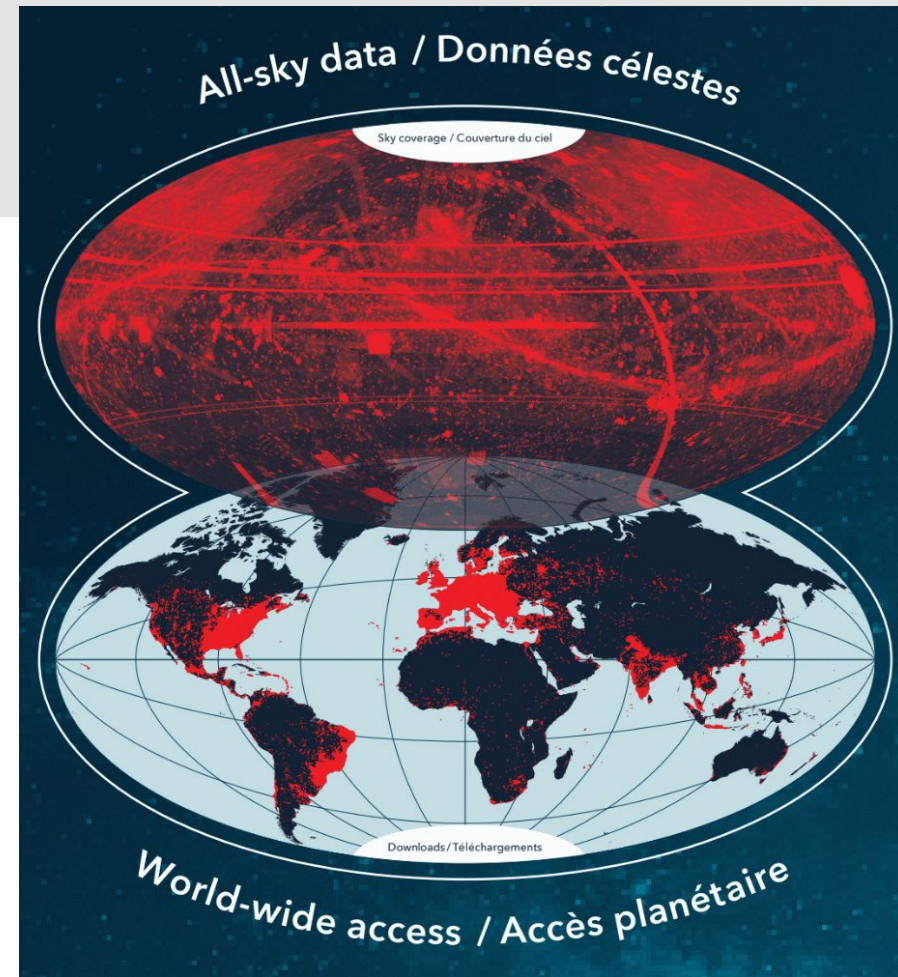


# The Canadian Astronomy Data Center: an Independent Data Access Centre for the LSST



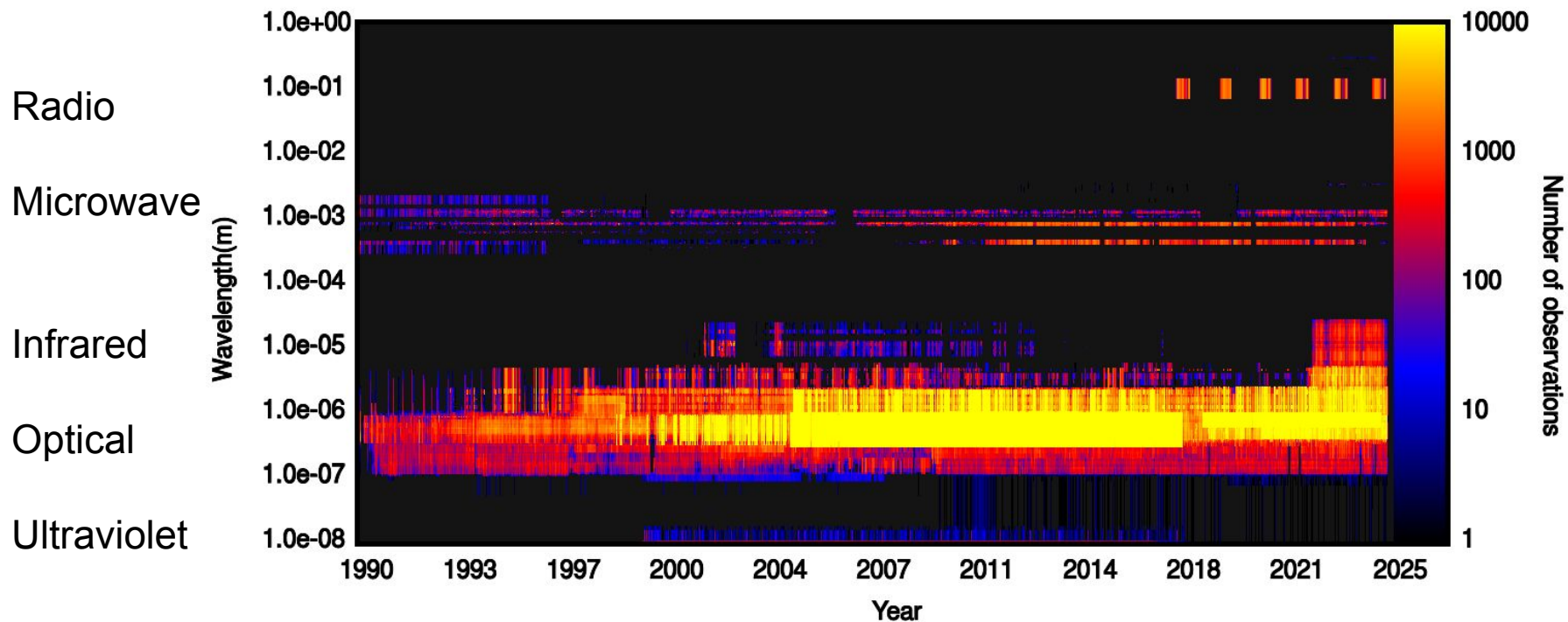
# CADC: Canadian Astronomy Data Centre

- Data distribution site for Canada's astronomical data
- We mirror some datasets (HST, JWST, Gemini, ...)
- We are the sole repository for other datasets (CFHT, JWST, OMM, NEOSSat, ...)
- 216 telescopes and instruments
- Current holdings:
  - 2.3 PB, 300 million files
  - Ceph ObjectStore back end
  - CADC layer on top: Storage Inventory
- Annual downloads:
  - 100 million files
  - 4.9 Petabytes
- Users are 15% Canadian, 85% international
- Upcoming datasets
