

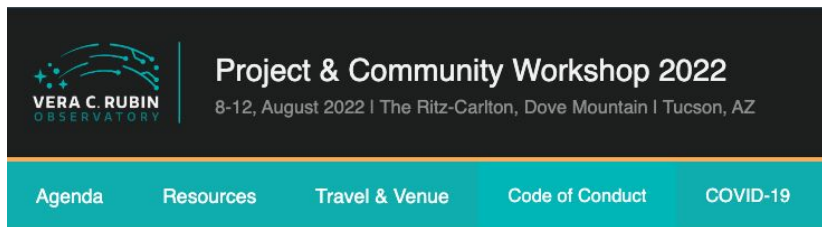


# Connecting the Community to IDAC and SPC Resources



U.S. DEPARTMENT OF  
**ENERGY**

# Friendly reminders - CoC & Covid




Home » Code of Conduct

## Code of Conduct

Harassment and unprofessional conduct (including the use of offensive language) of any kind is not permitted at any time and should be reported.

Rubin Observatory adheres to the principles of kindness, trust, respect, diversity, and inclusiveness in order to provide a learning environment that produces rigor and excellence.



Handshakes OK  
Fold Here

Elbow/Fist Bump OK  
Fold Here

I Need My Space  
Fold Here

Thank you for masking indoors!

Check name-tags for these contact comfort level stickers.

Use the confidential email [rubin2022-covid@lists.lsst.org](mailto:rubin2022-covid@lists.lsst.org) to request a test, report your test results, or ask questions.

## Reporting bullying, harassment, or aggression.

The Rubin 2022 Organizing Committee has appointed designated contacts:

- Ranpal Gill ([rgill@lsst.org](mailto:rgill@lsst.org))
- Andrew Connolly ([ajc@astro.washington.edu](mailto:ajc@astro.washington.edu))
- Melissa Graham ([mlg3k@uw.edu](mailto:mlg3k@uw.edu))

Contact via email, Slack, or the Community Forum.

# Friendly reminders - virtual participation



Virtual participants should be muted when they're not speaking.



In-person participants should speak into the room microphone(s), or the chair should repeat all questions into the microphone, so that the virtual participants can hear what is said.

In the **Rubin2022\_PCW Slack Space**, all participants can use the session's channel for Q&A and discussion.

The channel name is:  
#day4-thu-slot4a-in-kind-community

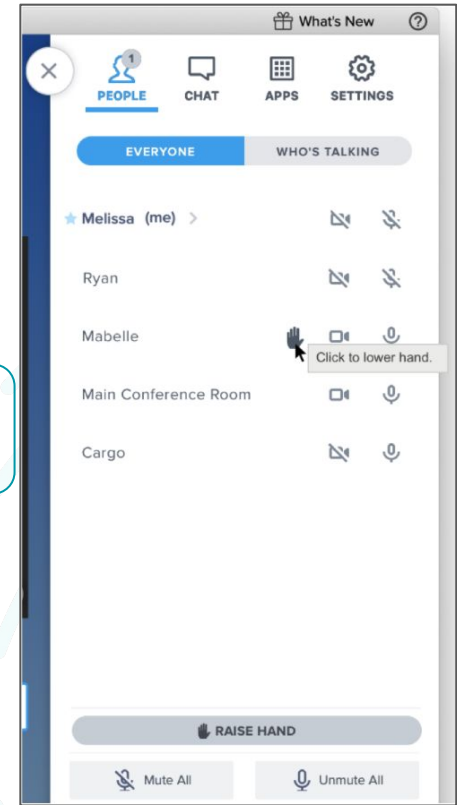


In BlueJeans, virtual participants should use the **BlueJeans chat** functionality to:

- **indicate you have a question and would like to unmute**, or

- type your question so that someone in the room can speak it.

BJ "raise hand" feature is hard for moderators to track, not preferred.



## Provide context

- What are Rubin Independent Data Access Centers (IDACs) and Scientific Processing Centers (SPCs)?
- How are IDACs and SPCs different from the Rubin Data Facilities?
- Why do we want IDACs and SPCs and when do we want them?

## Start connecting the community to IDACs and SPCs

- What capabilities will IDACs offer?
- What do scientists want from IDACs and SPCs?

