The DSA-2000 Radio Camera and the IVOA

www.deepsynoptic.org

Gregory Dubois-Felsmann, Caltech/IPAC
on behalf of the DSA-2000 team

DSA-2000

A world-leading radio survey telescope and multi-messenger discovery engine

- ~2000 x 5m dishes (19 x 15 km)
- Spring Valley, Nevada
- Frequency: 0.7 - 2 GHz band
- Spatial resolution: 3.3 arcseconds
- **Highly optimized for surveys**

- First light: 2027, key surveys: 2028 – 2033

Schmidt Sciences

DSA-2000

Schmidt Sciences

Unparalleled Survey Speed

31,000 deg^2 to 500 nJy (AB: 24.7)

>1 billion radio sources (IQUV)

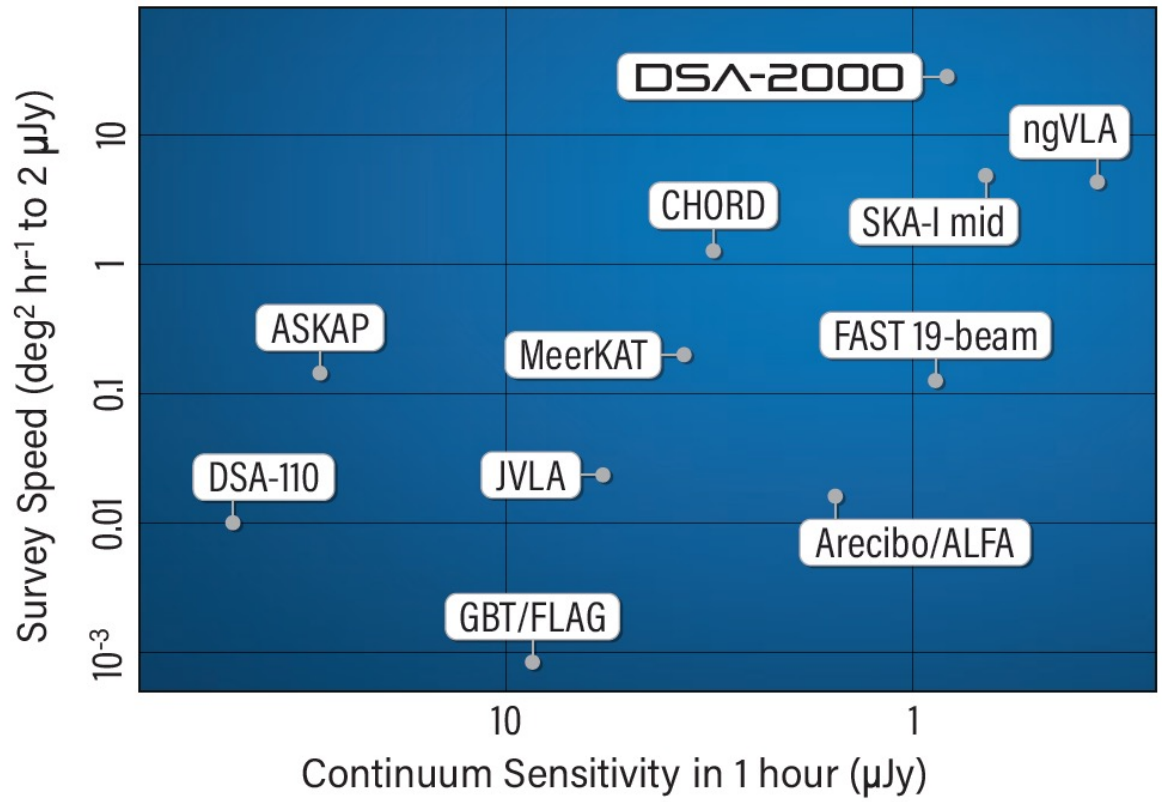
~few million galaxies in HI

