

The NSSDC Users Group (NUG) met on 16 March, 2009 at NASA/Goddard Space Flight Center in Greenbelt, MD. The group members in attendance were: Drs. Fred Bruhweiler (CUA), Ian Richardson (NASA/GSFC/UMD/CRESST), George Rossano (Aerospace Corp.), Robert Benson (NASA/GSFC), and Paul Lowman (NASA/GSFC). Presentations by members of the NSSDC staff discussed a variety of topics ranging from the Space Physics Archive Search and Extract (SPASE) effort to preservation of the non-digital data in possession of the NSSDC. The NUG recommendations, based upon the meeting, are summarized below.

NSSDC's Analog Archive. The NSSDC has made progress in organizing and cataloging its analog materials, but much remains to be done. The lack of a comprehensive electronic, searchable catalog, the analog nature of these materials, and their storage at multiple locations (GSFC, Iron Mountain or the Federal Records Center (FRC)) make exploiting this important resource difficult for users.

At a minimum, the NSSDC needs to complete the effort underway to thoroughly catalog what they have and where it is located, and make this information available to the users. In addition, the NSSDC must ensure that their analog holdings are in a safe condition. By safe condition, it is meant that there is little risk of slow deterioration or of sudden catastrophic loss. In particular, the NUG expresses concern over the quantity of NSSDC film and microfiche data records stored at the FRC and the poor quality of that environment for archival storage, especially the lack of climate control. We encourage the NSSDC to explore the possibility of moving these records to GSFC, if a proper archival environment can be identified, or Iron Mountain. In particular, high priority should be given to removing photographic material from the FRC.

Looking beyond this minimum step, it is recommended that the NSSDC better organize the analog materials, and perform the analog-to-digital (A/D) conversion of these records. The latter will require a strategy for prioritizing the A/D conversion, and to determine the best approach to digitization. It is recommended that the task of prioritization include not only the scientific timeliness of the material, but also the risk of loss due to the material being "overtaken" by changes in technology or the stability of the media, and also the historical value of the material. The goal would be to develop an efficient means to protect the analog datasets, convert them to digital media, and make them available in a digitized format to the scientific community.

It is suggested that the NSSDC form an advisory group to help them in the above process, similar to the approach that has successfully prioritized the NSSDC tape holdings for conversion to newer media. This process should start with cataloging the available analog datasets, and then making the catalog accessible on the NSSDC website. Tasks for the advisory group would include: prioritizing holdings for scanning as appropriate; identify as low priority those microfilm/microfiche holdings that duplicate digital databases (e.g., contain lists of data on the tape, or plots of data); and identify a cost-effective method of scanning large maps and photos at the highest possible resolution.

It is recommended extreme restraint be shown in regard to the disposal of duplicate material from the analog archive prior to the archive being assessed by the proposed advisory group.

It is appreciated that converting the analog datasets to a digitized state represents a significant effort. The NSSDC is encouraged to work with other groups, who might be interested in these data, to provide additional leverage to accomplish this work. The NSSDC has been successful in the past in getting external funding to accomplish similar tasks and such efforts should be continued.

Analog Data Storage at the NSSDC. The NUG recommends that the NSSDC needs to immediately examine the storage of analog materials at Goddard. These include some color material that is ~40 years old and shows signs of degradation. The room housing the photo archive is better climate controlled than the FRC and likely cleaner, but is still sub-standard for the archival preservation of photographic materials. The space has no temperature or humidity monitoring or control other than the building air handlers, and the air supply appears to be unfiltered. The humidity range is wider than recommended for a photo archive, and proximity to a loading dock can produce rapid fluctuations in conditions in the storage room. Although a dedicated heating/cooling system has been requested, GSFC has not provided it. Many of the storage containers are not of photo-archival quality. Of greatest concern is that the sprinkler system is tied to the fire control system for the building as a whole, and a small fire remote from the storage area will result in a deluge in the archive. Such an incident would result in the loss or damage of much of the archive since much of the material is not in watertight containers. In particular, there would be damage to material contained in cardboard boxes throughout the archive, many of which are stacked on top of cabinets. Fire safety and protection from smoke and water damage in the storage area needs immediate attention.

Data Migrations. The NUG congratulates the NSSDC on the progress since the last NUG meeting to prioritize the astrophysics, space physics and planetary science tape holdings for restoration onto newer storage media with the help of an advisory team of scientists and NSSDC staff members. We are encouraged at the high rate of success at reading “unreadable” tapes achieved by the contractor, JLB.