## **IDAC and SPC Contributors:**

**UK Data Facility** – Bob Mann, George Beckett

**Argentina** – Mariano Dominguez, Diego Garcia

**Australia** – Jarrod Hurley, Sarah Brough,  
Stuart Ryder

**Brazil** – Carlos Adean, Luiz Da Costa, Julia  
Gschwend

**Canada** – JJ Kavelaars, Renée Hložek, Wes  
Fraser

**Croatia** – Tomislav Jurkić, Lovro Palaversa

**Denmark** – Christa Gall, Radek Wojtak, Hans  
Kjeldsen

**Spain** – Cristóbal Padilla, Ramon Miquel,  
Nacho Sevilla

**Japan** – Hisanori Furusawa, Naoki Yasuda,  
Masahiro Takada, Satoshi Miyazaki, Yutaka  
Komiyaama

**Korea** – Chang Hee Ree, Narae Hwang,  
Byeong-Gon Park

**Mexico** – Octavio Valenzuela, Luis Arturo  
Areña-López

**Poland** – Adam Zdrożny, Agnieszka Pollo,  
Paweł Pietrukowicz

**Slovenia** – Andrej Filipčič, Andreja Gomboc

# More introductions

Rubin In-Kind Program Coordinators

| Name (affiliation)                        | IPC Role  |
|---|---|
| <b>Steve Margheim</b> (NOIRLab) ✨         | Telescope time and Datasets (primary contact)                         |
| <b>Steve Ridgway</b> (NOIRLab) ☎          | NOIRLab/CSDC liaison, Telescope time and Datasets (secondary contact) |
| <b>Knut Olsen</b> (NOIRLab) ✨             | <b>IDACs and other computing resources</b>                            |
| <b>Aprajita Verma</b> (Oxford) ☎          | Software Development & Science Collaboration interactions             |
| <b>Greg Madejski</b> (SLAC) ✨             | Software Development & Community Engagement team Liaison              |
| <b>Phil Marshall</b> (SLAC) ✨             | Rubin Recipients  |
| <b>Agnes Ferte</b> (SLAC) ☎               | TBC (New IPC team member)   |
| <b>Sierra Villarreal</b> (Argonne/SLAC) ✨ | <b>IDACs Technical Coordinator (starting next week!)</b>              |

# What are IDACs and SPCs?

- A type of international In-Kind contribution (see In-Kind session 3:30 PM today for overview of In-Kind program by Verma, Margheim, and In-Kind team)
- IDACs are defined in [Handbook for Proposal Teams](#) and [Guidelines for Rubin IDACs](#)
- Centers that enhance or add value to the Rubin science enterprise through added computing, storage, data access, and software service options
- Types of IDACs:
  - Scientific Processing Center (Croatia)
  - Lite IDAC (Argentina, Australia, Brazil, Canada, Denmark, Spain, Japan, Korea, Mexico, Poland, Slovenia)
  - The UK and France “Full IDACs” are Data Facilities on par with the US Data Facility
- Each IDAC type has different requirements outlined in Appendix C of [Guidelines for Rubin IDACs](#); all must be accessible and respect Rubin Data Rights

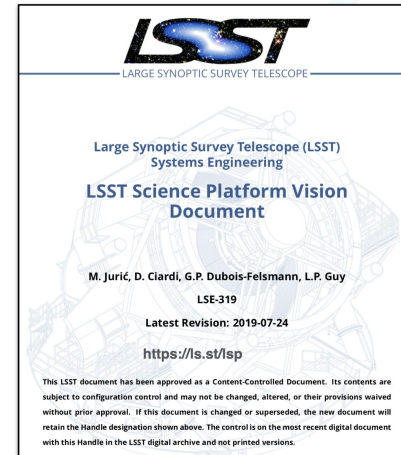
- Data Facilities (US, France, UK) host full copies of the Rubin data and run the Rubin Science Platform and all of its services
- Lite IDACs are scoped to store a subset of the data, typically a subset of the columns in the Object Catalog (so-called “Object Lite”, definition TBD)
- Offer a mix of CPUs, storage, GPUs, other datasets, and software services
- Many would like to deploy RSP services, but have to be careful here – RSP is built with full scale of data in mind, and Rubin is not scoped to support development and deployment of RSP at Lite IDACs
- Lite IDACs and SPCs will not simply be alternate portals to LSST data
- Strength of IDACs and SPCs will be the unique capabilities and expertise that they can offer

