

# Sunday, November 5

Week 45 of 2023

- Home
- Moon
- Work
- UA Holidays
- NOAO
- Siri Suggestions
- ADASS 2023 Calendar
- Telescopes

November 2023							December 2023						
S	M	T	W	T	F	S	S	M	T	W	T	F	S
			1	2	3	4					1	2	
5	6	7	8	9	10	11	3	4	5	6	7	8	9
12	13	14	15	16	17	18	10	11	12	13	14	15	16
19	20	21	22	23	24	25	17	18	19	20	21	22	23
26	27	28	29	30			24	25	26	27	28	29	30
							31						

WORKSHOP: Software Carpentry		11/4 to 11/5
8AM		
9AM		
10AM		
11AM		
Noon		
1PM	<b>1:00PM TUTORIAL T001: Wide-field multi-wavelength images and mosaics: creating a visually appealing presentation and building a value-added catalog</b> Where NORTH BALLROOM <a href="https://union.arizona.edu/in-fodesk/maps/sumc_maps.php?lev-">[https://union.arizona.edu/in-fodesk/maps/sumc_maps.php?lev-</a>	<b>1:00PM TUTORIAL T002: LSDB and HiPsCat: Joint Distributed Analysis of LSST-Scale Datasets</b> Where NORTH BALLROOM <a href="https://union.arizona.edu/in-fodesk/maps/sumc_maps.php?lev-">[https://union.arizona.edu/in-fodesk/maps/sumc_maps.php?lev-</a>
2PM		
3PM	<b>3:10PM TUTORIAL T003: Exploring big data efficiently with the Astro Data Lab science platform</b> Where NORTH BALLROOM <a href="https://union.arizona.edu/in-fodesk/maps/sumc_maps.php?lev-">[https://union.arizona.edu/in-fodesk/maps/sumc_maps.php?lev-</a>	<b>3:10PM TUTORIAL T004: Scientific Data Processing with the Pegasus Workflow Management System</b> Where SOUTH BALLROOM <a href="https://union.arizona.edu/in-fodesk/maps/sumc_maps.php?lev-">[https://union.arizona.edu/in-fodesk/maps/sumc_maps.php?lev-</a>
4PM		
5PM		
6PM	<b>6:00PM ADASS RECEPTION</b>	
7PM		
8PM		
9PM		
10PM		

• All-Day Events

■ **11/4 to 11/5 WORKSHOP: Software Carpentry**  
**Location:** Where  
 KUIPER SPACE SCIENCES, RM 330  
[\[https://www.lpl.arizona.edu/about/contact/kuiper\]](https://www.lpl.arizona.edu/about/contact/kuiper)  
**Notes:** Event Description

Software Carpentry [\[https://software-carpentry.org/\]](https://software-carpentry.org/) aims to help researchers get their work done in less time and with less pain by teaching them basic research computing skills. This hands-on workshop will cover basic concepts and tools, including program design, version control, data management, and task automation. Participants will be encouraged to help one another and to apply what they have learned to their own research problems.

\_For more information on what we teach and why, please see our paper "Best Practices for Scientific Computing [\[https://doi.org/10.1371/journal.pbio.1001745\]](https://doi.org/10.1371/journal.pbio.1001745)".\_

WHO: The workshop is aimed at graduate students and researchers at all stages of their education and careers who wish to learn computational tools to increase the reproducibility and efficiency of their work. Participants should have some knowledge of programming (not necessarily Python) and have some familiarity with the shell command line (i.e. navigating directories on the shell command line). Specific knowledge of Python and git are not required.

REQUIREMENTS: Participants must bring a laptop with a Mac, Linux, or Windows operating system (not a tablet, Chromebook, etc.) that they have administrative privileges on. They should have a few specific software packages installed, see details:

<https://abostroem.github.io/2023-11-03-adass/>

---

• Timed Events

■ **1:00PM to 2:50PM TUTORIAL T001: Wide-field multi-wavelength images and mosaics: creating a visually appealing presentation and building a value-added catalog**  
**Location:** Where  
 SOUTH BALLROOM  
[\[https://union.arizona.edu/in-fodesk/maps/sumc\\_maps.php?level=level3\]](https://union.arizona.edu/in-fodesk/maps/sumc_maps.php?level=level3)  
**Notes:** Event Description

With the advent of modern wide-field surveys in different wavelength domains and wide availability of mosaic detectors on ground-based and space telescopes, some of the important challenges are: how to visualize the data and make it easily accessible; how to process multi-wavelength datasets originating from different facilities and having different depth, resolution, and coverage in a uniform way to assemble source catalogs, which would be useful for science. In our tutorial we will present a suite of state-of-the-art technological solutions and algorithms to perform efficient (post-)processing and analysis of wide-field images and present them in a visually appealing way.

