

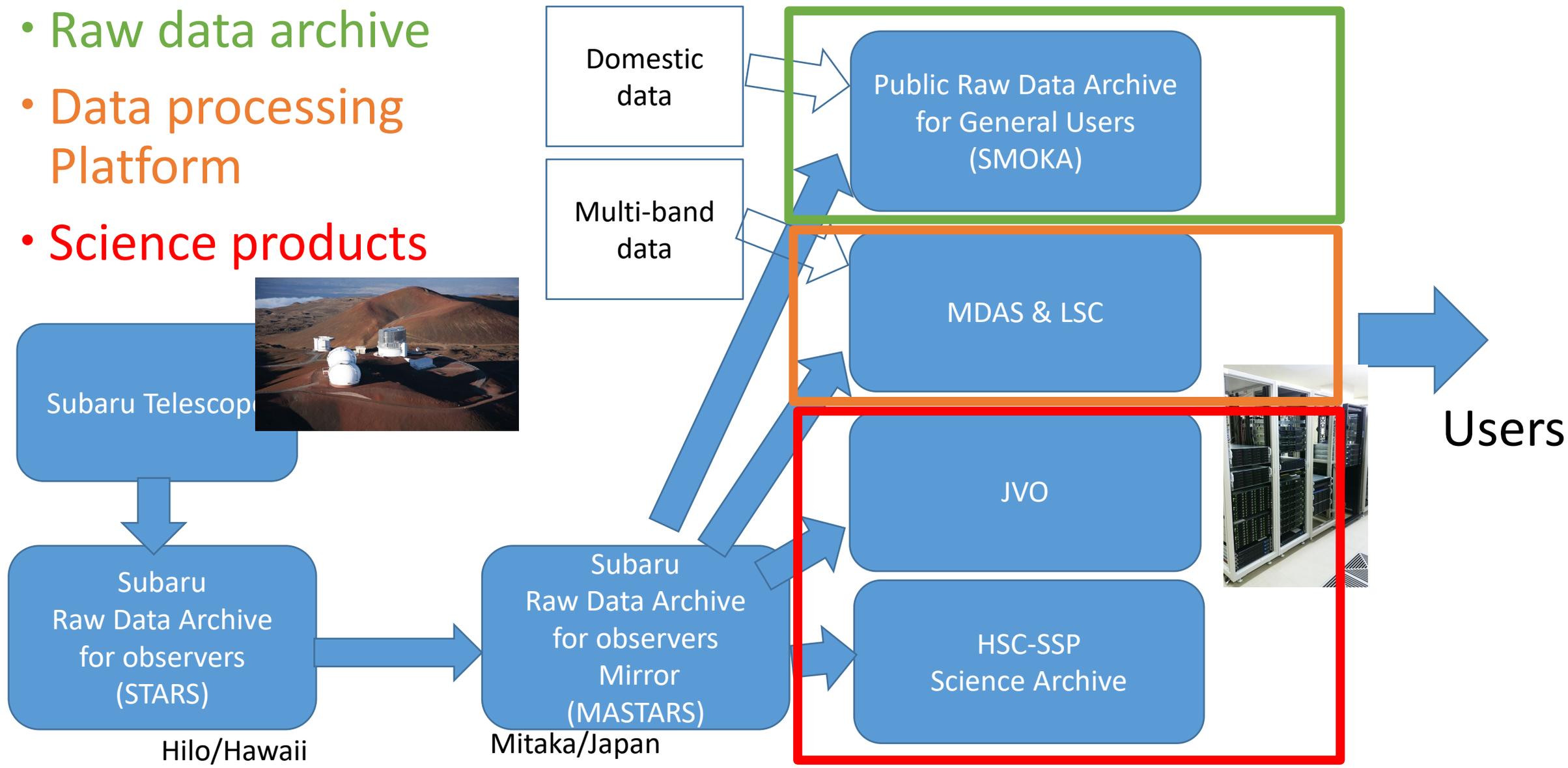
Science User Support for Subaru Telescope Observing Data at Astronomy Data Center (ADC)

Hisanori Furusawa

for ADC/Subaru Telescope open-use support team
1/11/2022 Subaru Users Meeting

Data Flow and Services for Subaru Data Sciences

- Raw data archive
- Data processing Platform
- Science products



Raw Data Archive

- STARS/MASTARS
 - Observatory's archive for observers and engineering
 - infrastructure, conveying data to public archive
- SMOKA (see Onozato-san's talk on Day2)
 - Public raw data archive for general users
 - Subaru Tel. and other domestic instruments
- Subaru & ADC next-gen. raw data archives (see Takami-san's talk; ~2024—)
Under discussion for:
 - Efficient validation and quality management of data
 - Unified workflow for consistent data management throughout data lifetime (from acquisition to archive)
 - Better usability of data and services

