NOAO Mini-Workshop: Mining Observatory Archives

Friday January 6, 2017   2 - 3:30 PM,   San Antonio 4

ABSTRACTS

Andy Adamson and Andre-Nicolas Chene

Gemini Observatory
Gemini Observatory Publications, Statistics and Archive

We describe the statistics of Gemini refereed publications, including relative productivity and impact of instruments and observing modes, and overall statistics such as the total publication count. We identify factors which may influence the probability of a publication emerging from a given observing program. At present, only a small fraction of publications arise purely from archival data. We present some of our plans for post-observing community support, and solicit input on various options for increased productivity in archival research.

Harry Teplitz
IPAC

Astrophysics Archives at IPAC

The Infrared Processing and Analysis Center (IPAC) at Caltech provides vital science and data archives for astronomy missions. IPAC operates the NASA/IPAC Infrared Science Archive (IRSA), the NASA/IPAC Extragalactic Database (NED), the NASA Exoplanet Archive, and the Keck Observatory Archive. Through these projects, IPAC operates archives for NASA, NSF and privately funded projects, including: Spitzer, WISE, NEOWISE, the W.M. Keck Observatory, the Zwicki Transient Facility, and the US archive for Planck. NED provides efficient access to an extensive synthesis of extragalactic data combining and standardizing key measurements from NASA missions, sky surveys, and the astrophysics literature. The NASA Exoplanet Archive is an online astronomical exoplanet and stellar catalog and data service that collates and cross-correlates astronomical data and information on exoplanets and their host stars, and provides tools to work with these data. The guiding principle of IPAC's archives is to enable cutting-edge science through strategic response to the evolving needs of the user community.

By supporting multiple astronomy archives at one center, IPAC has the advantage of the scientific and technical synergy between them. We provide powerful services to researchers that combine data sets across the infrared sky. IPAC services offer interoperability with other archives through program-friendly interfaces and the use of Virtual Observatory protocols. In total, IPAC manages more than 10 PB in the data center, including databases containing almost 200 billion rows. The data volume is expected to double in the next five years. IPAC meets these "big data" challenges through creative solutions and technical innovation.

Scott Fleming
STSci

MAST

The Mikulski Archive for Space Telescopes (MAST) is a NASA-funded project to provide, and support, a variety of astronomical data archives, with a primary focus on UV, optical, and near-IR data. MAST archives data from more than 20 different missions, including GALEX, Gaia, Hubble Space Telescope, Kepler, K2, PanSTARRS, Swift, and XMM. Looking to the future, we will be the primary archive for the James Webb Space Telescope,

TESS, and WFIRST. In this presentation, I will highlight some recent services and data products designed to enable archival astronomical research. I will introduce three tools that enhance GALEX and HST data in new ways: gPhoton, the Hubble Legacy Archive, and the Hubble Source Catalog. I will also take you on a tour of the MAST Discovery Portal, which enables cross-mission searches and data discovery across not only the MAST-supported missions, but data available at CDS and the Virtual Observatory as well.

K. Olsen, M. Fitzpatrick, M. Graham, L. Huang, S. Juneau, D. Nidever, R. Nikutta, P. Norris, and P. Wargo
NOAO

The NOAO Data Lab

The NOAO Data Lab aims to provide infrastructure to maximize community use of the high-value survey datasets now being collected with NOAO telescopes and instruments. Upon its release in mid-2017, the Data Lab will allow users to access and search databases containing large (i.e. terabyte-scale) catalogs, visualize, analyze, and store the results of these searches, combine search results with data from other archives or facilities, and share these results with collaborators using a shared workspace and data publication service. Prototype versions of Data Lab capabilities are being used at AAS 229 to support the SMASH DR1 data release, including a custom Data Discovery tool, database access to the SMASH catalog, a Python query interface to the database, an image cutout service, and a Jupyter notebook server with example notebooks for exploratory analysis. With the public release of the Data Lab, we will provide a capable framework to support science analyses of select high-value datasets, as well as provide templates to aid users in publishing their own NOAO survey-scale datasets.