Astrophysics
Division

Report of the
Astrophysics Archival Senior Review
For the
Astrophysics Division, Science Mission Directorate

June 17-19, 2008

Senior Review Panelists

Joel N. Bregman (Chair), Aaron Evans, Terry Herter, Kenneth Janes, Matthew Malkan, John A. Nousek, Michael Nowak, James Schombert, and Mel Ulmer

Introduction

NASA's Science Mission Directorate (SMD) regularly requests comparative reviews of Mission Operations and Data Analysis (MO&DA) programs with the goal of maximizing the scientific return. The MO&DA programs include both operating missions as well as archival data centers and both components of the portfolio were reviewed by separate panels, with this review directed to the archival component.

The review comes at a time when there are a number of changes occurring in the use of astronomical data sets. The scientific community continues to use more diverse datasets in multiple wavelength regimes to perform their research, so there is an increased importance in the effectiveness with which this range of data sets can be utilized. Also, some of the NASA Great Observatories are reaching the stage where Legacy archive data products will be delivered to data centers. These Legacy products are likely to be of enormous value, so the production and curation of optimum data products may require particular care.

A broad review of NASA archival centers was undertaken recently by the National Research Council, which led to the report "Portals to the Universe". This report, which is aligned with the views of the Science Mission Directorate, recognizes that the Archive Centers are a valuable and lasting resource. To optimize the value of these resources, this report recommended a change in the paradigm of peer reviewing and funding. In particular, a "level of effort" is no longer a viable model for archival centers, and resources must be justified through specific goals and tasks. Furthermore, these centers should continue to move in the direction where they facilitate science, as opposed to just provide spinning bits.

The Purpose and Charter of the Review

There are several avenues through which the Archive Centers plan and optimize their activities. Most of them have working groups, users groups and user surveys that make recommendations to improve the holdings and analysis tools offered by individual centers. These centers may also respond to the needs of a mission or to requests or suggestions by the SMD. A complement to these activities is provided by the Archival Senior Review, which occurs every 2-4 years. The Senior Review evaluates Archival Center proposals for continued and augmented funding for a number of significant and important projects. The Review evaluates the proposals in an absolute sense as well as in a comparative sense, which is necessary in times of constrained funding. In The Charter, the SMD instructs the Senior Review to carry out the following:

(1) In the context of the science goals, objectives, and research focus areas described in the Science Mission Directorate's Science and Strategic Plans, rank the scientific merit on a "science per dollar" basis – based upon the expected 2009 and 2010 returns from these reviewed projects.

(2) Assess the cost efficiency, technology development and dissemination, data collection, archiving and distribution, and education/outreach as secondary evaluation criteria, after science merit/usefulness.

(3) Based on (1) and (2), provide findings to assist with an implementation strategy for Astrophysics Division data curation and archiving for 2009 and 2010, including an appropriate mix of:

- continuation of projects as currently baselined;
- continuation of projects with either enhancements or reductions to the current baseline;
- consolidation of projects and activities to enhance efficient management of limited budgetary resources.

(4) Make preliminary assessments equivalent to (1), (2), and (3) for 2011 and 2012.

The Senior Review reports to Jon Morse, Director, Astrophysics Division, Science Mission Directorate, NASA Headquarters.

Programs to be Evaluated

Eight archive centers were contacted to prepare proposals for the Senior Review. They are:

- The High Energy Astrophysics Science Archive Research Center (HEASARC) and the Legacy Archive for Microwave Background Data Analysis (LAMBDA)
  (These two have merged and are considered together.)
- The SAO/NASA Astrophysics Data System (ADS) and the USA-SIMBAD astronomical database (SIMBAD)
  (These two are closely coupled and are considered together.)
- The NASA/IPAC Extragalactic Database (NED)
- The NASA/IPAC Infrared Science Archive (IRSA)
- The NASA/IPAC/MSC Star and Exoplanet Database (NStED)
  (These three provided a single document with subsections for each of the centers and they made separate presentations, the total of which was equivalent to a single center. However, each component was evaluated individually.)
- The Multimission Archive at STScI (MAST)

The Review Procedure

Each of the Archive Centers provided a proposal in which they described their current state, including their holdings, the services and tools offered, some examples of recent scientific successes, the synergies with other aspects of NASA science, and the relationship of the archive to the high-level goals of NASA. The proposals present the new and ongoing initiatives for the next four years, along with budgets, and in most cases, descriptions of the FTE levels and costs for hardware.