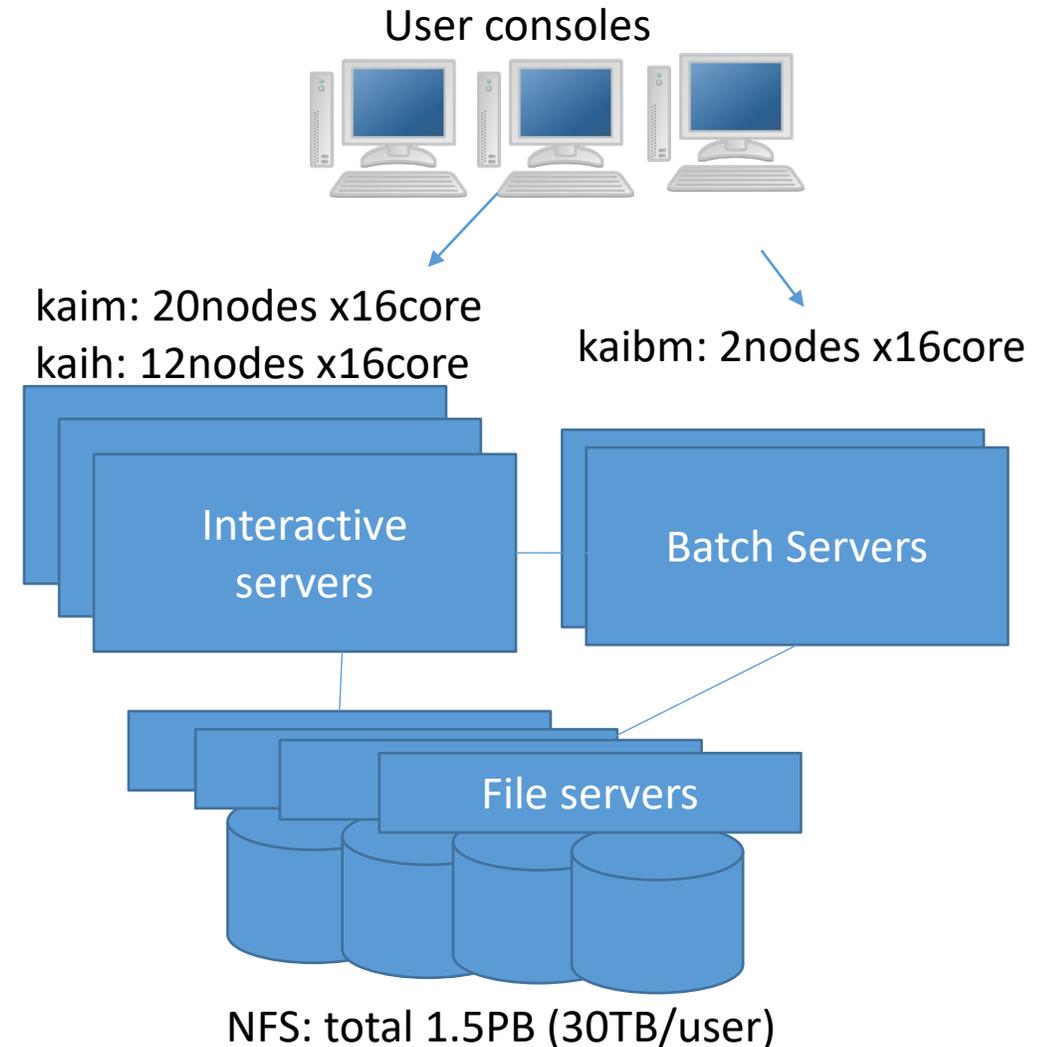
The screenshot shows the STARS 2.0 Subaru Telescope Archive System interface. At the top, there is a header with the text "STARS 2.0" and "Subaru Telescope Archive System" alongside logos for Subaru and the archive system. Below the header, there is a navigation menu with "MySTARS", "Search by", "Utilities", and "Exit". A search bar is present with the text "Search by: Proposal_ID". Below the search bar, there are several input fields for search criteria: "Proposal ID (o20406)", "FrameID Range (RCA00000010)", "ObsDate Range (2007-02-07)", and "Calibration Date Range (2007-02-07)". There are also radio buttons for "Instr." (ALL, HSC, VGW) and "Calb: none". A "Store Query" button is visible at the bottom left of the search area.

The screenshot shows the SMOKA Archive Advanced Search interface. At the top, there is a title "SMOKA Archive Advanced Search" and two links: "Click here for SUP Search (Suprime-Cam data Search)" and "Click here to know how to search". Below the title, there is a "Search Conditions" section with several input fields and buttons. The "Object Name (for name resolve)" field has a "Resolve" button. The "Coordinate System" section has a dropdown menu set to "Equatorial" and a "Circular" radio button. The "Field of View (arcmin)" section has a dropdown menu set to "auto" and a "Rectangle" radio button. The "Observation Date" section has an "Observation Date" input field. The "Exp Time (sec)" section has an "Exp Time" input field. The "Observer" section has an "Observer" input field.