  - LSST
  - Euclid
  - SKA



# Canadian Astronomy Data Centre

An alternate view of the archive





# A brief history of the CADC

- Started in 1986 to host a copy of HST data. The idea was scientists would travel to Victoria to process data.
- Expanded its mandate to include other observatories
- Starting providing access to catalogs, both standard and user-provided
- Started processing data for users, providing science ready data products
- Evolved to 3 main activities:
  - Archives
  - Data expertise
  - Providing compute facilities



# Canadian Astronomy Data Centre:

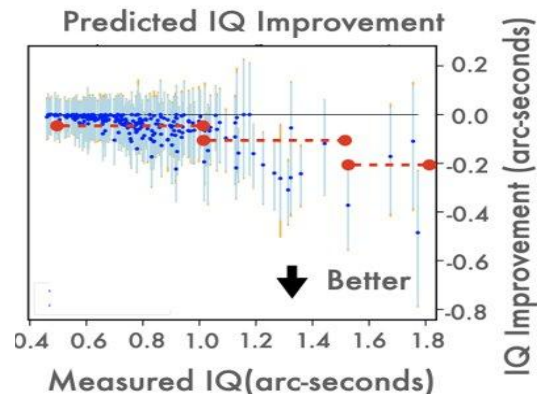
## Processing data and providing machine learning expertise