Generation of ASCII Tape Copies of Datasets. Many data sets currently on tape are written in binary of various formats, e.g., VMS, EBCDIC, and it is often difficult, frustrating or impossible for a potential user to extract the data from a direct copy of such a database provided by the NSSDC or convert the binary data into a more user-friendly form such as ASCII. However, such conversions are made routinely at the NSSDC. It is suggested that the NSSDC investigate the possibility that ASCII versions of a binary datasets held on tape could be generated during the process of copying the datasets (still retaining the binary format) to newer storage media. Generating ASCII versions of such databases would make them far more accessible to potential users, and the NUG therefore supports efforts to explore whether this is feasible. We recognize that there are several aspects that should be considered: Most importantly, the integrity of the ASCII data sets generated should be carefully monitored to ensure that they are faithful to the original binary versions and scientifically useable. Sufficient documentation should also be available to allow the user to interpret the parameters in a given data set. Ideally, the ASCII data sets should be “peer reviewed”, either by an NSSDC staff member with some familiarity with the type of data set, or by other persons with such expertise, before being made available from the NSSDC. Such a review process could introduce significant delay before the “approved” data set is released. The alternative, to post the ASCII version “as is” with caveats about its quality, may

lead to a reluctance for potential users to download the data, or result in scientific errors if there are unrecognized problems in the dataset. We suggest the impact of the routine production of ASCII versions of binary data sets on tape on the rate of tape migration should be examined, and that a practical method of reviewing such versions be developed. If the impact on tape migration is significant, it may be appropriate to convert to ASCII only those binary data sets that were judged to be of the higher priority in the recent review of data sets held on tape or only those datasets where such conversion would be most beneficial to the user.

Science Archives Workshops. The NUG considers it essential that the NSSDC presents its methods, practices and proposed technological approaches at national and international science archives meetings. Such meetings are a critical forum for gaining information about new and more efficient approaches to archiving and for informing the general scientific community about the activities within the NSSDC. Therefore, the NUG strongly endorses the NSSDC's plans to hold a second Science Archives Workshop at the University of Maryland in the fall of 2009. This will include a mix of invited talks, oral presentations and posters to discuss best practices, lessons learned, policies, problem areas, novel approaches and promising technologies.

Status of Space Physics Archive Search & Extract (SPASE) in the NSSDC. The development of SPASE at NSSDC has progressed significantly, and has made a useful and helpful contribution to the overall mission of the Center. Furthermore, the NUG suggests that the NSSDC be a strong advocate in the use of SPASE. The NUG also commends the joint efforts of the NSSDC and UCLA to provide SPASE tools and migrate the web-site to the NSSDC. The NSSDC should continue to work closely with the heliophysics community to further enhance tools for using SPASE. The adoption of SPASE for the Heliospheric Virtual Observatories (VOs) is a testament to the success of SPASE and the efforts of the NSSDC. The NSSDC is encouraged to explore if SPASE can be adopted by yet a further audience.

In the most recent data model draft (a 120 page document) the introduction states: "... there has been little corresponding unification of data access, formats and tools, resulting in a great deal of time being spent finding, retrieving and reformatting data. The key to reducing this inefficiency is a uniform way to describe adequately what exists. This is the purpose of the SPASE Data Model..." These are features that should be attractive to scientists using such data sets.

However, such documentation, which appears to be aimed at computer programmers, does not successfully convey how SPASE can benefit the scientist interested in using it, and how it could be implemented.

Trustworthy Archives. The NSSDC has made recommendations for standards that resident and active archives should adopt, but no definition exists of what constitutes a "trustworthy" archive. The term implies that a user can trust an archive's content. If this turns out not to be the case, it will (rightly or not) reflect poorly on the NSSDC and its reputation. On the other hand, since the datasets in an archive are provided by experimenters, there is inevitably some variation in the care at which these experimenters correct errors in their data and make them usable to others. At some point, it may be impossible for anyone else (e.g., the NSSDC staff) to independently verify the accuracy of a data set. Consequently, there will always be some ambiguity in what is meant by a trustworthy archive, and an element of "buyer beware".

Despite this caveat, it is recommended that the NSSDC establish a definition and requirements for what constitutes a trustworthy archive and publish these on its web site. Only archives that meet those requirements should be pointed to with the attached designation “trustworthy.” This term in effect becomes an NSSDC “seal of approval” for an archive. The NSSDC can point to archives that do not meet the standard for a trustworthy archive, but the designation “trustworthy” should then be withheld from those archives.

The NSSDC Web-site. It is very important to make the web-interface with the NSSDC more user friendly. One way to promote SPASE is through the web-site. However, the website doesn't seem very friendly to the scientist trying to find out whether SPASE might be useful to their work. Considering the new importance associated with the analog lunar data (photographic/microfilm/microfiche holdings), it is a high priority to provide a reliable listing of these assets on web-site such that they might become useable to potential investigators.

Other Comments or Concerns. The NUG continues to be concerned over the inefficiency introduced by splitting the NSSDC staff among different buildings.

For the NSSDC Users Group:

Fred Bruhweiler
Robert Benson
Paul Lowman
Ian Richardson
George Rossano

6 April, 2009