Primary learning objectives The audience will learn how to:

- \* Access wide-field survey and archival images using VO protocols from Python, download cutout images
- \* Reproject the cutouts, post-process them (creating RGB color images, perform de-noising and color enhancements)
- \* Configure and run modern image analysis software on them (SourceExtractor++) and create catalogs of point and extended sources
- \* Generate HIPS maps to display in Aladin/AladinLite (including a web-page integration)

Click for [instructions for tutorial T001 participants](#)

# Sunday, November 5

Week 45 of 2023

[<https://docs.google.com/document/d/1To6dPWP3miTYuvf8m-mx8jb9DBXSN992yFJWivhKxdCl/edit>]

## 1:00PM to 2:50PM TUTORIAL T002: LSDB and HiPSCat: Joint Distributed Analysis of LSST-Scale Datasets

**Location:** Where  
NORTH BALLROOM  
[[https://union.arizona.edu/infodesk/maps/sumc\\_maps.php?level=level3](https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3)]

**Notes:** Event Description

The present decade will be marked by growth of large survey catalogs, both in their number and scale. Joint analysis of such catalogs has historically shown itself to be tremendously useful (e.g. enabling multi-wavelength or time-domain studies), with its importance likely to rise even further. Yet, with the increase in scale towards PBs of data, joint analysis – even at a catalog level – becomes a complex data management problem that few astronomers are equipped to tackle with present-day technology. Here we present HiPSCat, a format for efficient and queryable storage of large datasets, and LSDB (Large Survey DataBase), a Python framework that enables distributed cross-matching and analysis of astronomical datasets at LSST scale (O(10B) sources). The HiPSCat format – framework-independent and built

as an extension of the well-known IVOA HiPS standard – provides intelligent (balanced) spatial partitioning and enables scalable serving of PB-scale datasets (via HTTP) using Parquet for efficient storage. The LSDB framework enables distributed computing and cross-matching on HiPSCat-formatted datasets. Leveraging broadly adopted community libraries such as astropy, Pandas, and Dask, LSDB presents a user-friendly API approachable to astronomers. The goal of LSDB is to enable the user to focus on the science aspects of their tasks, leaving the difficult data management aspects (distribution, resiliency) to the framework.

You can find the tutorial notebooks here:  
[https://github.com/swyatt7/ADASS\\_LSDB\\_tutorial](https://github.com/swyatt7/ADASS_LSDB_tutorial) in the nb/ folder. There is also a short presentation in the pres/ folder which will be a rough overview of HiPSCats for the first section of the tutorial.

Primary learning objectives:

- \* How LSDB can achieve your large scale science: Out of the box, LSDB comes equipped with the potential to perform various spatial analysis (like cone-searching, and cross-matching), along with time-series analysis (e.g. large scale lightcurve analysis). Not only will LSDB be equipped with these methods, but since it is built upon the `dask.dataframe` distributed framework, users will be able to define their own functions and map them across the catalogs. If a user wants to use their own cross-match algorithm, they can easily tie it into our HiPSCat framework through the `dask.dataframe` library, which we will provide adequate documentation on.
- \* The strengths of HiPSCat partitioning structure: It enables storage of astronomical datasets in a way that equalizes the number of rows per partition, yet keeps spatially adjacent objects together. Once two catalogs are partitioned in the HiPSCat manner, distributed, joint-spatial analysis is trivial between them.

Thank you to everyone interested in HiPSCat/LSDB! If you are here, you are excited about exploring concepts of LSST-sized catalog analysis with our python libraries. Prior to the tutorial, we will have most of the infrastructure already set in stone, so that it will be essentially seamless to get everything started.

### INSTRUCTIONS FOR TUTORIAL T002 PARTICIPANTS

All you will need is:

- \* A computer with an internet browser + internet access
- \* An account on <https://lsst.dirac.dev/> jupyter hub. To get an account please send me (PG4gdWVycz0iem52eWdiOmZxamxuZ2dAaGoucFolj5mcWpsbmdn-QGhqLnxadWvbj4=) an email with your github account username or email associated with the account, and I will grant you access.

Thats it! On the jupyter hub, the users will spin up there own computing instances. There we will have the python environment already

set up, along with the credentials to access to the sample datasets in the cloud. The tutorial notebooks will be git-cloned from [https://github.com/swyatt7/ADASS\\_LSDB\\_tutorial](https://github.com/swyatt7/ADASS_LSDB_tutorial), and the users will follow along from there.