### The CADC processes raw data into science-ready data products

- Makes the data significantly easier for scientists to use
- Increases data re-use
- Two modes
  - Off-the-shelf / prêt-à-porter processing of entire archives
  - Bespoke / tailored processing for particular projects
- Started with HST data in the early 1990s
- Expanded to CFHT in the early 2000s

### The CADC provides machine learning expertise to astronomers

- Detection of post-merger galaxies in (CFHT)
- CFHT image quality forecasting
- Extragalactic star-formation studies (SDSS)
- Detection of gravitational lenses (HST)
- Image assessment and recommendation (CFHT)
- Photometric redshifts (HSC)
- Adaptive optics predictive controls
- Extragalactic metallicity studies
- Moving object detection (CFHT, HSC)
- Spectral analysis of nearby galaxies (MaNGA IFU)
- Delirium problem in COVID-19 patients



# CANFAR:

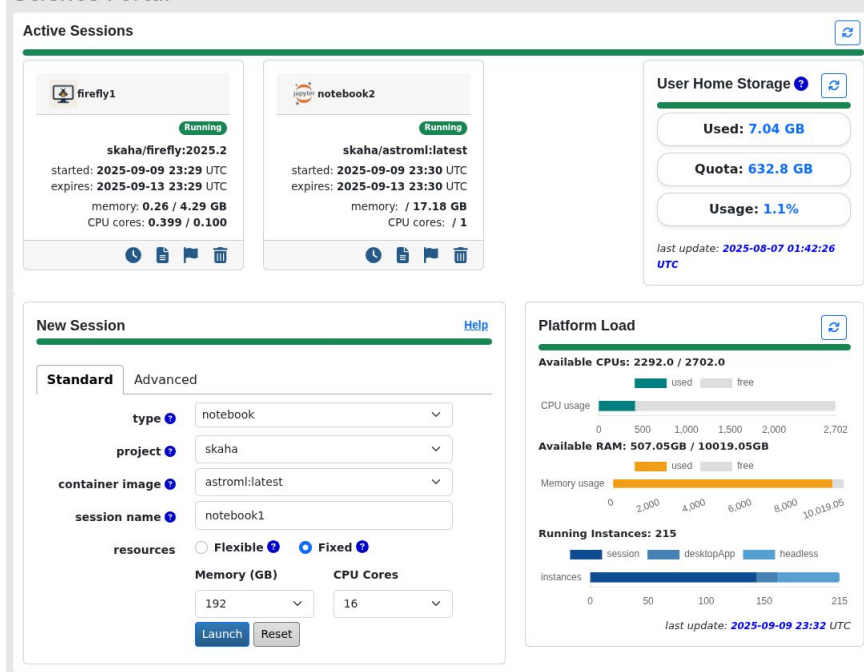
## Canadian Advanced Network For Astronomical Research



### The CANFAR Science Platform:

- Kubernetes-based
- Notebooks
- Browser-based VNC desktops
- Data visualization
- Scalable resources
  - up to 192 GB RAM / 16 cores per instance
- Multiple software container options (key!)
  - ALMA / JWST / LSST / ...
  - user generated
  - ability to share containers
- Batch processing (now uses Kueue)
- Very successful: most Canadian astronomers are users