~10^5 FRBs and pulsars

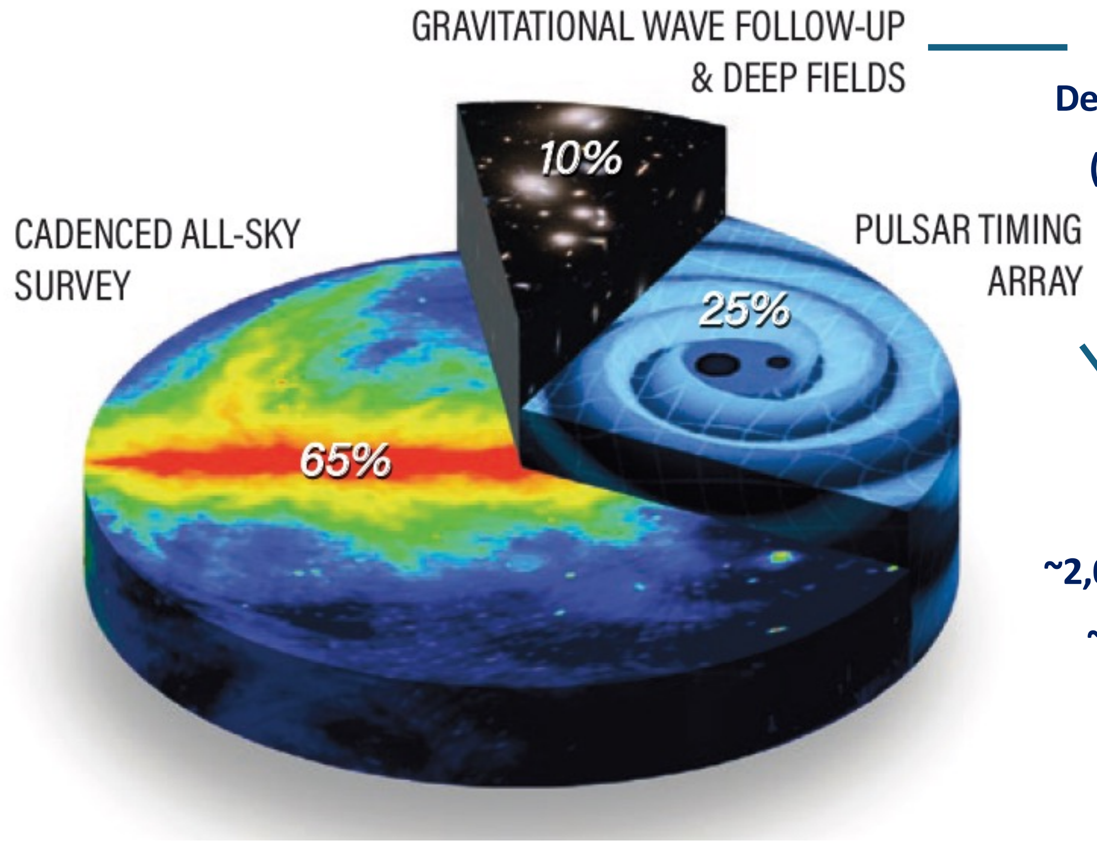
~10^6 “slow” transients

Enabled by two key technologies:

- A “radio camera” digital back-end
- A cryo-free antenna/receiver



5-Year Survey



30,000 deg^2 to 500 nJy

deep reference epoch

12 additional epochs

logarithmic cadence

GW (Triggered)

Deep: 30 deg^2 to 50 nJy

(or confusion limit)

~2,000 deg^2 to 200 nJy

~weekly cadence

Continuous and Commensal

Every ~10.3 minutes

Continuum

2 μJy rms per epoch (10 spectral windows)

Spectral Line

Galactic HI	960 channels	*0.2 km/s*
HI (<100 Mpc)	4192 channels	*1.7 km/s*
HI (z<1)	5600 channels	*27.5 – 55 km/s*