## **Important to understand the use cases for the specific IDACs and SPCs**



**A set of integrated web applications & services deployed at Data Access Centers through which the scientific community will access, visualize, subset and perform next-to-the-data analysis of Rubin Data products.**

- Enable peta-scale analysis of LSST data
- Exploratory analysis via browsing & visualisation
- Enable discovery – ‘bring the analysis to the data’
- Supports User-Generated product creation
- Integration with extant archives via IVOA protocols
- Collaborative working environment
- Provision of backend computation and analysis



[lsc-319.lst.io](https://lsc-319.lst.io)

*Slide credit: L. Guy*

## Portal Aspect

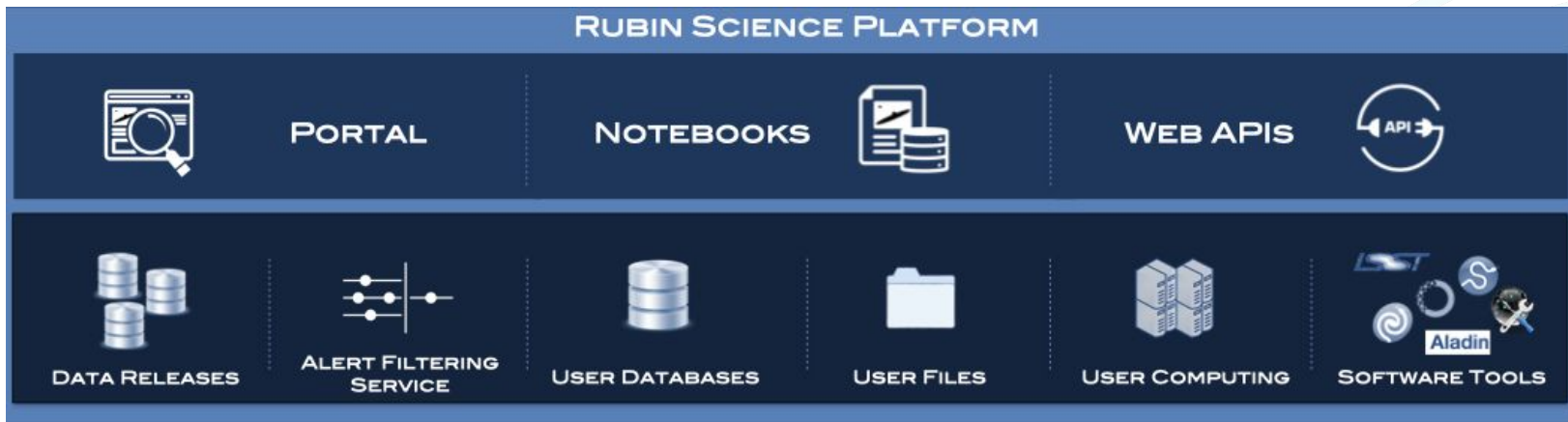
Exploratory analysis and visualization of the LSST archive

## Notebook Aspect

In-depth 'next-to-data' analysis and creation of added-value data products

## API Aspect

Remote access to the LSST archive via Virtual Observatory interfaces



Slide credit: L. Guy

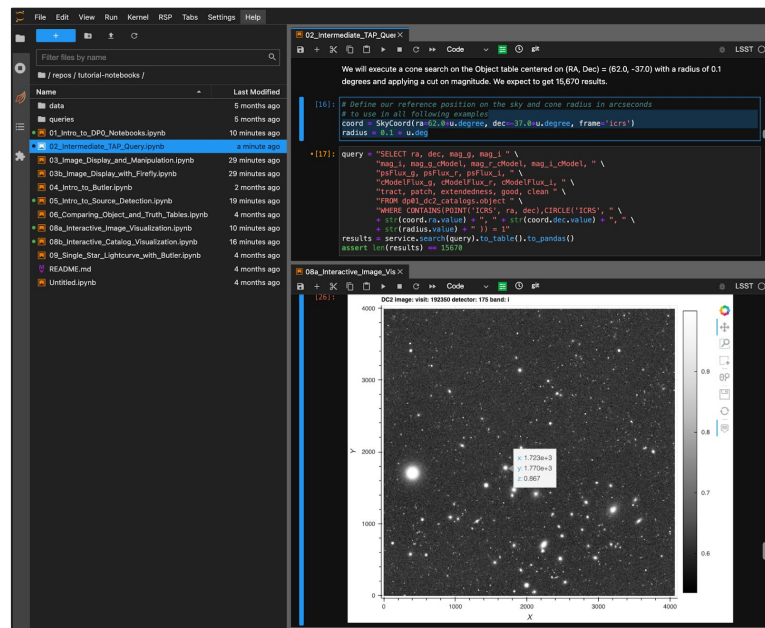
## Sophisticated ‘next-to-data’ analysis