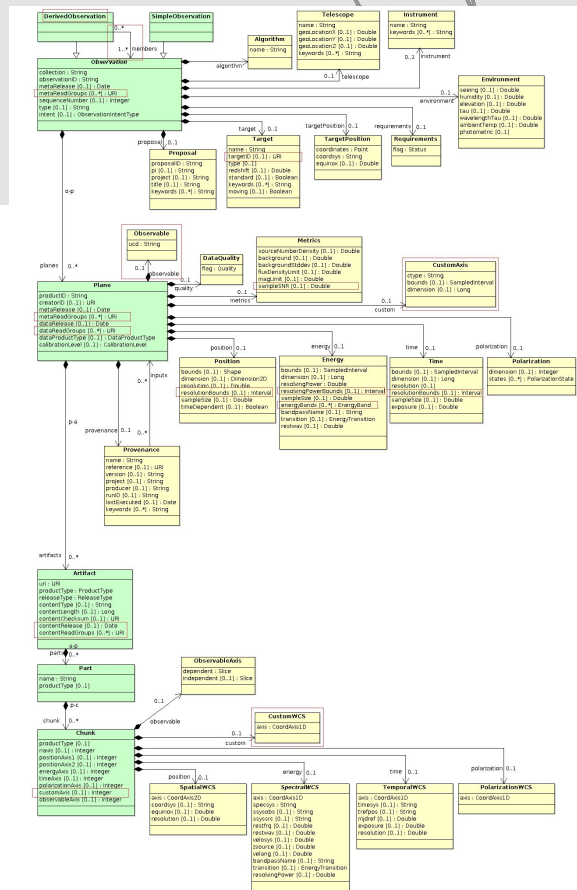
### Science Portal



# Astronomical metadata

## Metadata indexing

- 219 telescopes and instruments / 1 metadata system
  - CAOM2: Common Astronomical Observation Model
  - Very comprehensive
- Reduces effort needed to manage existing archives
- Low overhead to ingest new archives
- Not all data findable at via CADC search resides at the CADC
- Allows "one stop shopping" for astronomical data
- Initial development lasted ~1 year
  - multiple stakeholders
- Significant changes will be required as we pivot towards radio astronomy
- Each new telescope/instrument requires a period of data engineering (DE)
- Initial model: Astronomers (domain experts) drive DE
- Current model: Developers (model experts) drive DE



## Unified Modeling Language diagram of CAOM2



# CANFAR:

## User storage

### Cavern / arc:

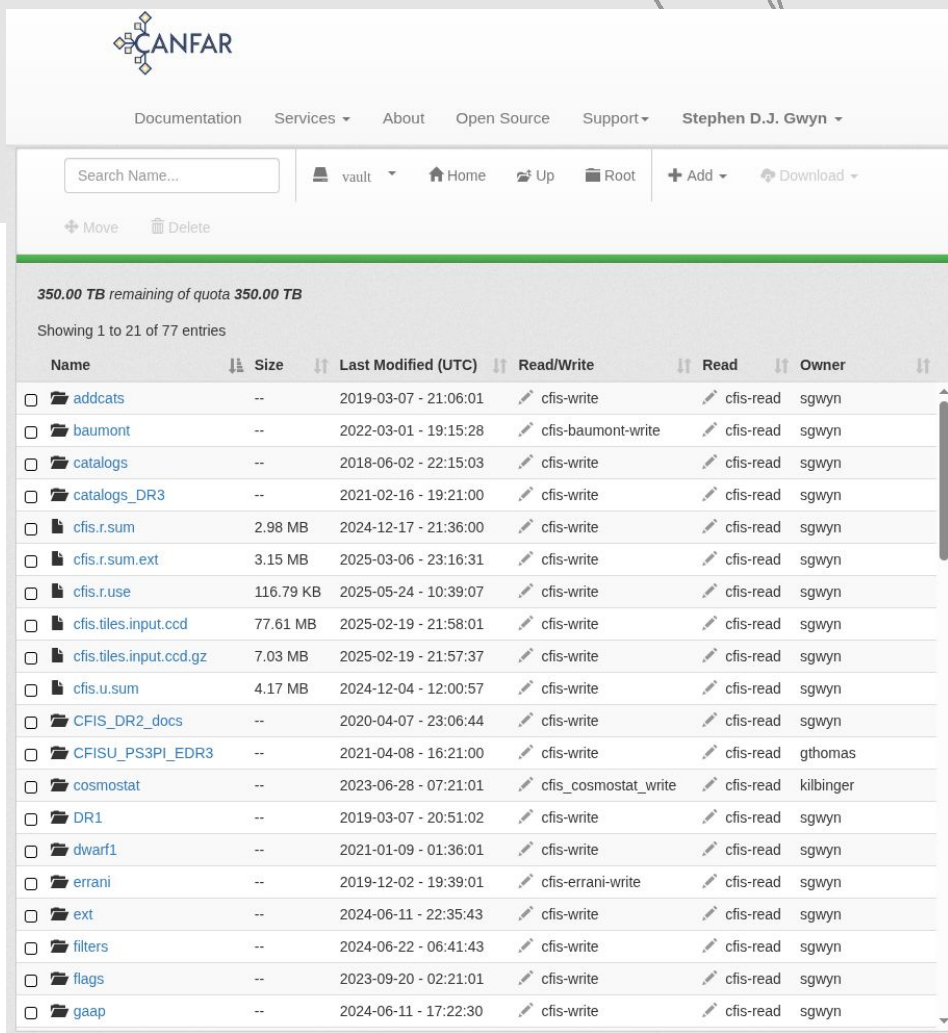
- Suitable for processing
- Each user gets a small home directory
- Project directory on request, quota can be expanded
- Persistent, stable, but not suitable for archival
- Looks like a POSIX file system
- Backed by CephFS

