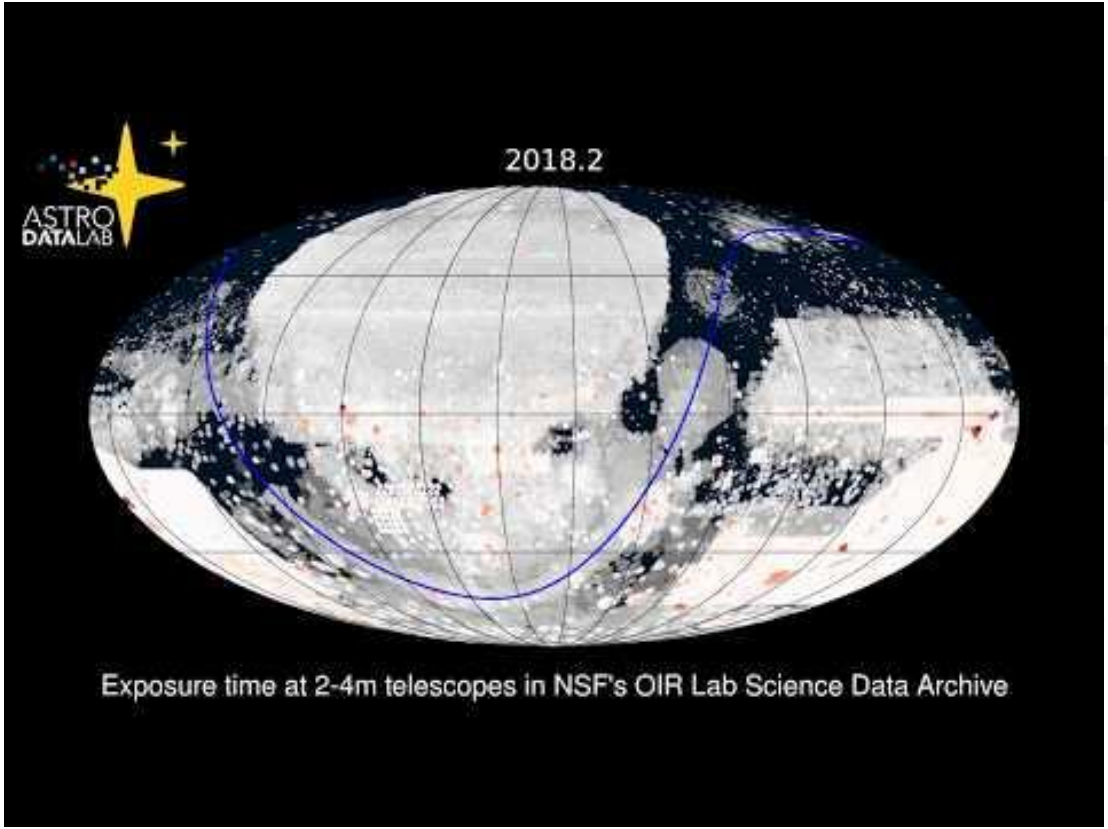
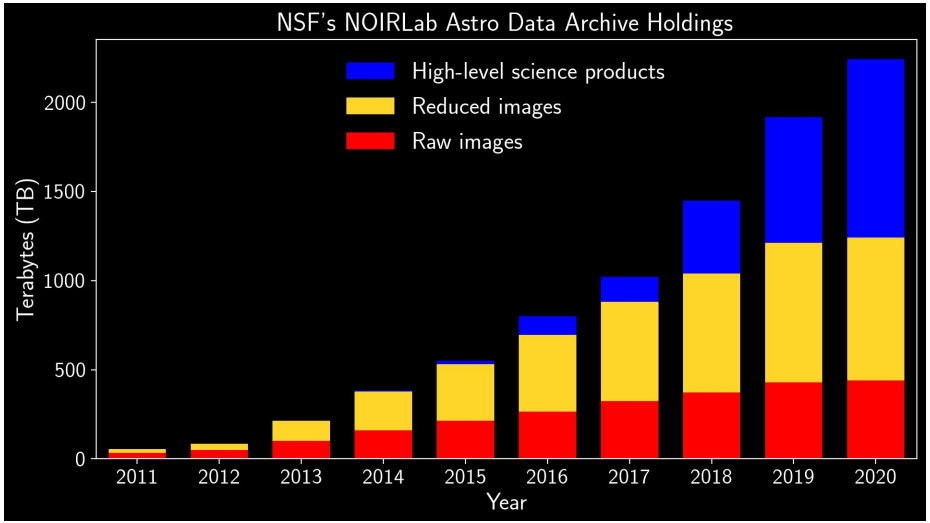
Astro Data Lab
Mission, Data, and Services