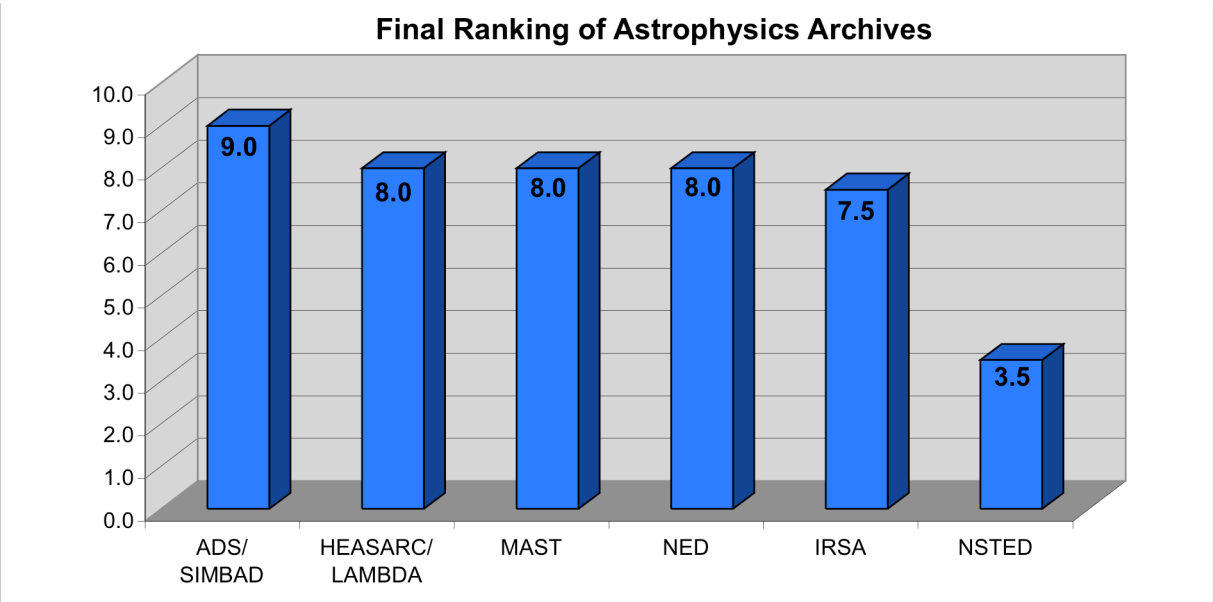
The review began with a Division Overview presentation by Jon Morse and followed by a detailed set of instructions for the review by Douglas Hudgins. Budget guidelines were provided, along with a discussion of a variety of budgetary considerations. An in-guide budget was given to each mission, which was usually a continuation of previous activities with room for modest expansion, at best.

The separate archival centers gave presentations over a two-day period on June 17-18. Each archive center (or grouping of sub-centers) was represented by 3-4 people who gave presentations that were scheduled for 90 minutes each, including questions. In practice, the presentations were interrupted with questions and the discussion usually extended beyond the nominally scheduled time. An executive session followed each presentation and if further questions emerged, the presenters were called back to address them. The committee felt that this led to a very thorough examination of each program, with no major questions left unanswered.

At the end of the second day, a preliminary scoring was carried out, based on the In-Guide budget. Subsequently, the values of Over-Guide activities were examined for each of the missions and the most important and cost-effective components were selected for additional funding, should it become available. Some of the Over-Guide activities were deemed to be critical to the long-term survival of certain centers. A final ranking and set of recommendations were made on the last day of the review.

The committee examined the procedures used in the review. We conclude that the length of the proposals was sufficient for most of the important issues to be expressed. Some proposed budgets were insufficiently detailed. For example, some budgets lacked a listing of the number of FTEs and their general tasks. Future reviews might have more in-depth treatment of the software being maintained for data reduction. Likewise, budgets could provide greater detail concerning proposed hardware purchases, including providing cost-benefit analyses of alternative systems. The amount of time for presentations and questions was ample, in contrast to some previous reviews where many more presentations were scheduled into the same number of days. Finally, the guidance by NASA officials was clear and questions, when they arose, were answered rapidly.