### Vault / vos:

- Suitable for data publication and long term curation
- Data mirrored at two sites
- Backed by Ceph Object Store
- Available by request
- Self-serve DOIs also available

### Both:

- Accessible via
  - VOSpace API
  - Python Client
  - Web UI
- Fine grained access control



The screenshot displays the CANFAR web interface. At the top is the CANFAR logo and a navigation bar with links: Documentation, Services, About, Open Source, Support, and a user profile for Stephen D.J. Gwyn. Below the navigation bar is a search bar labeled 'Search Name...' and a set of icons for file management: vault, Home, Up, Root, Add, and Download. A secondary row of icons includes Move and Delete. The main content area shows a status bar indicating '350.00 TB remaining of quota 350.00 TB' and 'Showing 1 to 21 of 77 entries'. Below this is a table listing files and directories. Each row includes a checkbox, an icon, the file name, size, last modified date, and read/write permissions. The table is scrollable, showing a list of files like 'addcats', 'baumont', 'catalogs', and 'cfis.r.sum'.

Name	Size	Last Modified (UTC)	Read/Write	Read	Owner
<input type="checkbox"/> addcats	--	2019-03-07 - 21:06:01	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> baumont	--	2022-03-01 - 19:15:28	cfis-baumont-write	cfis-read	sgwyn
<input type="checkbox"/> catalogs	--	2018-06-02 - 22:15:03	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> catalogs_DR3	--	2021-02-16 - 19:21:00	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> cfis.r.sum	2.98 MB	2024-12-17 - 21:36:00	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> cfis.r.sum.ext	3.15 MB	2025-03-06 - 23:16:31	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> cfis.r.use	116.79 KB	2025-05-24 - 10:39:07	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> cfis.tiles.input.ccd	77.61 MB	2025-02-19 - 21:58:01	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> cfis.tiles.input.ccd.gz	7.03 MB	2025-02-19 - 21:57:37	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> cfis.u.sum	4.17 MB	2024-12-04 - 12:00:57	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> CFIS_DR2_docs	--	2020-04-07 - 23:06:44	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> CFISU_PS3PI_EDR3	--	2021-04-08 - 16:21:00	cfis-write	cfis-read	gthomas
<input type="checkbox"/> cosmostat	--	2023-06-28 - 07:21:01	cfis_cosmostat_write	cfis-read	kilbinger
<input type="checkbox"/> DR1	--	2019-03-07 - 20:51:02	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> dwarf1	--	2021-01-09 - 01:36:01	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> errani	--	2019-12-02 - 19:39:01	cfis-errani-write	cfis-read	sgwyn
<input type="checkbox"/> ext	--	2024-06-11 - 22:35:43	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> filters	--	2024-06-22 - 06:41:43	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> flags	--	2023-09-20 - 02:21:01	cfis-write	cfis-read	sgwyn
<input type="checkbox"/> gaap	--	2024-06-11 - 17:22:30	cfis-write	cfis-read	sgwyn