**Enables science discovery** by ‘bringing the analysis to the data’, avoiding the download of large volumes of data.

**User environments** with pre-installed libraries:

- LSST science pipelines, Anaconda, AstroPy, etc.
- Users can install additional tools.
- Subsetting via forms, ADQL\*.

**Data Access Centers** provide computation & analysis resources.



The screenshot shows a Jupyter Notebook environment. The top panel displays a code cell with the following content:

```

[10]: # We will execute a cone search on the Object table centered on (RA, Dec) = (62.0, -37.0) with a radius of 0.1
degrees and applying a cut on magnitude. We expect to get 15,670 results.

# Define our reference position on the sky and cone radius in arcseconds
# to use in all following examples
coord = SkyCoord(ra=62.0*u.degree, dec=-37.0*u.degree, frame='icrs')
radius = 0.1*u.degree

[11]: query = """SELECT ra, dec, mag_p, mag_1, \
    mag_2, mag_3, mag_4, mag_5, \
    psf_fwhm_r, psf_fwhm_x, psf_fwhm_y, \
    model_fwhm_r, model_fwhm_x, model_fwhm_y, \
    tract_patch, extendedness, good, clean """
FROM astrometry_catalog.objects """
WHERE CONTAINS(POINT('ICRS', ra, dec), CIRCLE('ICRS', \
    %s, %s, %s)) = 1 """ % (coord.ra.value, coord.dec.value, radius)
results = service.search(query).to_table().to_pandas()
assert len(results) == 15670
    
```

The bottom panel shows a visualization of the star field resulting from the query. The plot is titled "DC2 Image: visit\_192389 detector: 173 band: i". The axes are labeled 'X' and 'Y', both ranging from 0 to 4000. A color bar on the right indicates intensity from 0.6 to 0.8. A white box highlights a specific star with the following coordinates:

```

(1.723e+3, 1.770e+3, 0.667)
    
```

Slide credit: L. Guy

# Why do we want IDACs and SPCs?

- User computing at DACs sized to “10% of the total LSST data processing capacity (computing and storage) for the purpose of scientific analysis of LSST data and the production of user Data Products”
- We want to expand available resources beyond this 10%
- IDACs and SPCs host key data products and provide additional **free** capacity in computing and storage

|                             | Storage TB/User | Cores/User |
|-----------------------------|-----------------|------------|
| US-funded US User Resources | 0.8             | 2          |
| In-kind US User Resources   | 4.8             | 4          |
| New US User Resources       | 5.6             | 6          |
| Increase factor             | 7               | 3          |

- Access to specific data sets and services
- A network of collaborating providers with valuable expertise
- Ability to tailor to needs of specific use cases and communities

# When do we want IDACs and SPCs?

## Rubin Operations Top Milestones

- **2021-06-30** Deliver Data Preview 0.1 (DP0.1) (L1-RO-0040)
- **2022-06-30** Complete Delivery of Data Preview Zero (L1-RO-0050)
- **Mar 2024 - Jul 2024** Complete Delivery of Data Preview One (DP1) (L1-RO-0060) (= Optical testing on TMA complete + 6 months)
- **Jun 2024 - Oct 2024** ("Survey"/"Full") Operations Begins (L1-RO-0100) (= Operation Readiness Review Complete + 1 day)
- **Jun 2024 - Nov 2024** Survey Start (L1-RO-0110) (= ("Survey"/"Full") Operations Begins + 1 months)
- **Dec 2024 - Mar 2025** Complete Delivery of Data Preview Two (DP2) (L1-RO-0070) (= Mini Survey 2 Complete + 6 months)
- **Oct 2025 - Jan 2026** Complete Delivery of Data Release One (DR1) (L1-RO-0120) (= Survey Start + 12 months)