Our evaluations and recommendations for the individual archive centers follow in order of decreasing rank. In addition, there were certain topics relevant to more than one data center, and they are discussed after the individual archive center sections.



SAO/NASA ASTROPHYSICS DATA SYSTEM, ADS, and USA-SIMBAD (Rank: 1)

The NASA (ADS) provides bibliographic information on articles and preprints in astronomy, instrumentation, and physics to the great majority of U.S. astronomers. It is the premiere astronomical search engine for finding publications on any astronomical subject with links to abstracts and articles. It provides a number of additional services such as citations, reference lists in many formats (including user defined ones), and full text searching on scanned journals. It allows customization to the needs of individual users through myADS which can provide weekly notices on articles satisfying a set of individually programmed criteria.

Relevancy:
ADS is so extensively used by the entire professional astronomy community that it is hard to imagine existing without it. By one calculation referenced in the proposal, the efficiency increase to astronomy research in 2002 through the use of ADS is estimated to be approximately 736 full time researchers (compared with otherwise obtaining the information in libraries). The ADS is an enabling tool for research in all of NASA astronomical research programs. ADS incorporates and adds substantial value to the astro-ph archive.

Science strengths:
As it should be, the ADS is designed for the expert user who requires speed and simplicity but also the power to do meaningful searches. The ADS team is to be commended for having an extremely efficient, low-cost system. By virtue of its pioneering work, ADS was ahead of the curve of astronomical publishing houses in its early years. The team is to be commended for its continued dialog with publishers (as they evolve) to get online data for the ADS database. Most notable is their success in getting Nature and Science to agree to indexing and the publishing of abstracts. By working with the major scientific publishing houses, ADS obtains tables of contents, abstracts, references, and links to articles. ADS also interacts with NED and SIMBAD for object name resolution. US SIMBAD access is arranged via an activity sponsored at SAO in coordination with the ADS activities.

Proposal weakness(es):
As noted in the 2004 review simplicity and speed make the ADS fast but may make it difficult to "separate the wheat from the chaff." ADS still needs to find a way to educate (busy) astronomers on its secondary capabilities and search features to enable them to do this. This may be helped by providing alternate interfaces (new looks?) to the user. The problem of a stable long-term location for a "Dark Archive" is important, but not yet solved. ADS may be reaching the point of diminishing marginal returns in incorporating more and more journals on topics peripheral to astrophysics.

Overall assessment and recommendations:

The Panel applauds the ADS team for providing an outstanding service to the astronomical community and for extending the ADS service to many international mirror sites. The ADS team made the compelling case that the system was "aging." In addition to new hardware, a new (open source) database system should be implemented to ensure reliability, maintainability, and long term stability of the ADS. The Panel recommends that NASA continue to fund the ADS at the full level including the "over-guide" budget.

HIGH ENERGY ASTROPHYSICS SCIENCE ARCHIVE RESEARCH CENTER, HEASARC (Rank: t-2)

HEASARC (High Energy Astrophysics Science Archive Research Center) is the original (1990) multi-mission science data archive for NASA missions and provider of data analysis tools. The HEASARC continues to be highly effective in serving the community by delivering data, tools and services that utilize a standard approach, by maintaining scientific expertise in missions, by providing a standard data analysis environment, and by providing a standard way for storing and accessing calibration information for a wide variety of high energy astrophysics missions.

LAMBDA (Legacy Archive for Microwave Background Data Analysis) serves as the archive and analysis support site for data from NASA's Cosmic Microwave Background survey missions COBE and WMAP, as well as for data from IRAS and SWAS. LAMBDA also serves as a repository for CMB data from past and upcoming suborbital experiments. The archive provides access to analysis tools (e.g., CMBFAST online), and links to mission papers and web sites. As of 2008, it was decided to formally merge LAMBDA with HEASARC. As such, the two archives were evaluated jointly, but we urge the next Senior Review to assess the success of this merger.