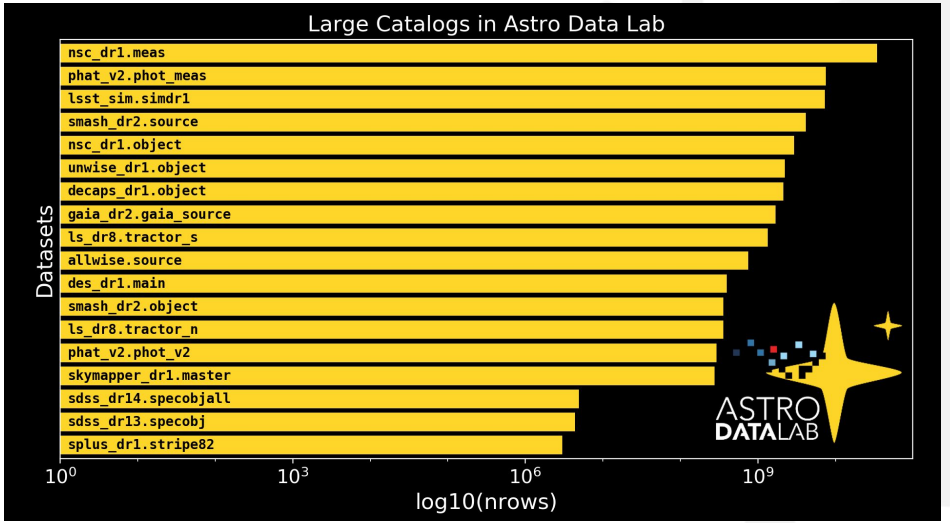
Robert Nikutta, Project Scientist
For the Astro Data Lab team

The Data Avalanche



Data growth at CSDC





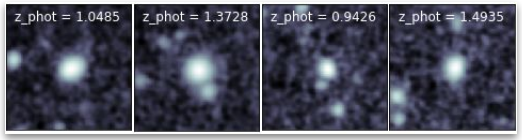
Data complexity, connectivity

Currently at Astro Data Lab ...

Catalogs
(2D tables)

	ls_id	ra	dec	dered_mag_r	dered_mag_g
0	8797229232750724	286.604936	43.783519	19.3421	20.7393
1	8797229232750718	286.602226	43.780599	22.8721	23.0592
2	8797229232750733	286.603586	43.786786	22.9804	23.4134
3	8797229232750742	286.612393	43.790177	18.8789	20.2648
4	8797229232750743	286.612561	43.791592	20.5371	22.0037
5	8797229232750735	286.607780	43.788338	19.2442	19.7413

Images
(2D arrays)

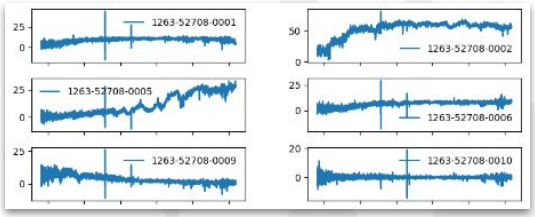


Heterogeneous
data collections
(file service)

```
1 print(sc.ls('gogreen_dr1://',format='long'))
```

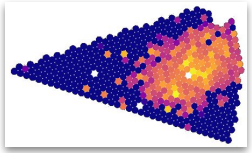
```
drw-rw----  gogreen_dr1       0  13 Aug 2020 17:54  CATS
drw-rw----  gogreen_dr1       0  13 Aug 2020 17:54  PHOTOMETRY
-rw-rw----  gogreen_dr1    5429  13 Aug 2020 17:54  README
drw-rw----  gogreen_dr1       0  13 Aug 2020 17:54  SPECTROSCOPY
drw-rw----  gogreen_dr1       0  13 Aug 2020 17:54  Scripts
```

1D spectra
(queryable)



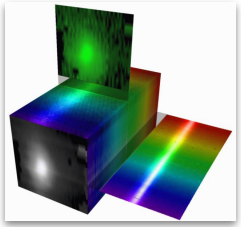
… in the future

DESI



2D spectra

IFU cubes &
complex data



GMOS-IFU,
MaNGA,
US-ELTs,
...