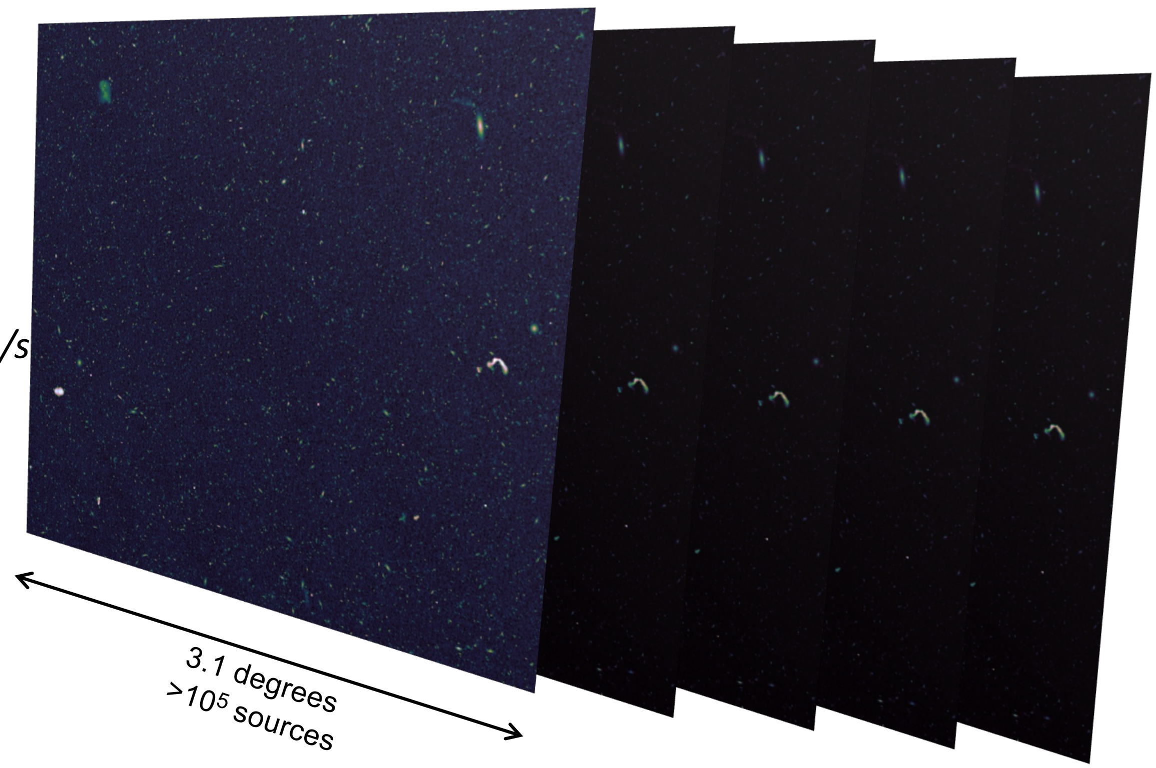
Polarization

Stokes IQUV images (500 x 2.6 MHz)