Science strengths:
HEASARC's usage statistics clearly demonstrate that it is an indispensable tool to the high energy astrophysics community and beyond. It offers excellent and easy access to a very large number of data sets, as well as very useful tools for viewing and analyzing them. Specifically, the common format of the data sets and the fact that the analysis tools can be used on nearly all of them is a great strength. This uniformity of access to current and past data sets is very important to enable comparison of high energy emitting sources across different epochs.

HEASARC's importance is clearly demonstrated by the very large breadth of its holdings (with data from 30 space missions, including 7 operating ones). During the past four years HEASARC has smoothly continued service while quadrupling in size.

The HEASARC web page is easy to navigate yet quite powerful. It allows users to create simple or advanced searches and the results are displayed in pages with tabs that include both the query and results information. Results can be sorted, cross-correlated, plotted, or exported, to mention a few of the post-search options. This is an excellent web site for archival data exploration, and has several capabilities that are unmatched by the other archival data sites.

The panel was impressed that arguably all successful high energy astrophysics mission proposals now pro-actively contact and coordinate with the HEASARC on their own initiative, based on the proposing teams' appraisal of the importance of providing a HEASARC compatible archiving strategy.

COBE and WMAP provide unique datasets that have had a profound impact on Cosmology. WMAP will continue to provide improved sky maps, and the inclusion of suborbital data will further strengthen the importance of LAMBDA. Additionally, the archival CMB data can potentially be used for non-CMB science.

Relevancy strengths:
HEASARC is relevant to NASA's goals, as it archives and manages NASA mission data, as well as other data that are scientifically relevant to the analysis and interpretation of NASA mission data.

Data accessibility:
The data are very accessible, and the interfaces to the data are clear and well documented. The HEASARC supports the Virtual Observatory protocols for data access (and was presented in usage statistics prepared by another project as the most frequently utilized VO data source).

Proposal weakness(es):
The merger of the HEASARC and LAMBDA has occurred too recently for any meaningful assessment of the effects of the merger. The panel has prepared its assessment based on a unified HEASARC, but suggests that attention be paid during the next Archive Senior Review to the success of the LAMBDA portion within the HEASARC environment.

The panel was not completely convinced about the proposed 'Topical' archive approach, but the requested augmentation is modest, and the panel is willing to defer to the advice of the HEASARC User Group to support this initiative, subject to review at the next Archive Senior Review.

The panel encourages the HEASARC increase its innovation efforts. Its dedication to long term software and data products is admirable, but this philosophy should not prevent innovation in future development.

The panel suggests that the LAMBDA archive actively coordinates with other NASA archives holding CMB data to prevent unnecessary duplication.

Overall assessment and recommendations:

The panel's overall finding is that the HEASARC is an excellent service. We recommend full funding for the In-Guide budget at the requested level. We specifically endorse the requested hardware renewal proposed for the FY2012 year.

The panel also felt the HEASARC's proposed enhanced activities are important and should be funded if Astrophysics Division funding allows. In particular, the Near-term International Mission archives should receive high priority for full funding. The Science-Based Topical Archive is desirable but at a lower level of priority. The Visiting Scientists enhancement was not considered by the panel because the Astrophysics Division is considering a comprehensive treatment of such scholarly programs, so the panel neither endorses nor rejects this request.

The formal merger with HEASARC will allow LAMBDA to remove some redundancies from its servers (e.g., coordinate converters), share infrastructure support, and incorporate the CMB data and images into the HEASARC BROWSE and SkyView services. The panel supports the goals of this merger, and encourages the teams to continue to look for ways in which the merger can be used to increase efficiencies and reduce costs, while expanding services and avoiding redundancies between centers holding CMB data.

MULTI-MISSION ARCHIVE AT SPACE TELESCOPE, MAST (Rank: t-2)

The Multi-mission Archive at Space Telescope Science Institute (MAST) is NASA's UV/Optical space astrophysics data archive center. While its primary charter is the maintenance of thirteen missions, it is also the source of several data analysis development projects (HLA tools, GalexView).

Science strengths:
MAST provides public access (via the internet) to multiple missions (e.g., HST, FUSE, GALEX, EUVE). The scientific importance of the rapidly growing GALEX archive is clearly very large. New datasets from the XMM Optical Monitor, Kepler, and the Swift UV/Optical Telescope are being made available to the community for a relatively small investment of resources.