Astro Data Lab mission

To empower astronomers with an open-access platform for data-intensive science with the large astronomical survey data sets of today and tomorrow

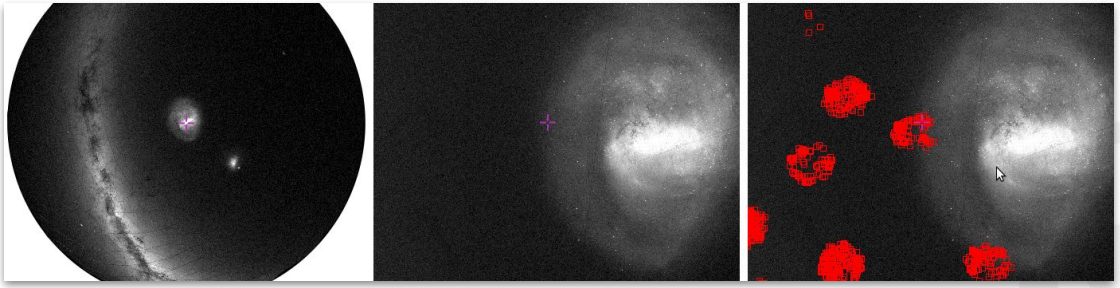
(one version)

High-level goals

- *Enable easy exploration of very large data holdings*
  → Catalogs, pixels, spectra, survey file collections…
- *Connect the various data products, joint analysis*
  → E.g. find interesting objects in catalogs, then find good images of them
- *Enable remote analysis*
  → Bring your code & algorithms & your data to the Big Data;
  execute code on our servers; analyze; visualize; publish
- *Enable easy user collaboration*
  → Sharing of query results, data sets, notebooks, group databases & storage
- *Help to train the community on big data analysis*
  → Prepare for the era of Rubin/LSST & Co.

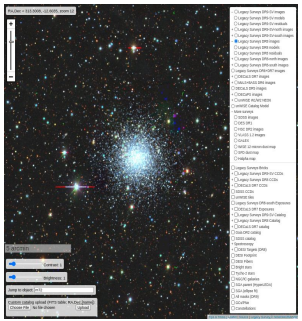
(Visual) data exploration

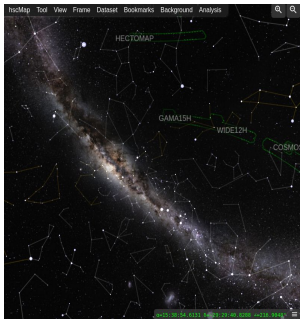
Web survey viewer
(based on Aladin Lite)

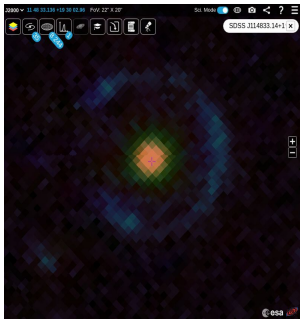


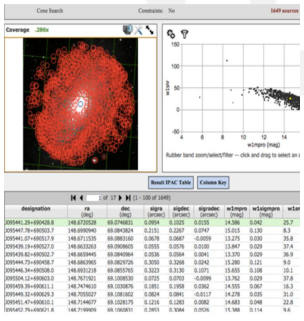
Survey footprint navigation ***Pan & Zoom*** *(here on the LMC)* ***Catalog overlays*** *(here SMASH fields)*

Also looking into other viewers, e.g.
- *Legacy Sky Viewer*
- *hscMap*
- *ESA Sky*
- *Firefly*
- *WWT*
- *...*





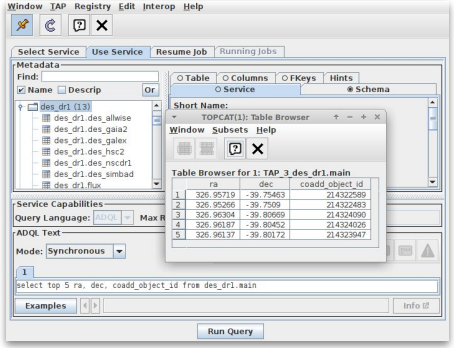




Getting to catalog data

- SQL-like queries via TAP→ PostgreSQL and ADQL
- Both sync and async queries→ Submit & wait / Submit & check later
- Can query both DL catalog holdings and user's MyDB
- Clients:

