AAS 237 - Astro Data Lab Special Session

Topic: 224. The Data Lab Science Platform and Open-Data Ecosystem at NSF's NOIRLab

# Date / Time:

Tuesday Jan 12, 2021 **04:10 PM Eastern Time** (US & Canada) (= 2:10pm AZ)

# Schedule

1- **Robert Nikutta**: "Astro Data Lab Overview: Mission, Data & Services"

2- **Dara Norman**: "Moving toward Integrated NOIRLab Data Services"

3- **Chien-Hsiu Lee**: "Time-domain analysis with the ANTARES event broker"

4- **Knut Olsen**: "Connecting catalogs to image pixels with the Astro Data Archive"

5- **Raga Pucha**: "Joint Spectroscopic & Photometric Analyses of Low-Redshift Galaxies"

6- **Alex Drlica-Wagner**: "Science with the Dark Energy Survey at NOIRLab"

7- **Decker French**: "Using Data Lab to identify Tidal Disruption Events”

8- **Ginny McSwain:** "Undergraduate Research with the Data Lab”

# Session Description

As we keep progressing into an era of increasingly large and rich astronomy datasets, we augment our opportunities for new transformative discoveries. Wide-field surveys generate datasets in tens to hundreds of Terabytes in volume. The number of objects in today’s survey catalogs reach billions per survey, which can be prohibitive for individual researchers’ resources. The Data Lab science platform was therefore developed as a core piece of NOIRLab's data mission to allow researchers to bring their analysis to the data. The “open-sky, open-data” vision of NOIRLab drives our efforts to strengthen the astronomy community by broadening the reach and accessibility of not only the data themselves, but also of powerful software tools and tutorials. Working together, NOIRLab's data services will empower users from all backgrounds to contribute a diverse set of ideas and skills and therefore maximize scientific discoveries from the astronomy community as a whole.

Since Data Lab opened its doors in 2017, we have added high-value datasets, new services, and have seen a growing number of registered users. In conjunction with AAS 237, our latest dataset additions include the second release of the Dark Energy Survey (DES DR2, to be announced at AAS 237), the ninth and final release of DESI imaging Legacy Surveys (LS DR9), and the second release of the nearly all-sky NOAO Source Catalog (NCS DR2). NSC includes almost 4 billion unique objects with 68 billion individual measurements, making it the largest photometric dataset before the arrival of Rubin Observatory's Legacy Survey of Space and Time. Our data holdings can be accessed through various services including catalog query tools, image cutout service, cross-matching capabilities, dedicated user storage space, workflows through a Jupyter notebook server, and via scripted analysis.

This Special Session will start with an overview of the Data Lab, its mission, latest advances and dataset additions before expanding to describe the broader data-intensive context at NSF’s NOIRLab. We will show how the various data-oriented services connect to form an open-data ecosystem. Namely, this session will introduce the new Astro Data Archive (formerly the NOAO Science Data Archive) interface, with its fast and flexible search capabilities, and showcase how to employ it directly with the Data Lab. We will further show how to connect time-domain capabilities from the ANTARES stream alert broker to the Data Lab. The session will also feature highlights from our user community including scientific results as well as an education application in a classroom setting. Data Lab aims to contribute to train the next generation of researchers at the crossroads of astronomy and big data science. We will conclude with a short Q&A during which we welcome questions and suggestions from the attendees.

We hope that the session attendees will leave with a greater understanding of the role of science platforms in astronomy, but also with ideas and inspiration for new research opportunities. Current users are invited to share their research making use of Data Lab or data services at NOIRLab in the associated poster session and/or during the Q&A. This session is open to astronomers, data scientists, and educators from all levels and backgrounds. We seek to keep improving our services to serve the broad astronomy user community.

# Abstracts

**1- Robert Nikutta: Astro Data Lab Overview: Mission, Data & Services**

Astronomical surveys now routinely release terabytes of catalog data, and their image sets are measured in petabytes. Download, storage, and processing are out of reach for individual researchers and many organizations. The natural remedy to this data avalanche are Science Platforms, which combine data archive functions with co-located compute resources. Astro Data Lab, developed at the Community Science and Data Center (CSDC) within NSF's National Optical-Infrared Astronomy Research Laboratory (NOIRLab) is one of the first fully-featured science platforms. We welcome professional astronomers and data scientists, students and educators, and citizen scientists. Our main mission is to provide open access to large survey data products, and to powerful data services through which all members of the community can conduct research at the forefront of astronomy. I will give an overview of the datasets and services offered by Astro Data Lab, and outline how together with their integration into the larger NOIRLab data ecosystem they enable world class research to anyone with a good idea and access to a web browser.

**2- Dara Norman: Moving toward Integrated NOIRLab Data Services**

Two potentially paradigm-shifting trends are taking place in astronomical research. The first is the move towards the use of surveys, large datasets and catalogs that enable a wide range of topics to be studied. The second is the recognition that the astronomical and astrophysical community of researchers must become more inclusive in order to realize the best scientific innovation and productivity. The NSF’s National Optical and Infrared Research Laboratory has been established to maximize support for both these trends through data services. I will discuss how the Community Science and Data Center (CSDC) at NOIRLab is well positioned to be a foundation for these efforts into the 2020s and beyond.