Both usage statistics and science output demonstrate the great utility of MAST. The panel was convinced that archival research with the MAST datasets will indeed have at least as wide a scientific impact as the initial work done by the guest observers.

Providing the same quality of high level processed data that is available for the `Legacy' projects will enable a great deal of science for the remaining 97% of the data archive. The development of the HLA advanced rapidly when a window of opportunity opened briefly before SM-4, while the leading experts are still available at STScI. The live demonstration of the power and speed of the HLA interface convinced panel members that it may be the best multi-mission data browser in existence, and could usefully include other mission footprints. The Flex-based GalexView appears to be an excellent platform for development. The use of VO standards is a major plus.

Proposal Weakness(es):
HLA functions such as searching the database for variability, or by using keywords, would be useful, but might be considered an add-on during a constrained budget environment.

While there was enthusiasm for HLA and GalexView, the panel expressed concern that the construction of the HLA will indefinitely freeze the Hubble datasets at a level that may satisfy a majority of the community, but fail to address some future science projects. Past missions have also demonstrated that locking down of a dataset results in the end of software reduction/calibration support, and the fatal loss of institutional memory on issues such as details of the conversion of Level 0 data (raw data), or even Level 1 data (basic processing), to science (Level 2 or 3) data. The top priority must be fully archiving the data and software for the longest possible lifetime.

There was some criticism of the current MAST web page in that resources can be confusing or difficult to find. For example, “High-Level Science Products”, which includes “HST Treasury, Archival Legacy Programs” is entirely different from the “Hubble Legacy Archive” button on the same page, despite the name similarity. It can be unclear where a user should begin for some HST data queries, as there are several possibilities: the basic MAST search; buttons for “Hubble”, “Hubble Legacy Archive”, “HSTonline”; and “High Level Science Products”. There is little guidance as to how best a user should proceed. Different lines of inquiry can lead to different versions of a data product (e.g., WFPC2 Associations), and for some queries, there can be apparent duplications (sets of ACS data combined into more finished products). We hope that the current user survey will lead to improvements in the web site.

The science benefit of devoting (a modest amount of) resources to Google Sky/World Wide Telescope collaboration was not clear.

Overall Assessment and Recommendations:

MAST provides vital services to the Astronomical community, and the Panel recommends continued support for this important facility through the in-guide budget. We also recommend Over-Guide funding to allow for a substantial part of the continuation of the Hubble Legacy, including GalexView. On the other hand, some augmentations, such as time-domain developments, were less compelling because superior time-domain data will be coming from ground-based observatories.

NASA EXTRAGALACTIC DATABASE, NED (Rank: t-2)

The NASA Extragalactic Database (NED) is a powerful tool to "organize and provide access to both the known contents and cumulative knowledge about our Universe".

Science Strengths:
The NED database, as well as the power to search that database, is continually expanding. The goal of incorporating as many multi-wavelength catalogs of ground based data as possible, if achievable within manpower constraints, is very useful to the scientific community. The enhanced tools that NED has made available are quite useful: velocities relative to the heliocentric; the local super cluster; etc.. The proposal effectively laid out the hardware requirements to achieve these goals.

The proposal clearly laid out the need for migrating to an open source software system, and the need to make enhancements to the search queries. The software development plan to achieve this goal was well-justified.

Relevancy Strengths:
The NED database facilitates astrophysical research directly relevant to NASA goals. NED provides the cornerstone to the extragalactic component of the NVO.

Proposal Weaknesses:
The panel is concerned that NED may be trying to ingest too much data too quickly. The case for closing the ingestion gap to 4-6 months was not made. Possible overlaps in providing direct data transfers from the journal publishers in collaboration with the ADS were not clearly identified.

There was insufficient discussion made of only archiving a subset of enormous catalogs in order to reduce costs in manpower and hardware. The panel found that all the archives are using very high-quality data storage with 24/7 maintenance contracts, which has a significant cost impact. The panel would like the archive centers to consider whether it is sensible for large but less essential data sets to be served with less expensive storage and less comprehensive maintenance contracts (i.e., 9-5 M-F). While this means that not all data would be available all the time, the impact on scientific research might be minor.