Fast Time Domain

FRB search

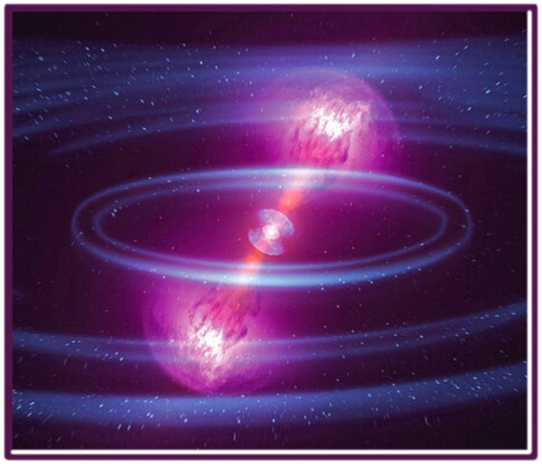
pulsar search

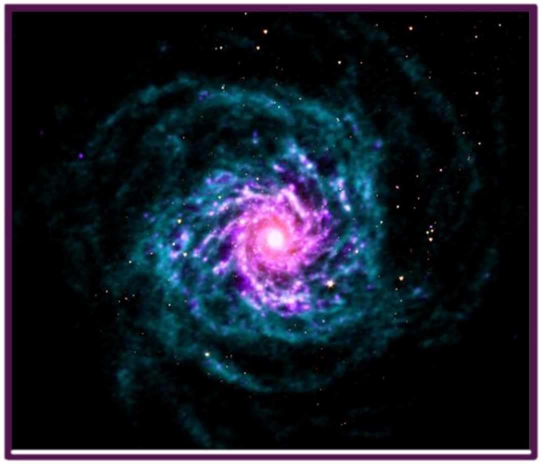


Fully public data with no proprietary period – Archived and served by IPAC

DSA-2000 Key Science

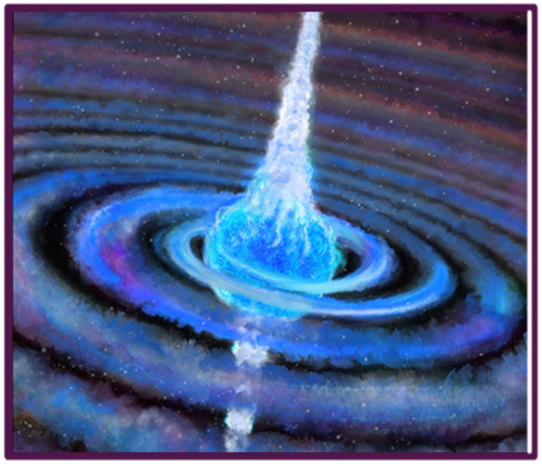
Multi-Messenger Astronomy

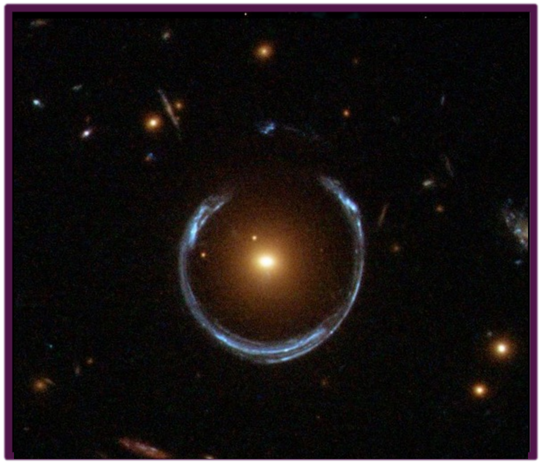




Our Cosmic History

The Dynamic Radio Sky





The Dark Sector and Strong Gravity

DSA-2000

Schmidt Sciences

Multi-Messenger Astronomy

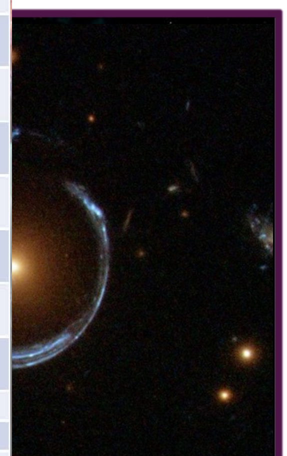
KSG1: Characterizing the nanohertz gravitational-wave universe via pulsar timing.
Q4, Q8, Q9

KSG2: Discovering the counterparts to compact binary coalescences detected in gravitational waves.
Q1, Q2, Q3, Q8, Q22

The Dynamic Radio Sky

KSG5: Determining the distribution of matter in the circum- and inter-galactic medium with FRBs.
Q2, Q10

KSG6: A new window on binary-driven mass loss and relativistic jets in stellar explosions.
Q2, Q3, Q22