# Enhancements to infrastructure

## Currently:

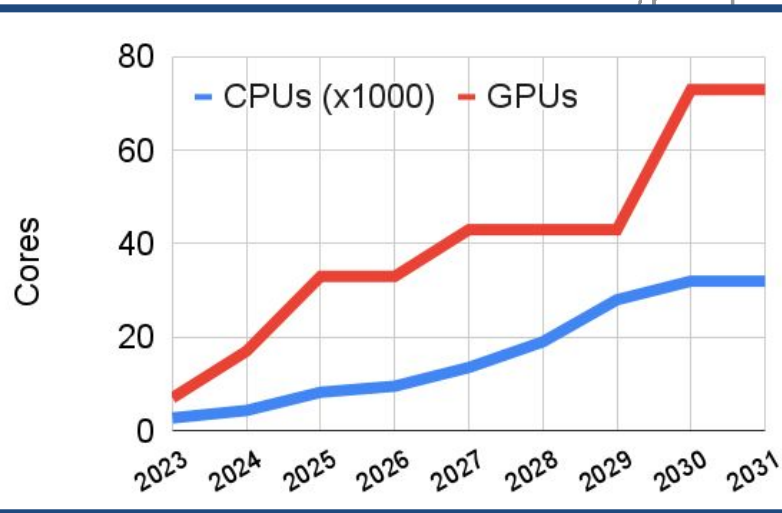
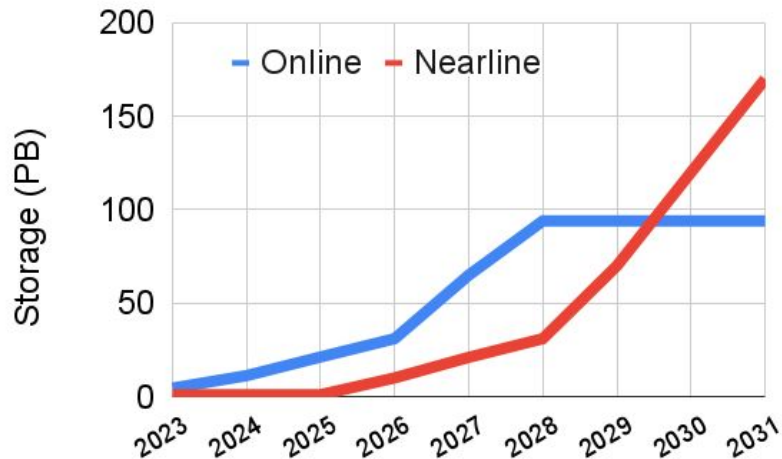
- 10PB
- 3000 CPUs
- 10 V100 + 1 A100 GPUs
- 20TB database

## Arrived for 2025

- 8PB
- 3900 CPUs
- 16 H100 GPUs
- 200TB federated database

## Steady state 2030

- 32000 CPUs
- 72 GPUs
- 90 PB online storage
- 170 PB nearline storage
  - (growing 50PB/year)



# Digital Research Alliance of Canada

- Provide research IT infrastructure to Canadian academics
- Long-standing partnership with CADC
- Initially through a resource allocation through university-based astronomers
- Expansion driven through funds transferred from CADC