The proposal also did not discuss the issue of quality or reliability flags. For example, the classification of a type of object, such as a galaxy versus a QSO, can have various levels of

reliability. Thus if a person is asking how many objects of certain classes exist above redshift of 0.4, the answer may be misleading without any methods to set quality flags.

Overall Assessment and Recommendations:

NED provides an important data service to the community that needs to be maintained. The panel specifically recommends that the hardware be maintained at least at the minimal (life support) level as described in the proposal.

The over guide tasks of migrating to an open source query package is strongly encouraged. The open source project should not be done in vacuum, however, and the panel strongly urges consultation, if not direct sharing of code with ADS, SIMBAD, HEASARC, MAST, IRSA and the VO. The panel also notes that NASA funds other programs, specifically ADP and AISRP, which create content and software that might be directly relevant to NED. The panel strongly encourages the NED team to investigate the possibilities of leveraging the work being done within the context of these other programs.

The panel is cognizant of the fact that the storage of new catalogs, especially those that come from the ground with nearly unlimited bandwidth capabilities, comes at a cost, which can perhaps grow nearly without bounds when datasets such as Pan-STARRS and LSST come on line. Thus, the cost of ingesting all of the data from these newer larger missions could become prohibitive. The panel therefore strongly urges the NED team to define some guidelines for a cost/benefit analysis to determine which databases should be incorporated locally.

INFRARED SCIENCE ARCHIVE, IRSA (Rank: 5)

The Infrared Science Archive (IRSA) is the NASA repository for IR and sub-mm data sets obtained by NASA's IR missions. IRSA will incorporate the Spitzer Archive Data center, which will have a major impact on the community.

Science strengths:
IRSA provides access to a wealth of data from past and existing missions such as IRAS, 2MASS, MSX as well as plans for integrating new missions from Spitzer and Planck. IRSA is located on the Caltech campus with a diverse collection of quality scientists and development teams. IRSA/IPAC has a long history with the IR community and a superior set of user groups for advice and support.

IRSA provides imaging, photometry, spectroscopy and time-domain data as well as a suite of analysis tools (e.g., SPOT, getCal). IRSA also maintains the Level 0 and Level 1 data reduction codes, which provides the community with information for science publications and proposal development.

IRSA has been designated to become the Spitzer data archive. The panel was in unanimous agreement that this task should be one of IRSA's highest priorities during the next funding cycle. This effort will require not only the absorption of the archive data itself, but also the expertise and knowledge of the Spitzer instrument suite currently provided by the mission.

Relevancy strengths:
IRSA's archiving efforts support a significant portion of the astronomical community and have demonstrated a strong impact on astrophysical knowledge across the entire range of astronomical topics. Publications from IRSA data are only superseded by MAST.

Data accessibility:
IRSA provides access to their data holdings through the usual web portals. The web portal was redesigned in the past few years and it is easier to use with a straightforward results page that can be queried further. The panel reports a number of "glitches" during normal usage. Also, in some ways, the front end tools do not match the flexibility and power demonstrated by other data centers (e.g., complex queries; post-query services).

IRSA is rich in the diversity of its datasets. While this provides the strengths of shared costs, economies of scale and depth of knowledge, it also presents a lack of coherence in its presentation to the community. This has resulted in a number of different software interfaces that are needed to process Level 0 and Level 1 data from the various missions. This is perhaps a symptom of different development teams following differing user group recommendations. The IRSA was never charged to provide a uniform set of data processing tools, yet the current variety of software presents an unwelcome barrier to the user.

Proposal weaknesses:
By far the most difficult part of this proposal was the lack of any plan or outline for the Spitzer archive within the In-Guide request. Whereas the panel understood that IRSA is being directed to absorb the entire Spitzer holdings, the lack of any contingency plan for integrating some portion of Spitzer expertise and tools within the In-Guide limits places the entire archive/data center budget in jeopardy. Their response to archiving the Spitzer data is an over-guide request for a doubling of the budget. There seemed to be a lack of community input on this plan.

The panel questions the wisdom of splitting the Planck archive from the instrumental expertise that generated the data.

Spitzer data and other IR data are commonly accessed via two different tools, *RADAR* and *Spot*. The panel suggests that the archive provide a unified method of data access.