**3- Chien-Hsiu Lee: Time-domain analysis with the ANTARES event broker**

With the avalanche of alerts delivered by the Zwicky Transient Facility (ZTF) and Rubin observatory, it is not possible to manually inspect each alert. Instead, we will rely on brokers to automatically select intriguing alerts that warrant follow-up in a timely manner. Here we present ANTARES, a time-domain broker that is currently digesting and broadcasting ZTF alerts in real time that incorporates user-defined filters to hunt for the rarest of the rare events in the time-domain. In this talk, we will provide an overview of the ANTARES system, how users can use the Data Lab science platform to develop their own filters to identify intriguing transient sources, to connect to the streams of alerts broadcasted by ANTARES, and to explore the ANTARES alert database for long-term variables.

**4- Knut Olsen: Connecting catalogs to image pixels with the Astro Data Archive**

Science with surveys starts with the pixels, and thus critically depends on robust archival services to deliver the pixels where they need to go. The Astro Data Archive is NOIRLab’s new platform for storing and serving petabytes of pixel data to a variety of users. It transports, regularizes, validates, and ingests the observations from all telescopes at CTIO and KPNO, including science observations and calibrations from 18 different instruments across multiple telescopes. It serves raw pixel data to several automated processing pipelines, and ingests the reduced data products back to the Archive. It delivers these processed data to PIs, archival users, and survey teams, who in turn run measurement pipelines from which survey catalogs are derived. In this talk, I will present how the Astro Data Archive enables users to quickly search the millions of files in its petascale volume, including by constraints on any header keyword. Example uses of these searches include retrieving sections of full frame images for comparison with catalog objects and generation of coverage maps for survey footprints, both of which are often important for getting science out of massive catalogs.

**5- Raga Pucha: Joint Spectroscopic and Photometric Analyses of Low-Redshift Galaxies**

The era of big data astronomy has changed how science is done. Astronomers now have access to impressively large spectroscopic and photometric datasets, which offer new opportunities for a broad range of discoveries. However, such increased data volume and complexity also introduces new challenges. Integrated science platforms such as the Astro Data Lab provide a unique opportunity to access and efficiently combine spectroscopic and photometric datasets for astronomical analysis and visualization. In this talk, I will showcase a new spectral access service that is over 200 times faster than the previous tools in retrieving spectra from large collections such as the Sloan Digital Sky Survey (SDSS). I will show how to use the Data Lab to join the SDSS spectroscopic catalogs with the other available photometric datasets to study low-redshift galaxies. In addition, I will also demonstrate how to use Data Lab Jupyter notebooks to visualize images and spectra of these galaxies together. These Data Lab capabilities are going to be even more useful in the coming years with the advent of large surveys such as Rubin Observatory Legacy Survey of Space and Time (LSST) and the Dark Energy Spectroscopic Instrument (DESI) survey.

**6- Alex Drlica-Wagner: Science with the Dark Energy Survey at NSF’s NOIRLab**

The Dark Energy Survey is a six-year optical/near-infrared survey of the southern Galactic cap using the Dark Energy Camera on the 4m Blanco Telescope at Cerro Tololo Inter-American Observatory. DES was designed to probe the nature of dark energy by mapping hundreds of millions of galaxies, detecting thousands of supernovae, and searching for patterns of cosmic structure. The second DES data release, DES DR2, consists of deep coadded imaging of >5000 sq. deg. and >600 million cataloged astronomical sources reaching a coadded depth of >24 mag. These data have already led to revolutionary advances in studies of the outer solar system, the Milky Way galaxy, and the structure and evolution of the Universe. I will provide a snapshot of some of the science that can be done with the DES through the NOIRLab Data Lab interface.

**7- Decker French: Using Data Lab to identify Tidal Disruption Events**

Identifying Tidal Disruption Events (TDEs) with Rubin Observatory will be challenging due to their low rates compared to other transients like supernovae and due to the large variation in the properties of TDEs observed to date. Nonetheless, Rubin Observatory has the capability to dramatically increase the sample of known events from the dozens currently known, discovering about 10 new TDEs every night. One method for identifying TDEs in Rubin Observatory is to use information about the host galaxies to identify likely candidate TDEs for spectroscopic follow-up. TDEs occur at a high rate in post-starburst (or “E+A”) galaxies, at rates of 2x10-4 — 3x10-3 per year per galaxy. I will discuss how a strategy of host galaxy-informed follow-up will enable the discovery of a large number of new TDEs and enable the discovery of a wider variety of TDE behaviors. I will also discuss methods for identifying likely TDE host galaxies in the southern hemisphere for which spectroscopic information will be limited. These programs require large photometric datasets across multiple wavelengths, the ability to cross-match large numbers of sources, and a platform to analyze and classify objects; all of which can be achieved with Data Lab.

**8- Ginny McSwain: Undergraduate Research with the Data Lab**

The NOIRLab Data Lab science platform can be a powerful entry tool for undergraduates beginning research in astronomy. Many of my students are using it to learn Python and introductory data science techniques using the included Jupyter notebooks. The science examples provided on the platform are particularly useful for beginners. A wide variety of science is possible with the Data Lab since users can access any public dataset, not just the NOIRLab legacy surveys, or build their own theoretical models within the platform. The Data Lab is also an important tool for increasing diversity and inclusivity within astronomy research. Students don't have to be able to afford a high-end computer with a fast processor and plenty of hard drive space to be able to process large datasets, and faculty at small teaching colleges don't need to purchase extra computers for their research group. Finally, the Data Lab is easily accessible from anywhere, which has been especially helpful during the pandemic.