## Canada's Advanced Research Computing Platform



Digital Research  
Alliance of Canada



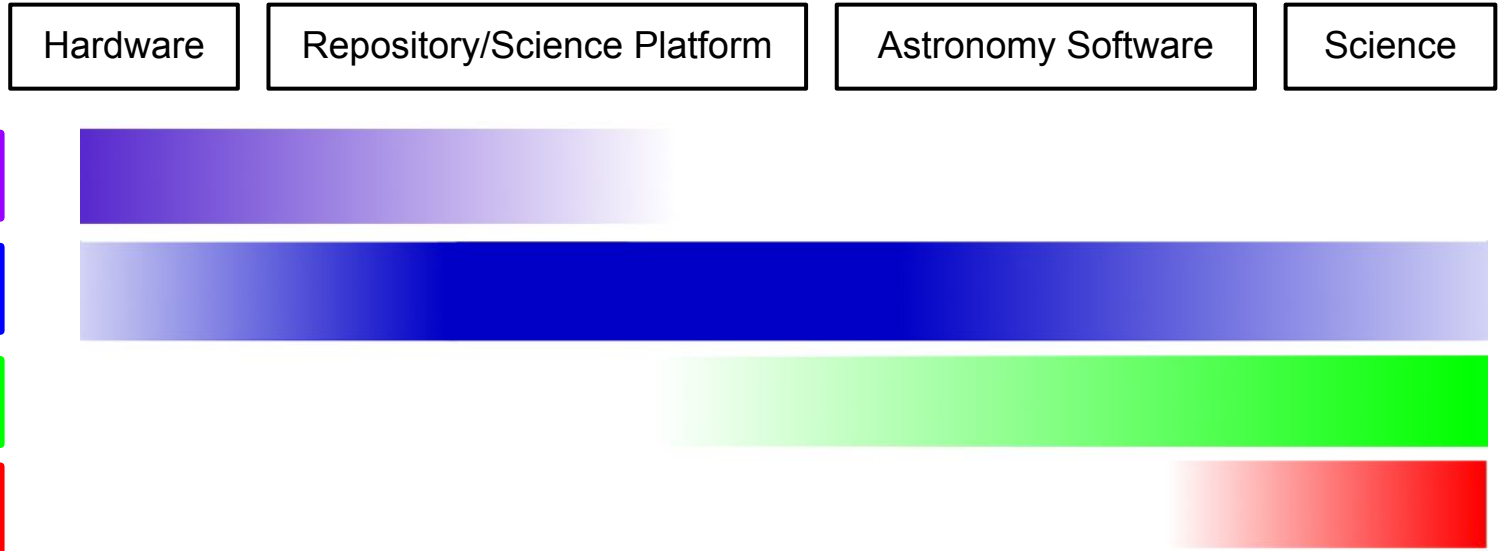
National Host Sites



Local Support Sites



# A spectrum of competencies



# The Canadian IDAC

- The CADC is building an IDAC
- DP1 and DP2 will be available through CANFAR
- For the DRs we will host
  - coadded images: 2.7 PB
  - ObjectLite catalog
  - providing 3000 cores (may change some to GPUs)
  - providing 2PB of user storage
- One set of resources for Rubin data-rights holders
- One set of resources for the wider astronomical community when the data becomes public





# The Canadian IDAC

## Interfaces:

- CADC tools
  - Data storage / retrieval
    - StorageInventory
    - VOSpace
  - Database interface
    - CADC TAP service
  - Processing
    - CANFAR Science Platform
- LSST tools
  - DataButler
  - LSSTpipe container
- Data visualization
  - Rubin Science Portal



# CADC's place in the LSST computing landscape

- Based on our experience with CANFAR, usage is evenly distributed in log space
  - most users require very few resources: Rubin Science Platform
  - some users require some resources
  - a few users require a lot of resources: IN2P3, NERSC, EPCC
- For the Rubin data rights holders
  - we best serve users who require a more power than RSP provides
  - not (currently) optimized for large single-user allocations
  - “First come, first serve” provides maximum flexibility
- For the public releases, the “first come, first serve” method is the only option
- Besides LSST, the CADC is hosting:
  - SKA
  - Euclid
  - CFHT (UNIONS = LSST of the North)
  - All available through a common set of tools and interfaces



# Current status

- DP1 is ingested into the CADC
  - Access via CADC tools
  - Access via Rubin tools / DataButler
- Access control is by CADC username
  - Need to be manually linked to Rubin rights
- Working on a general solution:
  - Same login as RSP will get you access to the CADC
- Rubin users currently have full access to all CANFAR resources
  - Separate allocations not yet in place
- Test users welcome!