TAP-aware (e.g. TOPCAT)



queryClient.py (notebooks, scripts)

```
query = 'select ra, dec, coadd_object_id from des_dr1.main limit 5'
print(qc.query(query))
```

```
ra,dec,coadd_object_id
326.957189,-39.754627,214322589
326.952661,-39.750899,214322483
326.963039,-39.806693,214324090
326.96187,-39.804521,214324026
326.961371,-39.801715,214323947
```

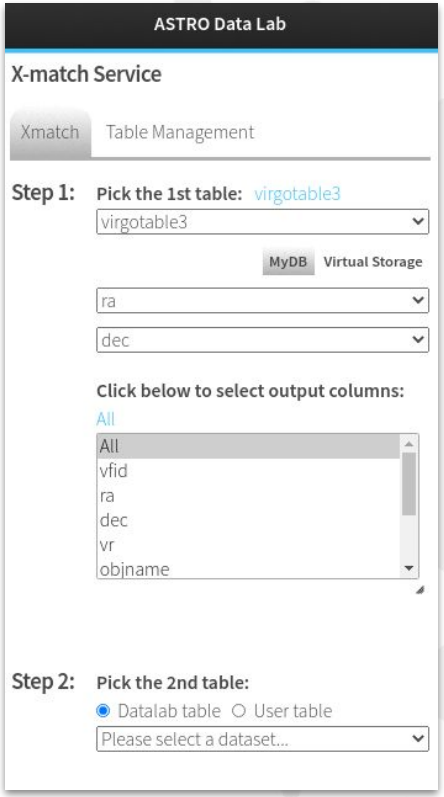
datalab CLI (on local computer)

```
$datalab query sql="select ra, dec, coadd_object_id from des_dr1.main limit 5"
ra,dec,coadd_object_id
326.957189,-39.754627,214322589
326.952661,-39.750899,214322483
326.963039,-39.806693,214324090
326.96187,-39.804521,214324026
326.961371,-39.801715,214323947
```

Query from DL website

Cross-matching catalogs

- Python API *(e.g. in Jupyter notebooks)*
- Positional cross-matching web tool *(uses same API)*
- On the backend: (Quad Tree Cube, Q3C); *Very fast!*
  (Koposov & Bartunov 2006)
- Tens of millions of rows in user table *are no problem*
- Also: hundreds of *pre-computed* cross-match tables



Connect catalogs with pixels

See Knut’s talk in a few minutes...

File services: public VOSpace

Public (read-only) file services to serve heterogeneous survey file collections, e.g.

- Arbitrary directory structure
- Weight masks, images, catalog files
- Documentation files
- “Aux” files… anything goes

Access through *`storeClient.py`* and *`datalab`* CLI

```
print(sc.services())
```

name	svc	description
chandra	vos	ChaMPlane: Measuring the Faint X-ray Bin ...
cosmic_dawn	vos	Cosmic DAWN survey
deeprange	vos	Deeprange Survey
deep_ecliptic	vos	Deep Ecliptic Survey
dls	vos	Deep Lens Survey
flamex	vos	FLAMINGOS Extragalactic Survey
fls	vos	First Look Survey
fsvs	vos	Faint Sky Variability Survey
ir_bootes	vos	Infrared Bootes Imaging Survey
lgs	vos	Local Group Survey
gogreen_dr1	vos	GOGREEN DR1 Survey
lmc	vos	SuperMACHO Survey
ls_dr1	vos	DECam Legacy Survey DR1
ls_dr2	vos	DECam Legacy Survey DR2
ls_dr3	vos	DECam Legacy Survey DR3
ls_dr4	vos	DECam Legacy Survey DR4
ls_dr5	vos	DECam Legacy Survey DR5
ls_dr6	vos	DECam Legacy Survey DR6
ls_dr7	vos	DECam Legacy Survey DR7
ls_dr8	vos	DECam Legacy Survey DR8
m31_newfirm	vos	M31 NEWFIRM Survey
ndwfs	vos	NOAO Deep-Wide Survey
nfp	vos	NOAO Fundamental Plane Survey
nmbs	vos	NEWFIRM Medium Band Survey
nmbs_2	vos	NEWFIRM Medium Band Survey II
nsc	vos	NOAO Source Catalog
sdss_dr8	vos	SDSS DR8
sdss_dr9	vos	SDSS DR9
sdss_dr10	vos	SDSS DR10
sdss_dr11	vos	SDSS DR11
sdss_dr12	vos	SDSS DR12
sdss_dr13	vos	SDSS DR13
sdss_dr14	vos	SDSS DR14
sdss_dr15	vos	SDSS DR15
sdss_dr16	vos	SDSS DR16
singg	vos	Survey for Ionization in Neutral-Gas Gal ...
smash_dr1	vos	SMASH DR1
smash_dr2	vos	SMASH DR2
sze	vos	SZE+Optical Studies of the Cosmic Accele ...
w_project	vos	The w Project
zbootes	vos	z-band Photometry of the NOAO Deep-Wide ...

