

**Report of the NSSDC Users Group (NUG) Meeting
Held on December 12, 2005 at GSFC**

Members:

Fred Bruhweiler (Catholic Univ. of America), Chair
Paulett Liewer (JPL)
Mal Neidner (NASA/ GSFC, Code 662)
Ian Richardson (Univ. of Maryland)
George Rossano (Aerospace Corp.)

The NUG recognizes NSSDC's chartered role as the permanent repository of NASA-funded space science data and strongly endorses the continuation of that role. The major recommendations by the NUG are summarized below. More detailed comments follow:

Summary of Recommendations

- 1.) The fledgling relationship between NSSDC and Resident Archives (RAs) is an important one which should be continued and strengthened. (Combine with next bullet as per detailed comments?)
- 2.) NSSDC's technical oversight role of RAs has the potential of producing a better data product and increasing efficiency. The NUG is concerned that, should this function become an increasing fraction of NSSDC activities, then NSSDC's budget should be increased accordingly.
- 3.) The NUG endorses NSSDC's plan to export AIP generation software (MPGA) to active archives, and resident archives. This procedure insures a more error-free product and reduces NSSDC ingest workload.
- 4.) NSSDC's unique datasets should be registered for use for Virtual Observatories (VOs). NSSDC should be playing a strong role in setting up VO standards.
- 5.) NSSDC should not accept commercial off the shelf database submissions. NSSDC should only accept those formats that are commonly used in the scientific community.
- 6.) Tracking of NASA-funded datasets is valuable, but large amounts of resources of NSSDC should not be expended on this task.
- 7.) Migration of legacy analog data to digital status should proceed at a reasonable rate, but guided by user demand and budgetary realities.
- 8.) The NSSDC's Educational and Public Outreach (EPO) involvement should not extend significantly beyond the level set by public requests of data. Expanded use of NSSDC infrastructure for EPO activities should be contingent on external funding.
- 9.) In any NASA/GSFC plans to reallocate NSSDC workspace in Building 28 at GSFC, the NUG strongly urges that that the NSSDC workspace be kept contiguous. Not doing so would greatly jeopardize basic NSSDC archiving functions.

Detailed Comments from the NUG.

NSSDC Relationships with Resident Archives:

The NUG considered the nature of existing and future relationships between the NSSDC and Resident Archives (RAs). The RA concept is new, and, at present, no formal relationship exists. The RAs are chartered to serve “active” data to the community during their (finite) lifetimes. The NSSDC is chartered to be the “deep” archive for this data and will presumably become the community source for the data when RAs are terminated. The committee strongly encourages a strengthening of the current fledgling relationship between the NSSDC and RAs. Specifically, the committee recommends that the NSSDC assume a role of providing technical oversight to (and possibly management of) the RAs in order to create better, more consistent data products, to make the RAs more efficient, and to make the transfer from the RAs to the NSSDC more efficient. Moreover, should this technical oversight function become an increasing fraction of the NSSDC activity as the number of RAs grows, then the NSSDC budget should be increased accordingly. NSSDC roles in both the technical oversight and management of the RAs are discussed in “White Paper: Resident Archives in the S3C Data Environment, “available on the NSSDC web site.] (“Suggestions” might appear to imply that they are NUG suggestions)

Exporting of Archival Information Package generation software:

The committee endorse NSSDC's plan to export Archival Information Package (AIP) generation software (Multifile Package Generator and Analyzer, MPGA) to active and resident archives and other data providers to facilitate the delivery of data sets to NSSDC that are already formatted as AIPs. This benefits NSSDC by reducing the ingest workload, including the interaction required between the provider and acquisition scientist, while the data provider benefits by delivering a product (data and supporting documentation) that fully incorporates the expertise and desires of the provider. The committee notes that while the number of AIP-delivered data sets is growing, these are currently largely limited to those with strong NSSDC connections, such as IMAGE and CDAWeb. The NSSDC is encouraged to promote the wider use of this software, ensure that it is as user-friendly and error-free as possible, and perhaps make its use a requirement in MOUs with individual data providers. This software is currently available in Solaris and Linux versions which the NSSDC believes will be useable by many providers. The development of other versions, for example running under Mac or Windows operating systems, would appear to be possible but the size of the potential user base is unknown. Noting that the STEREO project, for example, is standardizing around a Mac O/S, the committee suggests that a survey be made to assess the level of interest in additional versions of MPGA and hence whether it is worthwhile developing these versions versus generating the AIPs at NSSDC for such providers.