Astro 2020 Science Panel Questions
Q1. What are the mass and spin distributions of neutron stars and stellar mass black holes?
Q2. What powers the diversity of explosive phenomena across the electromagnetic spectrum?
Q3. What do some compact objects eject material at nearly-light-speed jets, and what is that material made of?
Q4. What seeds supermassive black holes and how to they grow?
Q5. What set the hot Big Bang in motion?
Q6. What are the properties of dark matter and the dark sector?
Q7. What physics drives the cosmic expansion and the large-scale evolution of the universe?
Q8. How will measurements of gravitational waves reshape our cosmological view?
Q9. How did the intergalactic medium and the first sources of radiation evolve from cosmic dawn through the epoch of reionization?
Q10. How do gas, metals, and dust flow into, through, and out of galaxies?
Q11. How do supermassive black holes form and how is their growth coupled to the evolution of their host galaxies?
Q12. How do the histories of galaxies and their dark matter halos shape their observable properties?
Q13. What is the range of planetary system architectures, and is the configuration of the solar system common?
Q14. What are the properties of individual planets, and which processes lead to planetary diversity?
Q15. How do habitable environments arise and evolve within the context of their planetary systems?
Q16. How can signs of habitable life be identified and interpreted in the context of their planetary environments?
Q17. How do star-forming structures arise from, and interact with, the diffuse ISM?
Q18. What regulates the structures and motions within molecular clouds?
Q19. How does gas flow from parsec scales down to protostars and disks?
Q20. Is planet formation fast or slow?
Q21. What are the most extreme stars and stellar populations?
Q22. How does multiplicity affect the way a star lives and dies?
Q23. What would stars look like if we view them like we do the Sun?
Q24. How do the Sun and other stars create space weather?

Our Cosmic History

KSG3: Neutral hydrogen census through half of the Universe's age.
Q10, Q12, Q17

KSG4: The cosmic-ray lifecycle: production, propagation and cooling amidst galactic and intergalactic magnetic fields
Q3, Q4, Q9, Q11

The Dark Sector and Strong Gravity

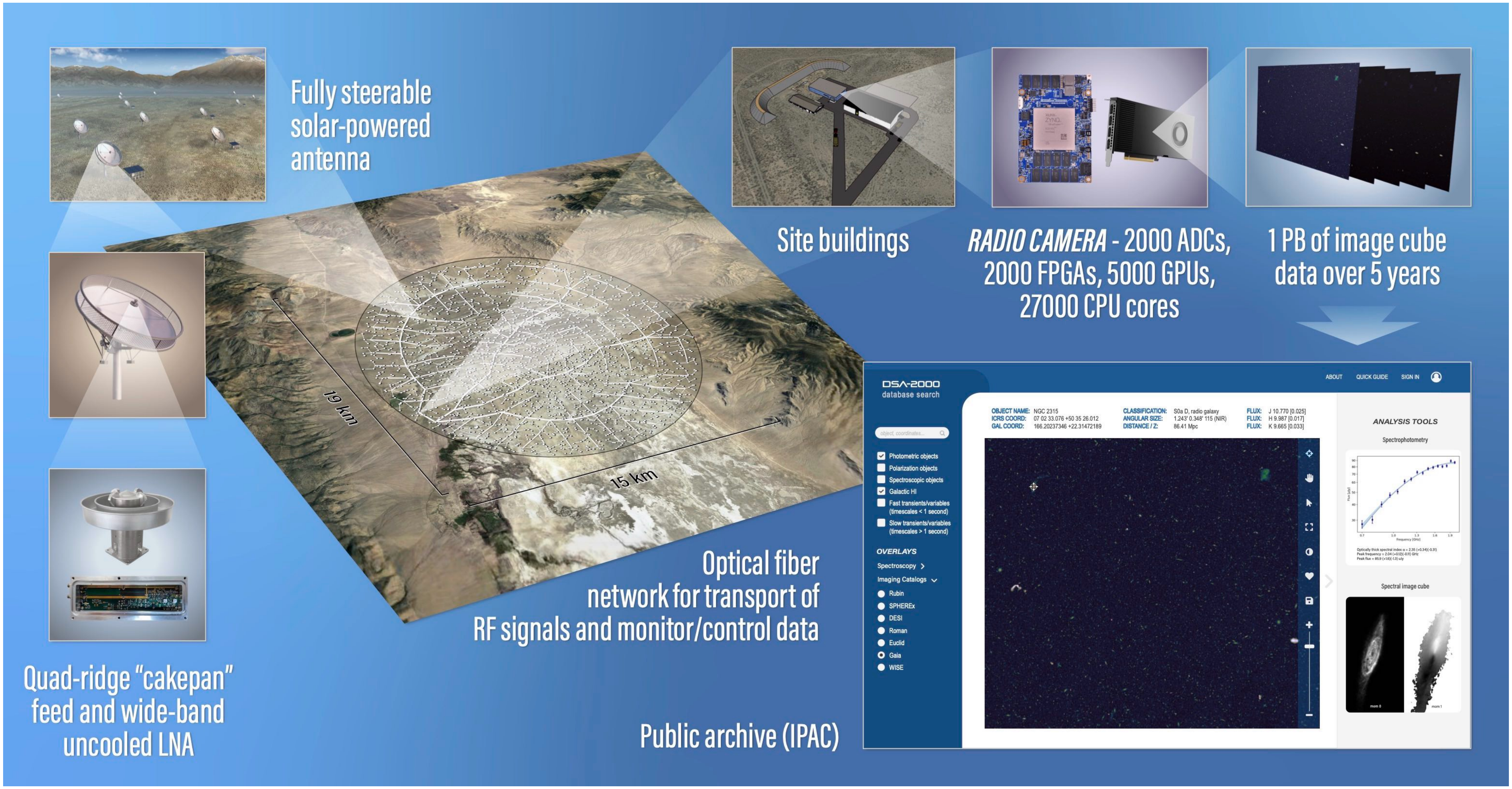
KSG7: Physical Characteristics / Fundamental Properties of Dark Matter and Dark Energy.
Q6, Q7, Q12

