A White Paper on US WEB-Based Astronomical Services

Barry F. Madore
Carnegie Observatories

While astronomy and astrophysics are intrinsically multi-wavelength disciplines, they have been driven in many instances by advances in detector technology that are highly wavelength specific. In addition, the combined opacity of and re-radiation from the Earth's atmosphere has required that certain parts of the electromagnetic spectrum can only be observed from space. Here again wavelength has largely divided (NASA-funded) space-based x-ray, gamma-ray, ultraviolet and far-infrared astronomy from (NSF and private-funded) ground-based optical, near-infrared, millimeter and radio astronomy. This has not only led to a community of radio astronomers, optical astronomers, x-ray astronomers, etc., but it has produced catalogs of radio sources, optical sources, and x-ray sources ... many (most) of which are the same objects discovered, named, re-discovered and re-named, time and again, with and without the knowledge of other astronomers working in other wavelength domains. The discovery lists and catalogs necessarily focus on wavelengths; the discovery science and physical interpretation should focus objects. Consolidating all of the observations of all of the celestial objects made at all of the currently accessible wavelengths has yet to be accomplished. Opportunities abound, and discoveries await the merger of datasets already in hand. The question is, given limited resources and an avalanche of data and datasets, how best to proceed?

For more than 20 years now NASA has been successfully operating a number of web-based archives supporting wavelength-specific missions and datasets, as well as funding and operating topical archives in support of multi-wavelength astrophysics. Under contracts with NASA SAO operates the Chandra X-Ray Observatory (CXO), AURA operates the Space Telescope Science Institute (STScI) and JPL/Caltech operates the Spitzer Science Center (SSC), all of which support community-wide electronic access to active mission datasets, pipeline processed and archived from operating missions. On the other hand, NASA's wavelength-specific archives, IRSA, MAST, LAMBDA and HEASARC, serve legacy datasets from completed missions such as IRAS, 2MASS, MSX, ISO, IUE, Einstein, FUSE, UIT, HUT, EUVE, XMM, RXTE, Swift, GLAST, WMAP, etc. The NASA Astrophysics Data System (ADS) provides unlimited on-line access to all of the current and

historical astrophysical literature. The NASA/IPAC Extragalactic Database (NED) provides multi-wavelength merged data and search capabilities covering all of the extragalactic sky, including the (space-based and ground-based) catalogs generated by IRAS, 2MASS, ROSAT, APM, NVSS, etc., and most recently 150 million sources from the Sloan Digital Sky Survey (SDSS), as well as all extragalactic objects published in the regular refereed astrophysical literature. The latest addition to NASA's on-line facilities is the NASA EXoplanet Science Institute (NExScI) and its on-line database service called NStED, concentrating on tools and datasets for the exoplanet community. Only SIMBAD, which provides basic data on stars (primarily), is not directly funded by NASA, although access to it by means of North American mirror sites is, in fact, paid for by NASA funds. Access to the sky and its contents is incredibly well covered by NASA's mission-specific, wavelength-specific and topical/theme-specific data centers.

These services are on-line and operating now. They have been well tested, they are widely used, and they are robust. They have their individual prime missions and mandates, but they intercommunicate and cooperate. The Astrophysics Datacenters Executive Committee (ADEC), which reports to NASA Headquarters, has been meeting on a regular basis for almost a decade now, and acts as a focal point for discussion and coordination of related efforts and as a means of directly informing and engaging NASA management and its advocates. All of the data centers and archives have independent user panels and they are all externally reviewed competitively every four years in NASA's on-going Senior Review process. This is a mature and functioning alliance of dedicated data providers serving the broader astronomical research community.

Some years ago NASA management solicited a plan for even more advanced interoperability between its archives and data centers. The ADEC prepared and presented that plan under the name of "Celestial Navigator". However, with the pre-emptive promise of the NVO and its proclaimed broader terrain, which encompassed both NASA and the NSF, that NASA-specific plan was not implemented. Ten years later the relevance and need for "Celestial Navigator" remains unchanged. The basic plan is still viable and could be implemented without delay. The decade-long NVO effort has provided protocols and broad aspects of a supporting infrastructure on a global scale. Early on HEASARC provided the community with a tool (Astrobrowse, aka DataScope) that interrogates existing databases "to find

everything that is known about a given target or region of the sky." Having the means to find data is not the problem; having the data on-line and in an intelligible and useable format is still problematic. What is required now is a broadly coordinated, but locally motivated, implementation of pipeline processing and on-line availability of science-ready data products at a site-specific level.

NASA is, and has been moving aggressively to provide rapid and broad-based on-line support for the access to and distribution of its science legacy and active-mission data products. It is clear that the next generation of ground-based telescopes need to move in this direction also. ALMA is clearly on this path as well. Other facilities, especially private observatories, have a wide (and expensive) gap to close. However, NSF-funded ground-based facilities (private and public) have a long way to go before matching the complexity and sophistication of NASA's data archives and mission centers. What is also lacking is a well defined link between highly processed datasets and published papers. To fully exploit the science content and discovery potential of legacy datasets, ground-based facilities (both public and private) as well as the publishers of scientific data need direct support and targeted funding.

What is clear from the NASA experience is that a highly structured, top-down approach does not work. The total failure of such systems (the original Astrophysics Data System in the US and its equivalent ESIS in Europe) in the late 1980's is testimony to this. The difficulties encountered in trying to manage AIPS++ as a globaly distributed software development activity may be relevant here too. A light-handed coordination of locally controlled, science-driven component parts (the individual data centers) does work and has been been successfully deployed across the country by NASA for more than a quarter of century. Rather than attempting to impose a massive new structure onto a successful existing system it would seem prudent to have ground-based systems emulate NASA's approach, and at the same time have both agencies additionally fund smaller, innovative activities that are designed to advance the scientific use of these facilities in a targeted way, building laterally as technology and science opportunities move ahead more rapidly than can be easily predicted.

Having a strong and reliable network of databases and archives, interconnected, but independently evolving, can serve the community and serve it well if there is also the means for anyone within the broader scientific/astronomical

community or within the existing centers to react quickly, build components and offer extensions to a system that is dynamic, open and ultimately science-driven. A long-term grant program designed to foster this new alliance within the astrophysics community, between the data providers, the data archivists, the data publishers and the data users (including theorists) would be a natural and welcome evolution of a mature and functioning system that is now and has been for many years well beyond virtual.

Is this just a plea for more infrastructure? No, it is a statement that we are at a unique point in the history of astronomy and astrophysics where legacy datasets spanning the entire electromagnetic spectrum are competitive with "the sky is our archive." What has already been archived can and should be preserved and productively brought to bear on current and future science. What remains to be observed must be able to easily tap into those precious datasets that cost so much time and money to acquire.

And we are tantalizingly close. By assuring that NSF and public/private ground-based facilities have the funding to pre-process and archive science-ready data, place it on-line and make it easily accessed and universally available, is essential. By enabling the journals to become more involved in the archiving of and/or linking to high-level data products as they are published would, in combination with NASA's existing mission-based and thematic archive systems, close the scientific loop. And then, from the ground up, we will have a system that makes astronomy and astrophysics truly panchromatic and multi-wavelength both in concept and in practice.