 2.1):

The panel was generally impressed with the VSPO and its capabilities and accepts VSPO's incorporation into SPDF. Specifically the panel was impressed by the functionality of the VSPO, which seemed to make it easier to find and to access data that are otherwise served and accessed by the existing CDAWeb interface. The panel finds that SPDF's plans to develop a new user interface for CDAWeb are not well justified and that it would be more prudent to extend the VSPO interface to provide the full CDAWeb functionality. As advertised, VSPO could then also, in time, be used to access data that are not served by CDAWeb, but by other sites as well, possibly in cooperation with the VxOs that are currently under development. The panel finds that SPDF's data holdings and underlying functionality should remain as is, and be improved where needed, as discussed below.

The panel finds that some of the data sets served by CDAWeb are duplicates of data sets that are served from other sites. One example is the Cluster data that are now fully available from the ESA Cluster Active Archive (CAA), but there are other data sets as well. In addition, CDAWeb holds data that are known to be deficient (by admission of

their creators.) The panel finds that SPDF should make efforts to no longer offer data sets that are (i) otherwise available and accessible from their originators, (ii) available from other archives that are closer to the data providers, and (iii) of low quality. The overriding principle should be the trustworthiness of the data, even when caveats are attached. The panel finds that only reduced funding is necessary to develop and provide this data service, because no new CDAWeb user interface needs to be developed.

Heliospheric Data Environment (HDE) Functional Services (WBS 1.3):

The panel had difficulty understanding what "cross-cutting functional services" actually means for a SPDF user. The panel finds that such services should not be part of SPDF's core mission, but that such services should be developed under competed grants. The panel did not understand what these services would be in detail, since the proposal is very vague on this issue.

HDE Coordination (WBS 1.4):

The panel finds that there is no need for a "VxO center" and that such an endeavor would be contrary to the very idea of how VxOs work. Such a center would, in fact, likely interfere with the separately funded VxO development efforts. SPDF personnel involvement in VxO related efforts is welcome, but only at a level that is equal to those of grant funded efforts. The panel finds that separate funding for such involvement is not necessary and that such efforts are already covered, for example, as part of the VSPO work.

Implement a common heliophysics data dictionary (WBS 1.5):

Because of its unique expertise from SPASE and VSPO, the panel finds that SPDF should engage in the implementation of a common heliophysics data directory (dictionary).

Sustain the vitality of the CDAWeb data service until superceded (WBS 2.1):
See comments for SPDF and HDE service (WBS 1.1 & 1.3) above.

Evolve the CDAWeb concept into a new generation of services (WBS 2.2):

Like WBS 1.3, this sub-project is not well defined and the panel cannot see how the community will profit. The panel finds that CDAWeb follow-on work should better be focused on supporting the VxOs. The panel finds that this effort should not be funded.

Maintain CDF and enhance its external interfaces (WBS 2.3):

Since CDF software is widely used the panel finds that the software should be maintained. However, the panel questions whether more than 1 FTE/yr is necessary for this effort, in particular, since no new releases are anticipated.

Support for the OMNI database (WBS 2.4):

The panel finds that the OMNI database widely used by the community, and its continued development should be supported as a core service.

Enhance the SSCWeb orbit and query service (WBS 2.5):

The panel finds that the orbit and query service is unique and should be supported under any circumstances. Further, the panel finds that this should be regarded as a core service.

Recreate ModelWeb's capabilities as services (WBS 2.6):

The panel finds that the SPDF should consider moving the models to the CCMC where they would make a better fit and augment the CCMC collection of numerical models.

Pursue the VSPO DataShop and Higher Order Query Service (HOQS) (WBS 2.7):

The panel finds that Datashop and the HOQS appear to be functions that really belong in the realm of VxOs. The panel finds that such projects should be competed.

Merge or retire old services (WBS 2.8):

The panel applauds SPDF's decision to retire obsolete services. The panel does not understand the rationale for the FTEs required to do this under the two funding scenarios.

Act as a Legacy Active Archive (WBS 3.1):

The panel finds that there is a need for the development of a clear data policy that delineates the role of the SPDF from that of the NSSDC. The SPDF should not serve data that are already served by someone else. In general, NSSDC should not have to duplicate the machinery to catalog (for VxOs) and serve data in the same fashion as the SPDF. However, the policy would outline the rules as to which data sets should be both permanently archived by NSSDC and at the same time be offered by SPDF. The panel finds that this topic may be appropriate to be addressed by the Data Operations Working Group (DOWG.)

Overall Summary Assessment:

The SPDF makes "integrated access to data" its key objective (Level-1 requirement). At present, however, it offers a number of services that do not appear to be integrated. The proposal, in its sparse details, also shows very little emphasis on integration. It focuses mostly on improving and extending the existing products and services. These products services cover a wide spectrum. Some data and services are not available anywhere else and thus are deemed essential, for example, the spacecraft orbit and OMNI data. However, other data and services may be of less value to the community. The panel is particularly concerned about duplication of efforts, and the SPDF's engagements in areas

that are better served by competed projects or community grass-roots efforts, as detailed below.

The panel also notes that the proposal was difficult to read, using too many buzzwords in place of specific details. Likewise, the presentation left some questions open in spite of detailed inquiries by the panel. The panel also sees a need to correct the breakout of "base" versus "optimal" activities because some of the "optimal" activities appear to be rather basic and vice versa.

To a large part, the SPDF is a "data center", which is by its very nature contrary to the idea of distributed data that make up a Virtual Observatory. Nonetheless, mostly for historical reasons, the SPDF should continue in this role where it makes sense. However, whenever data can be served by the PI or a mission active archive, SPDF should not serve those data any longer, but defer to that data provider, provided there is no break in the functionality of SPDF services. As guiding principle, data should reside as close as possible to its creators.

The panel finds that an "oversight committee" would be able to assist SPDF in critical choices, such as to which data sets to serve or discontinue to serve, or which services to develop and offer. The panel finds that a review every 3-4 years is not sufficient to provide enough guidance. On the other hand, a "user group" may make useful suggestions but is unlikely to observe day-today operations, and is generally not a body that the SPDF leadership would report to.

The panel also notes that only 15% of the SPDF staff are scientists. This low number appears to be having a negative impact of SPDF's choices and its utility for the community. Many of the proposed projects seem to be driven by computer scientists' and programmers' desires for newest software technologies rather than by the actual need of the science community. This is clearly reflected by the semantics of the proposal, which is full of computer and software buzzwords but hardly mentions specific community needs. In the process of this review the panel observed that the centers that have a larger fraction of scientists on their staff (SDAC, CCMC) also seem to achieve higher "user satisfaction" levels. The panel thinks that it is essential for the SPDF to build collaborations with providers and users to guide the development of new services and middleware.