KSG8: A Galactic census of radio pulsars to test theories of gravity, bulk nuclear matter, and the endpoints of stellar evolution.
Q1, Q2

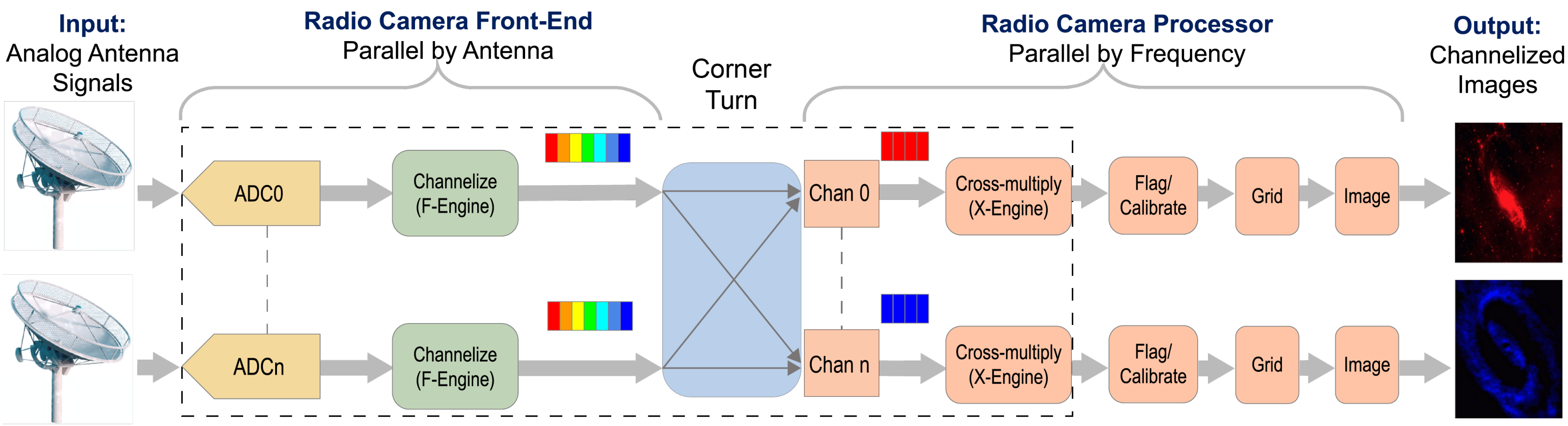
DSA-2000

DSA-2000 in a Nutshell

Schmidt Sciences



The First Radio Camera



50 Petabytes of image data over 5 years

~1 PB Public Archive

2000 x ADCs, FPGAs

iWAVE Zynq SoM

55x 100 GbE switches, cables

5160 GPUs (860 servers)

NVIDIA RTX 4000 (Ada architecture)