Search ObsCore Search Results Error ADQL Help

Download complete query results: VOTable CSV TSV Bookmark URL

Download Showing 2084 rows (2084 before filtering). Change Columns View in sky

Mark	Preview	Collection	Obs. ID	Product ID	RA (J2000.0)	Dec. (J2000.0)	Instrument
<b>Filter:</b>							
					H:M:S	D:M:S	
N/A		LSST.DP1	lsst_cells_v1-4848-60	lsst.deep_coadd-1	03:32:01.19	-28:45:20.0	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-70	lsst.deep_coadd-4	03:32:00.86	-28:35:20.1	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-71	lsst.deep_coadd-3	03:31:15.31	-28:35:23.9	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-71	lsst.deep_coadd-5	03:31:15.31	-28:35:23.9	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-73	lsst.deep_coadd-5	03:29:44.21	-28:35:28.6	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-80	lsst.deep_coadd-3	03:32:00.53	-28:25:20.2	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-81	lsst.deep_coadd-3	03:31:15.06	-28:25:23.9	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-83	lsst.deep_coadd-1	03:29:44.10	-28:25:28.7	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-83	lsst.deep_coadd-3	03:29:44.10	-28:25:28.7	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-90	lsst.deep_coadd-1	03:32:00.20	-28:15:20.3	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-90	lsst.deep_coadd-3	03:32:00.20	-28:15:20.3	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-90	lsst.deep_coadd-4	03:32:00.20	-28:15:20.3	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-91	lsst.deep_coadd-2	03:31:14.80	-28:15:24.1	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-91	lsst.deep_coadd-5	03:31:14.80	-28:15:24.1	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-92	lsst.deep_coadd-2	03:30:29.40	-28:15:26.9	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-93	lsst.deep_coadd-3	03:29:43.99	-28:15:28.8	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4848-93	lsst.deep_coadd-4	03:29:43.99	-28:15:28.8	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-68	lsst.deep_coadd-4	03:32:39.93	-28:45:23.8	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-78	lsst.deep_coadd-0	03:32:40.19	-28:35:23.9	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-78	lsst.deep_coadd-3	03:32:40.19	-28:35:23.9	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-79	lsst.deep_coadd-2	03:31:54.64	-28:35:20.1	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-88	lsst.deep_coadd-4	03:32:40.44	-28:25:23.9	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-89	lsst.deep_coadd-0	03:31:54.97	-28:25:20.2	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-89	lsst.deep_coadd-5	03:31:54.97	-28:25:20.2	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-97	lsst.deep_coadd-0	03:33:26.10	-28:15:26.9	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-97	lsst.deep_coadd-2	03:33:26.10	-28:15:26.9	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-97	lsst.deep_coadd-4	03:33:26.10	-28:15:26.9	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-98	lsst.deep_coadd-0	03:32:40.70	-28:15:24.1	LSSTComC
N/A		LSST.DP1	lsst_cells_v1-4849-98	lsst.deep_coadd-3	03:32:40.70	-28:15:24.1	LSSTComC

Date modified: 2025-03-24

# LSST activities in Canada

- Canadian LSST project lead: Renée Hložek, University of Toronto
- In-kind contributions
  - Directable development
  - Skilled software development (science pipelines)
    - Dark energy science
    - Galaxies science
    - Transients
    - Solar system science
      - RAFTS (Research Announcements For The Solar system)
  - CanDIAPL = Canadian Data-Intensive Astrophysics PLatform
  - 17 PI slots
- CADC IDAC: 30 PI slots
- Total 47 PI slots + 4x postdocs+students/PI
- ICRADA is (finally) signed





# Summary

- The CADC is the astronomical data repository for Canada
- CANFAR provides a compute/user storage platform to support multiple projects through a common set of tools
- CADC/CANFAR is growing by a factor of 30 to support SKA+LSST
- The CADC supports LSST with
  - data hosting
  - user processing and storage
  - software development
- CANFAR provides an intermediate level of resources relative to RSP or IN2P3 / NERSC / EPCC