Remote storage / user data co-location

User file storage: VOSpace

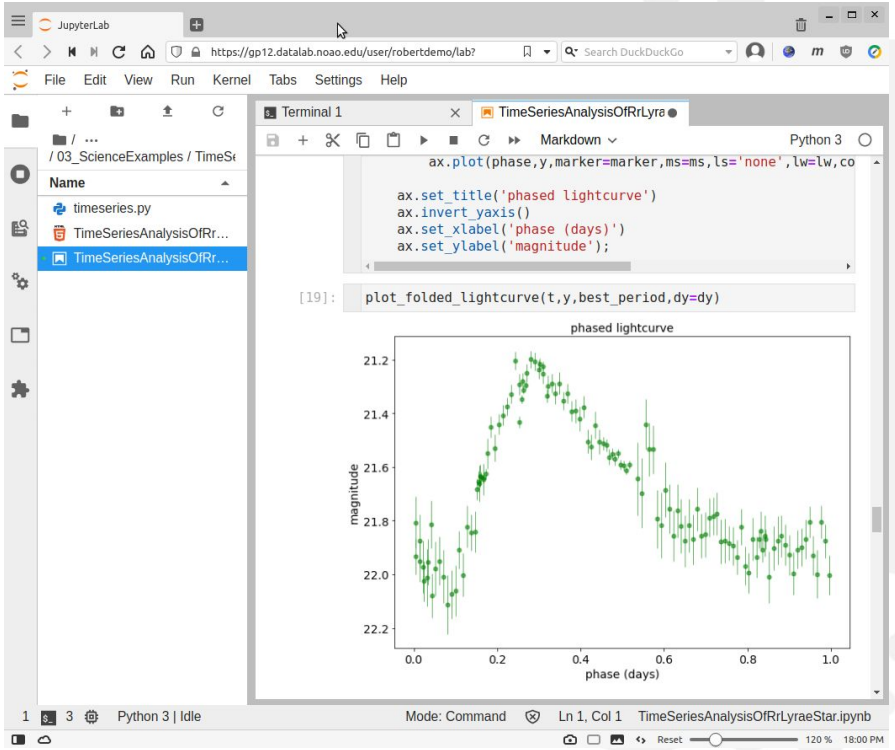
- 1 TB / user, r/w access, access from Jupyter notebooks and CLI
- `public/` subdirectory to share files with other users

User database: MyDB

- 250 GB / user, r/w, access from Jupyter and CLI
- also used for very fast positional cross-matching (web & programmatic)

Bring your analysis to the big data

- Remote computing, co-location with data
- No installation required (just a browser)
- Jupyter notebook server
- DL-curated NB suite + user-contributed NBs
- Full astro + data science S/W stack installed
- Planned: users *own* their containers, e.g. can install S/W
- Interfaces to data, to services, and to user storage (DB and VOSpace)



Also a service: User support

Sign up for a free Data Lab account:
https://datalab.noirlab.edu/

Get help from the DL team (**we solve every help request**)
Email: datalab@noirlab.edu
Helpdesk: https://datalab.noirlab.edu/help
User Manual: https://datalab.noirlab.edu/docs/manual
Ping us on Twitter: @DataLabAstro

Install Data Lab clients and CLI in your local computer:

```
pip install noaodatalab
```

The three things to take away

- A *Science Platform* combines big data, co-located / remote compute, data discovery, easy data access, analysis, visualization, and collaborative working.

- As part of the larger NOIRLab data ecosystem, the *Astro Data Lab* does all these, while hosting one of the largest collections of photometric data, image datasets, and spectroscopic capabilities. Importantly, users can *access to all data products from raw to HLSP*.

- We are a *community* Science Platform: users-first, open-source where possible, open-access (most data sets), open-standards (supporting IVOA protocols and interoperability).

Thank you

User base growth at Astro Data Lab