https://gitlab.com/dsa-2000/code/rci/rcl

DSA-2000

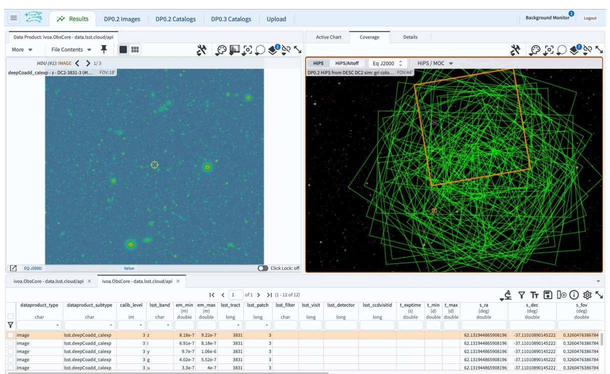
The DSA-2000 Public Archive

Schmidt Sciences

IPAC will be providing the spectrophotometric pipeline and catalog creation, as well as the public archive and data access environment for DSA-2000.

We will apply the same IVOA-centric approach that we do for our NASA archives and in the Rubin Observatory's "VO-first" Science Platform, and will support access to data through three mechanisms:

Portal (IPAC Firefly)



(VO) Data Services

ObsTAP

TAP

SSA

DataLink

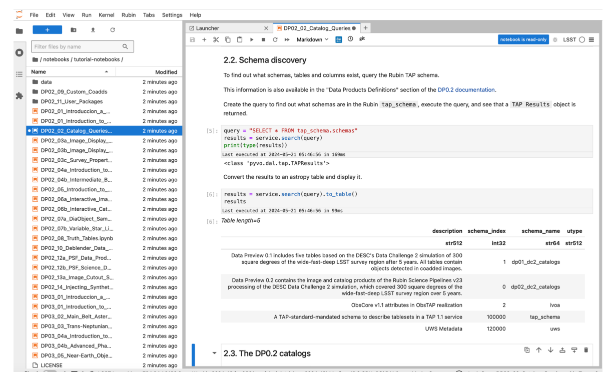
UWS

SIAv2/DAP

CAOM2

Data modeling

Notebooks (JupyterLab)



We will be evaluating cloud vs. on-premises tradeoffs during the upcoming final-design phase of the project.

Caltech ipac

Initial Public Data Products

The initial planned public data products are:

- Reduced images of ~30,000 deg^2 in 10 spectral bands and 16 epochs
- Object catalog with $O(10^9)$ rows with 10-band spectra

We will apply IRSA experience with SPHEREx (an all-sky 0.75-5um spectral survey, launching Feb. 2025), which has qualitatively similar all-sky image and catalog data products, to the design of the user interfaces and data services.

The imaging data will be made available in an "all-sky cube" visualization interface (we are very interested in working with CDS on enhancements to HiPS cubes) as well as in its native image files. A cutout service will be provided to package up all available imaging data for user-specified targets.

The catalog data will be made available for TAP queries, with DataLink annotations for the retrieval of spectral data, associated cutouts, and so on.

