

Reactions to the Senior Review of Data Centers, Fall 2009

Overall the report presents a fair and well-considered view of our current state and future directions. The call for clearer lines of authority and more definitive statements of policy points in the direction we have been continuing to pursue, and it is helpful to have confirmation that this is appropriate. The panel rightly noted the need for more significant community buy-in on issues such as the need for integrative approaches to data analysis. The issue of data quality is one that the HPDCWG has been struggling with, and it will be important to work to solve the various problems in this area. Stronger interactions between missions (from their inception) and Final Archives is desirable. Our interfaces for search, access, etc. certainly need work. Many of the specific suggestions for improvements are very helpful, and will be implemented.

There are a few points where we apparently did not communicate sufficiently clearly, or took for granted a very careful reading of the Data Policy. The main ones are:

(1) The report does not properly characterize the role of SPASE:

(a) SPASE is very important for providing uniformity of description, search, access, and registration, but it is not intended to provide interoperability of datasets. It does not replace all FITS, CDF, etc. headers, and will not solve the problem of the variety of formats. This is to be accomplished by tools such as DataShop, which enables format independent tools, including Dataset Runs on Request, Autoplot, and IDL access.

(b) We do not believe it will be difficult to provide SPASE descriptions of VSO datasets, and a number of these exist already in VSPO. There have been many conversations between VSO and HDMC on the issue of SPASE descriptions, and SDAC has taken a constructive position, with the problem being constrained resources. We expect to solve this problem in the next year.

(c) The “automatic translation” of CDF to SPASE is only partial and the creation of SPASE descriptions will never be seamless due to missing and incompatible metadata. Moreover, the idea is not to change CDF files into SPASE files, but rather to pull any possible SPASE information from CDF headers. FITS to SPASE translators will not be very useful, since the FITS headers are not intended to convey SPASE information, and the SPASE information required for most solar physics products is much simpler than that for many space physics products.

(2) There were a number of issues identified as problems that are already dealt with explicitly in the HP Data Policy:

(a) Missions are already required to have a PDMP that includes discussions of the archiving and serving of data, and they are already required to provide documentation of data, algorithms, data quality, etc. The enforcement of some of these is still something of

a problem, but each successive Senior Review of missions has included a deeper review of data issues that has greatly improved the availability and quality of datasets.

(b) RAs are already required to have a plan for transitioning their data to a Final Archive, although the details here might be spelled out more clearly.

(c) The required Mission Archive Plans already call for missions to determine what their legacy data products will be and how they will be documented so that the products will be useful when mission/RA support is no longer available, thereby maximizing long-term science value. Their performance in this regard is to be evaluated by the mission Senior Reviews.

(3) While it is stated that there is considerable duplication of efforts across the Data Centers, very little is spelled out, and we do not see much actual duplication. In the case of Browsing and Visualization, for example, SDAC has no efforts comparable to Autoplot, and HDMC and SPDF are working together to produce the best result. SPDF tools, such as in CDAWeb, do not readily generalize to datasets outside of those in that system, and thus these tools do not replace the need for Autoplot-like capabilities. SDAC, through VSO, is developing useful browse tools for image data, and these efforts will be taken into account in devising more general tools. All that said, we agree completely that more systematic coordination of efforts is needed, and we will work to achieve this.