## 3:10PM to 5:00PM TUTORIAL T003: Exploring big data efficiently with the Astro Data Lab science platform

**Location:** Where  
NORTH BALLROOM  
[[https://union.arizona.edu/infodesk/maps/sumc\\_maps.php?level=level3](https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3)]

**Notes:** Event Description

In the era of data-intensive astronomy the community needs to acquire skills to handle increasingly larger and more complex datasets, and to access high-performance computing and analysis tools. In this tutorial we will teach participants how to use data-proximate science platforms to conduct astronomy research. Using the Astro Data Lab science platform and the new SPARCL (SPectra Analysis and Retrievable Catalog Lab) service for spectroscopy, participants will first learn how to find documentation, information about all of Astro Data Lab's data holdings of over 100 TB of wide-field survey catalogs, 2.5 PB of imagery, and millions of spectra from DESI and SDSS, and how to access help from the Astro Data Lab team. We will then teach the group in an interactive mode how to use various data services and analysis tools at Data Lab, including how to crossmatch tables, build and submit catalog queries, search for images and create cutouts, search for and download spectra, and how to use the Astro Data Lab Jupyter notebook server. The participants will execute and modify a number of science-case example notebooks from various domains of astronomy focusing on data analysis. The tutorial will also make use of some amenities on science platforms, including remote file storage and remote user-owned database tables.

Primary learning objectives

- \* Construct SQL queries to query large datasets through a dedicated Jupyter Notebook server, web-interface, and command-line interface
- \* Use Astro Data Lab's X-match web service to crossmatch datasets
- \* Discover and query for Sloan Digital Sky Survey (SDSS) and Dark Energy Spectroscopic Instrument (DESI) spectra with the new SPARCL tool
- \* Create plots with the data obtained to realize the graphical and visualization capabilities within our notebooks
- \* Get image cutouts for a set of objects using an image discovery service and a cutout tool

Register for an Astro Data Lab account (ideally done in advance): <https://datalab.noirlab.edu/account/register.html>

Tutorial notebooks are available here: <https://github.com/astro-datalab/Tutorial-ADASS-2023>

Click for instructions for tutorial T003 participants  
[<https://github.com/astro-datalab/Tutorial-ADASS-2023/blob/main/README.md>]

## 3:10PM to 5:00PM TUTORIAL T004: Scientific Data Processing with the Pegasus Workflow Management System

**Location:** Where  
SOUTH BALLROOM  
[[https://union.arizona.edu/infodesk/maps/sumc\\_maps.php?level=level3](https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3)]

**Notes:** Event Description

Workflows are a key technology for enabling complex scientific computations. They capture the interdependencies between processing steps in data analysis and simulation pipelines as well as the mechanisms to execute those steps reliably and efficiently. Workflows can capture complex processes, promote sharing and reuse, and also provide provenance information necessary for the verification of scientific results and scientific reproducibility.

Pegasus (<https://pegasus.isi.edu>) is being used in a number of scientific domains doing production grade science. In 2016 the LIGO gravitational wave experiment used Pegasus to analyze instrumental data and confirm the first detection of a gravitational wave. The Southern California Earthquake Center (SCEC) based at USC, uses a Pegasus managed workflow infrastructure called Cybershake to

# Sunday, November 5

Week 45 of 2023

generate

hazard maps for the Southern California region. In 2019, SCEC completed the largest CyberShake study to date, producing the first physics-based PSHA maps for the Northern California region. Using Pegasus, they ran CyberShake workflows on three systems: HPC at the University of Southern California (USC), Blue Waters at the National Center for Supercomputing Applications (NCSA), and Titan at the Oak Ridge Leadership Computing Facility (OLCF), consuming about 120 million core hours of compute time. Pegasus orchestrated the execution of over 18,000 remote jobs using Globus GRAM, rvGAHP, and Condor Glideins, and transferred over 150 TB between the three systems. Pegasus is also being used in astronomy, bioinformatics, civil engineering, climate modeling, earthquake science, molecular dynamics and other complex analyses.

In 2020, we released Pegasus 5.0 that is a major improvement over previous releases. Pegasus 5.0 provides a brand new Python3 workflow API developed from the ground up so that, in addition to generating the abstract workflow and all the catalogs, it now allows you to plan, submit, monitor, analyze and generate statistics of your workflow. Since 2022, Pegasus has been a key part of the ACCESS support strategy (<https://support.access-ci.org/pegasus>)

Primary learning objectives

The goal of the tutorial is to introduce application scientists to the benefits of modeling their pipelines in a portable way with use of scientific workflows with application containers. We will examine the workflow lifecycle at a high level and issues and challenges associated with various steps in the workflow lifecycle such as creation, execution and monitoring and debugging. Through hands-on exercises in a hosted Jupyter notebook environment, we will describe an application pipeline as a Pegasus workflow using Pegasus Workflow API and execute the pipeline on distributed computing infrastructures. The attendees will leave the tutorial with knowledge on how to model their pipelines in a portable fashion using Pegasus workflow and run them on varied computing environments. The tutorial will also cover how to bundle application codes into a container and use them in workflows.

Click for instructions for tutorial T004 participants  
[<https://adass2023.lpl.arizona.edu/sites/adass2023.lpl.arizona.edu/files/documents/ADASS23-Instructions-for-Pegasus-Tutorial-v2.pdf>]

## **6:00PM to 10:00PM ADASS RECEPTION**

**Notes:** Event Description

The ADASS 2023 Opening Reception will be held at Gentle Ben's Brewing Company [<https://adass2023.lpl.arizona.edu/agenda#az-accordion--3>].