- Main change: LSST starts Fall 2024 (Project COVID-delayed by another 6 months)
- Research grant proposal writers and science planners: you can expect
  - DP1 (ComCam) data release Spring/Summer 2024 (ie. Year 1 of 3)
  - DP2 (LSSTCam SV) data release Winter 2025 (ie Year 2 of 3)
  - **DR1 (First LSST data) Fall 2025 (ie Year 3 of 3)**

*Slide credit: P. Marshall*

# Community-developed use case areas

---

- [LINCC Frameworks](#) team, with funding from Schmidt Futures Foundation and support from LSST Corporation, organized a workshop to sketch out areas for software development and computational needs
- Many interested audiences, including IDAC and SPC teams
- White paper now available on [arXiv](#) (also see Mandelbaum [Community post](#))
- Technical areas identified for discussion of software development: cross matching, selection functions, time series, image reprocessing, image analysis, and photometric redshifts
- Discussion of time series software in [“Data to Software to Science”](#) session earlier today (Connolly, Kubica, Mandelbaum)
- Discussion of [LSST photometric redshifts](#) Wednesday
- Some areas could be good fits to particular IDACs and SPCs

# What will IDACs and SPCs offer?

---

- Contributed talks by Croatia, UK, Brazil, and Canada

We appreciate that you will have questions, ways to get help

- Comprehensive [in-kind FAQs](#) are posted on [community.lsst.org](https://community.lsst.org)
- Also follow the [IDACs category on Community](#)
- **Office hours:** please [book slots here](#) for meetings with 1 or more IPCs
- Not sure whom to contact: write to [jikh@lsst.org](mailto:jikh@lsst.org)

### FAQ: In-kind Contributions and Data Rights Agreements

Support in-kind



drphilmarshall Phil Marshall LSST

2 20h

The In-kind Program Coordination (IPC) Team is working with the in-kind contribution teams, their Program Managers and recipient groups to implement the approved in-kinds. The [Handbook for In-Kind Proposal Teams, RDO-31](#), and the [Manual for In-Kind Contributors and Recipients, RDO-41](#), are useful references, but the IPCs are collecting FAQs as well. We'll update this post with more questions as we go.

#### Data Rights, Agreements and Access

- ▶ Q: What is the status of the data rights agreements?
- ▶ Q: Can PI-ship be reassigned within a team that asks for more than 1 PI?
- ▶ Q: What is the process for defining data rights holders, and how do we know how many PI slots we have?
- ▶ Q: Does 1 PI's worth of data access rights via in-kind contributions also come with 4 associated JAs, as in the previous MoAs with LSSTC?
- ▶ Q: Am I correct in thinking that our PIs, and their JAs, automatically get the same access to data processing facilities as US/Chilean astronomers?
- ▶ Q: On page 25 of the Handbook it says: 'Due to the diversity conditions of proposing groups and funding sources we do not distinguish between permanent staff or temporary staff in the exchange rate.' Does this imply that temporary staff who lead contributions will be considered as PIs if data rights are approved?
- ▶ Q: Could a JA be at any institution as long as the associated mentor is at an institution with data rights?
- ▶ Q: Did I see 2 different meanings for "IPC"? One for "In-kind", and one for "International"?
- ▶ Q: Once everything is sorted with whom will our data rights agreement actually be, Rubin Observatory, SLAC, NOIRLab or ...?
- ▶ Q: If a country has 5 different in-kind contributions to different SCs, will they have 5 different data rights agreements (DRAs) to sign?
- ▶ Q: With which agency or managing organization will IDAC DRAs be signed?
- ▶ Q: Will it be possible to extend a data rights agreement after it has been signed?
- ▶ Q: How often do you expect amendments to the data rights agreements?

#### Proposing Contributions



## What do scientists want from IDACs and SPCs?

For your use case:

- What LSST and other data do you need? Recall that Lite IDACs generally sized with a light version of the Object catalog in mind, but many have significant user storage volumes
  - What information should an Object Lite data product contain?
  - Database + TAP query vs. Parquet
- What software will you be running? Will you want an interactive environment or mainly be submitting batch computing jobs?
- Does the computational workflow place particular demands on the hardware or software services?