Virtual Observatories and the Role of the NSSDC:

The NUG considered how, and to what extent, the NSSDC's data holdings should be “available to” Virtual Observatories (VOs). This is an important topic in view of the increasing need for multi-facility/-wavelength/-discipline datasets in the examination of

complex research questions, and the likely heavy use of VOs in such studies. The NUG was briefed informally by the NSSDC Chief and his staff about the status of VOs, such as the National Virtual Observatory (NVO) in astrophysics and the Virtual Space Physics Observatory (VSPO). It was clear that VOs are either in a substantial study/development phase, or are actually operational at some level.

In consideration of the NSSDC's role in VOs, the NUG strongly recommends that any NSSDC unique digital datasets, i.e., those that are not primarily held elsewhere in an active, resident, or mission archive, should be registered with the appropriate VO. Second, the NSSDC should continue its discussions with the VO communities, expanding them if necessary, in order to promote strong VO data standards and a better understanding of overall needs. The Committee also wrestled with the question of whether important datasets which are held by NSSDC, but are not primary to it, should be VO-registered in a back-up mode to increase VO efficiency. The conclusion was that absent any known requirement by VOs to achieve a data retrieval success rate of (say) $> 98\%$, the NSSDC should concentrate on registering only those digital datasets for which it has primary/permanent responsibility. It was further noted that many of the NSSDC's uniquely held data are still in analog form, and that as they migrate to digital status—guided by peer-determined priorities and resource availability—they should be VO-registered in due course.

Efficient Use of Resources:

There are several areas in which the NSSDC has and should continue its policy of making efficient use of resources and guaranteeing the integrity of the archived database.

First, the NSSDC should not accept commercial off the shelf database submissions. NSSDC should only accept those formats that are commonly used in the scientific community. This guarantees that integrity of the established archive and also makes the data easily accessible to the general community without the user needing to make budgetary outlays to commercial proprietary software. Ultimately such costs would likely be transferred back to the funding agency, typically NASA.

Second, Identifying and tracking of NASA-funded datasets is valuable, but large amounts of resources of NSSDC should not be expended on this task. This task is certainly straightforward for the recognized, established active archives and here lies the NSSDC's primary responsibility. In the case of small NASA-sponsored experiments, the duty for transferring the data to the NSSDC archives should largely fall on those responsible for the funded experiment.

Thirdly, NSSDC is expected to take the responsibility for migration of legacy analog data to digital status. This should proceed at a reasonable rate, but guided by user demand and budgetary realities.

Educational and Public Outreach:

The NSSDC's Educational and Public Outreach (EPO) involvement should not extend significantly beyond the level set by public requests of data. In most cases, the role of for educational and public outreach for the various datasets is covered by Resident Archives. Only in the case of the legacy data maintained by the NSSDC does the NSSDC have a unique role. Thus, when requests for these data are made, the NSSDC has the responsibility to provide them to the public. Once a data set has been provided, the NSSDC, should, if possible, make the data set available to others. On possible way to provide easy and efficient access to these same data by the public would be to make them directly accessible on the Web.

The expanded use of NSSDC infrastructure for EPO activities should be contingent on external funding. The present involvement of NSSDC in programs such as Radio JOVE and Moon Trees is to be commended.

Workspace at NASA/GSFC for the NSSDC:

The NUG was made aware of NASA/GSFC administration plans to make available space at GSFC for NOAA which could impact the present workspace for NSSDC in Building 28. The NUG is greatly concerned that such plans may split up the present NSSDC workspace. Failure to keep the NSSDC work area contiguous would have a negative impact on the main mission of the NSSDC, namely data preservation. This function requires the periodic transferal (renewal) of data from older storage media to newer media. This entails the manual handling of old, problematic tapes, which often require a high degree of personal intervention to insure that information is transferred successfully. This process would be greatly jeopardized if staff must travel down a long corridor or up and down stairs to manually interact with tapes on their drives, or to transfer the tapes from the custom-designed, immovable, tape storage area to the drives. We therefore urge that any necessary reallocation of workspace within Building 28 should keep the NSSDC work area contiguous. [Any comment on moving analog storage area?]

For the NSSDC Users Group:

Fred Bruhweiler
16 January, 2006