Overall assessment and recommendations:

The in-guide request is endorsed by the panel. However, maintaining the long-term health of the Spitzer archive is an absolute priority that must be funded. The staffing of this archive was significantly higher than the panel judged was required, or allowed by the available funding. The panel estimates that the individual instrument support staff could be approximately cut in half compared to the Over-Guide levels.

NASA STAR AND EXOPLANET DATABASE, NStED (Rank: 6)

The NASA Star and Exoplanet Database (NStED) is a newly-created archive of data dedicated to extra-solar planets, to their host stars and to the much larger population of nearby stars. The purpose for the archive is to provide a resource both for planning exoplanet observing programs and for broader astrophysical studies of stars that might (or might not) harbor planets.

Science strengths:
In the near term, the database consists of a systematic collection of information on the nearby stars and a developing archive of exoplanet data from ground-based and space-based programs (MOST and CoRoT). Early US access to the micro-magnitude precision CoRoT data will enable detailed studies of stellar activity, rotation and seismology. By merging overlapping ground-based transit surveys, it may be possible to uncover more transiting planets. In the longer term, as the archive grows, transit and Doppler archives can be mined to uncover additional planets.

Relevancy strengths:
A major goal of NASA is to understand how planets and stars form and whether life exists elsewhere. NStED aims to support that goal. A proposed SMEX mission, TESS, plans to use NStED for its archive.

Data accessibility:
NStED is built on the model of (and with help from) NED and IRSA. It supports multi-faceted queries on a number of stellar and exoplanet parameters.

Proposal weakness(es):
NStED is a very small database compared to the other NASA archives, and is likely to grow only slowly for some time. At the same time, the proposed budget is large in comparison to the size of the database, the expected use of the archive and the near term demand for tools to use on these data. The proposal does not make clear how much value will be added to the archive with the CoRoT data, which will be publicly available, at least at some level, directly from the CoRoT science center. Although the exoplanet field is growing rapidly, it is still in its infancy, and the number of potential NStED users in the short term, at least, is small compared to users of the other archives. The Kepler archive will be located at MAST and other large space exoplanet projects are many years in the future.

Overall assessment and recommendations:

The rapid development of exoplanet research makes it difficult to predict what the needs will be more than a few years into the future. The proposal suggests a review of the NStED progress in two years. We endorse that suggestion and recommend that NStED be funded until then at a considerably reduced level from the requested budget, sufficient to continue the collaboration with the CoRoT project and some accumulation of ground-based transit and spectroscopic data.

The panel found the proposal to be weakest of all reviewed. Most of the panel recommends that NASA considers alternatives that include creating an exoplanet archival opportunity opened to a competitive selection process. Finally, we note that this effort is different from all other archive data centers. The other centers provide data and research tools, but do not define the science to be pursued by researchers, whereas NStED defines the science to be pursued at their site.

Issues of Relevance to Multiple Data Centers

A Common Philosophy of Archive Centers

The archive data centers share a common requirement that data, in Level 0 (and higher) form, be preserved far into the future. This ensures that the bits will be readable decades from now. However, there does not appear to be a uniform policy with regard to the software that one might use to reprocess Level 0 data. Different archive centers have taken different approaches.

The HEASARC maintains a software suite, FTOOLS, which is a set of software tasks primarily written in C and executable from a command line on a variety of platforms. This choice was made to provide a long life to the software and it appears to be working, partly because other high-energy missions were encouraged to and have adopted FTOOLS for their data processing.

The MAST site has software that can process the HST holdings, but many of these tasks were written under IRAF format, an environment that the astronomical community will not maintain into the future. They have rewritten some but not all of the tasks into a non-proprietary format that can be maintained more easily. This opens the possibility that a dozen years from now, certain software tasks needed to process Level 0 data, or even Level 1 data, will not exist. There may not be a great demand for some of these tasks, as MAST envisions that most needs will be served by the higher level data products in the Hubble Legacy Archive. The decision of which tasks to not rewrite appears to be based on demand.