IVOA support – easy items

Much of what we need to do is covered by existing IVOA standards

Use cases:

- Show me the imagery for a region of interest; browse by wavelength; browse by time
- Find sources in the catalog near my points of interest
- Extract cutouts or perform photometry on images at my points of interest

- ObsTAP/SIAv2/DAP/DataLink cover access to native image data well
- ObsCore and CAOM2 are sufficient to cover the initial public imaging products
  - They are much like optical data in many respects
  - Existing time and spectral metadata/query capabilities support the planned data
- UWS (ideally the post-PTTT version!) supports planned calculational services (cutouts, forced-photometry-on-demand)
- SODA provides a reasonable starting point for the cutout service; will have to add features

- TAP with UPLOAD is adequate for small-to-medium scale, targeted catalog queries (but see below)
- DataLink to spectral data and cutouts
- SpectrumDM 1.2 for per-target spectra

DSA-2000

Easy-ish / bleeding-edge items

Schmidt Sciences

Some enhancements seem needed to provide the desired user experience

Use cases:

- Show me an all-sky view; extract spectra at click; browse by wavelength; browse by time
- Find sources in the catalog near my points of interest

- We think HiPS cubes need work in order to support the data access (including science-grade pixel spectra) that we need. *We will also need this (earlier!) for SPHEREx at IRSA.*
  - We have **4D all-sky data (x,y,λ,t)**; what's a good UX for this? A good data format? HiPS-4D?

- We will be providing a multi-mission, multi-band all-sky data access environment, including many other survey datasets in infrared, optical, and ultraviolet wavelengths
  - **Improved annotation of spectral bands, filter curves**, etc. across all these datasets will be needed to support on-the-fly, interoperable interpretation of the combined data
  - Simple example: assemble SED-like data displays for cross-matched sources through "generic" community tools just using VO interfaces
  - There are tools for this in PhotDM, VO-DML/MIVOT etc. – we might not need new standards – but we expect issues to arise when we work out all the details in implementation

Caltech ipac

Technical/deployment challenges

A real trade study is planned as part of our final-design process, but…

There are lots of reasons to suspect we'll be led to a cloud-based deployment

- It would be helpful to not have to replicate historical features of the community's on-premises deployment infrastructure when transitioning to cloud services

One major highlight: **Can we use commercial cloud vendors' database-as-a-service offerings?**

- This has been challenging in the community, in no small part due to the use of spherical geometry extensions to databases which are not well-supported in cloud offerings – or which have vendor lock-in issues (c.f. the CADC/Google/Rubin demonstration of TAP-over-BigQuery)

- A full-featured open-source TAP-over-cloud-Postgres solution, supporting RegTAP 1.2 / MOC queries, and with rich annotation with DataLink service descriptors and MIVOT data, verified to work with multiple vendors, would be an immense contribution to the community!

DSA-2000

IVOA support – challenges (Big Data)

Schmidt Sciences

But… much of the power of this dataset will be in **wide-area statistical analyses**, including cross-matching with Rubin, Roman, Euclid, WISE, SPHEREx, UVEX, etc. data

We need to **support bulk data processing** for machine learning, application of user-specified cross-matching algorithms, "find me more objects like this one" searches, etc.

We think this needs **co-location of many datasets**, supported by

- standards for in-situ ("code to data") joint, parallel analyses of 10^9-10^{10} row (and beyond) catalog datasets (looking at the 'hipscat' effort here – perhaps with something like Execution Broker?)
- standards (or perhaps standard practices) for registering/discovering co-located copies of datasets
- standards for bulk (catalog) data access that are more big-data-centric than currently
  - E.g., TAP-like queries with results in the cloud (in Parquet?) ready for further parallel analysis

We also need standards support for **bulk access to spectra**

- There's a massive increase already beginning (DESI!) in the number of spectra available for retrieval – SpectrumDM and SSA are not ready for this

Caltech ipac

IVOA support – beyond the initial data

The project will have many other possible derived data products, including ones that are more "radio-unique" in their characteristics.

Some of these datasets are so large that "code to data" is the only plausible solution on the project's time scale.

We think our basic Science Platform architecture is well-suited to this, but data access metadata and mechanisms will need work.

We hope the project will be able to be part of advancing standards in this area, and we'll be looking at some of the specific-to-radio challenges and their match-up with existing and in-progress IVOA work during the upcoming final-design period.

Onward

Many of us face the same challenges – let's work together

https://www.deepsynoptic.org/

Questions?