The IRSA center has ingested software from several missions, which do not use a common software base. Not only is there a lack of uniformity in the software, but most of it is based on IDL, a proprietary language. These shortcomings are not the fault of the IRSA, who have dealt heroically with the software packages they were expected to ingest. Rather, the shortcomings have to do with software decisions made by other observatories with little regard for the function and needs of IRSA. Often, choices for software development are due to cost constraints during the early stages of a mission rather than to usability and maintenance costs a decade later. However, without a long-term commitment to the maintenance of this array of software, and the graphical user interfaces to which they are often attached, it may become difficult to process some Level 0 data, or even Level 1 data within the next decade.

The current situation has arisen because NASA and their partner institutions have not agreed upon a common philosophy in the development of software and the delivery to the archive sites. Also, they have not agreed upon the length of time in which the software should be maintained. There are possible solutions to both of these issues. For example, NASA may require that the data system delivered to archive centers be written in a non-proprietary, and perhaps common, format. NASA may desire to require that these systems, in as much as possible, be built from modular, publicly available components that are addressable by one of the commonly available open source scripting languages. Whatever the agreed upon policies, they should be long-lived and have international commitments. Plans should be formulated soon. The archival data sets are astronomical treasures and an inability to reexamine them at some future date should and can be avoided.

Database Management

The costs of commercial database management systems (DBMS) have risen in recent years with the expansion of holdings. This is forcing several sites to migrate their DBMS to open source

DBMS systems, a move that the panel strongly endorses. There are several possible choices for open DBMS systems and there might be an advantage of uniformity between sites. As some centers have already moved to an open DBMS, we encourage the other centers that are planning this move to talk to each other before making their choices.

Public and Private Software Systems

In addition to the software that is being developed within the archive centers, there are several other potential sources of software. In the private sector, Google Sky and Microsoft's World Wide Telescope are the most prominent examples. These already have great appeal to the lay public, so they are ideal for EPO activities. However, these developers have stated that they expect their software will be suitable for research astronomers as well. While this ambition is unrealized at the present time, this could change given the enormous resources at the disposal of these corporations. The archive centers are aware of the current capabilities of Google Sky and the World Wide Telescope and they should continue to monitor developments and take advantage of them if it improves the accessibility and functionality of the archive centers (e.g., maximizing the scientific return from archival data).

NASA funds software development programs, such as the AISRP. These are software development programs and they are not designed to produce "shrink-wrap" software. However, there are projects that have progressed to the level where they are widely used on many platforms, with documentation. Some successes include ds9 (formerly SAO Image; now part of the HEASoft package), COMBAT (for CMB data reduction), or Skyview (now part of the HEASARC). There is no procedure in place for selecting promising software projects and bringing them to a state suitable for inclusion in an archive center. There are additional costs associated with adopting software, such as maintenance and further development. We recommend that the archive centers, in conjunction with the relevant NASA officials, formulate a plan for tapping this potentially valuable resource.

Other types of software have been included or made available at some data archive centers, such as scripts that perform a series of complex tasks. This type of activity should be continued and expanded as possible.

The astronomical community has made effective use of open source software, although it has been in a passive mode. An alternative model is where the archive centers try to encourage volunteer programmers, such as through projects on the Sourceforge effort. Critical components should not be developed with this model, but more long-range or auxiliary projects could be developed in this way. This could be an avenue for bringing AISRP projects from an early state to a user state. This is an unconventional model that may require NASA to reconsider how it will spend resources (i.e., partial support for someone overseeing a public project), but it offers skilled programmers the opportunity of being involved in NASA endeavors, which may be quite appealing (note, for example, the number of people who became involved in reducing SETI data).

Toward A Virtual Astronomy Observatory

We support the concept of a Virtual Astronomy Observatory (VAO), and as there are proposals under review at this time, we will only comment upon this topic from a limited NASA-centric perspective. Each of the archive centers has been working on components of a VAO and many queries can now be performed using VO standards. One of the goals of a VAO is to be able to

access data holding across archive centers, using a single portal. The individual portals of each archive center have become more capable with time and the panel felt that one or more might be mature enough to serve this purpose. We encourage NASA, with input from the archive centers and the community, to select one or more portals for this type of development (or a merger of the best parts of different portals). While this is only a small part of the capabilities that would be offered by a VAO, it has the potential of being broadly used and valuable in advancing scientific endeavors. As a caution, a danger that must be avoided is that the selection of a global portal should not prevent the development of future portals that may offer a number of improvements.