Data Processing Platform

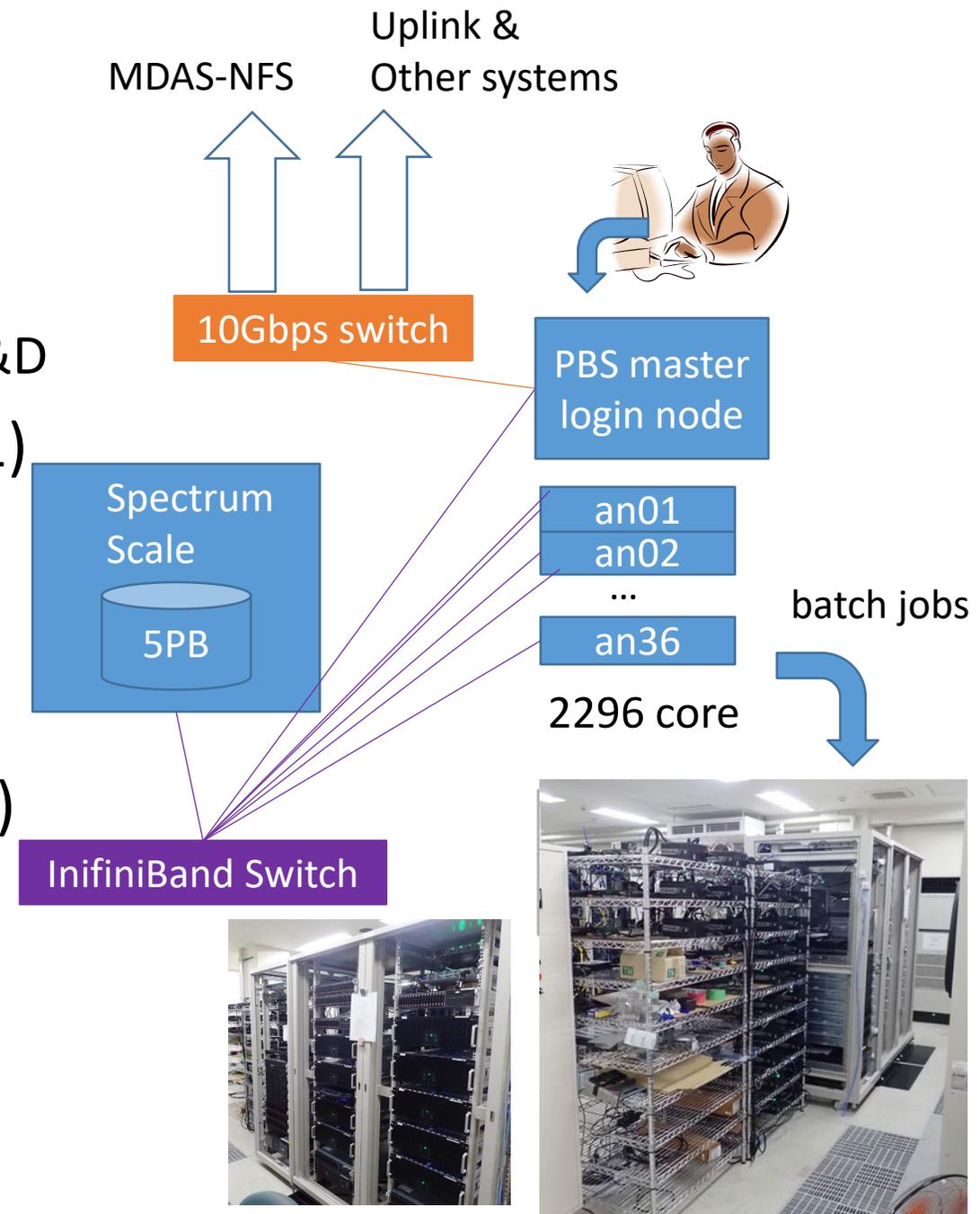
1) MDAS https://www.adc.nao.ac.jp/MDAS/mdas_e.html

- General **interactive data analysis** work
incl. scientific post-processing,
writing papers
- Plus a small set of batch jobs
- large shared area NFS
- VPN access
- Various popular/common software packages
- Any researchers can apply for use
 - Foreign researchers may require some approval processes (Japanese contact & import control)



2) LSC

- Intensive processing for huge data sets
 - Both for Open-use and NAOJ's production+R&D
- PC cluster (2296core/40node as of FY2021)
- Fast filesystem (5PB Spectrum Scale)
- Accounts and NFS area shared with MDAS
- Prioritized access to resources by user status through batch job system (PBS)
- Has basic software packages for HSC data (hscPipe, associated calibration files)



Batch Job Queues and Priorities

- Resources assigned based on job queues with different priority
- Open-use observers assigned relatively-high priority for 1 year

Queue Name	User	N.Core	Job Lifetime	Term	Priority
qssp	Subaru/ADC	~2000	unlimited (~2months)	x 2/ year	High
qm	Open-use HSC PIs	112	15days	1 year	Mid
ql	General	32	7days	-	Low
qhm (Large mem)	General (TBD)	32	15days	-	Mid
Test	test	4	10min	-	High

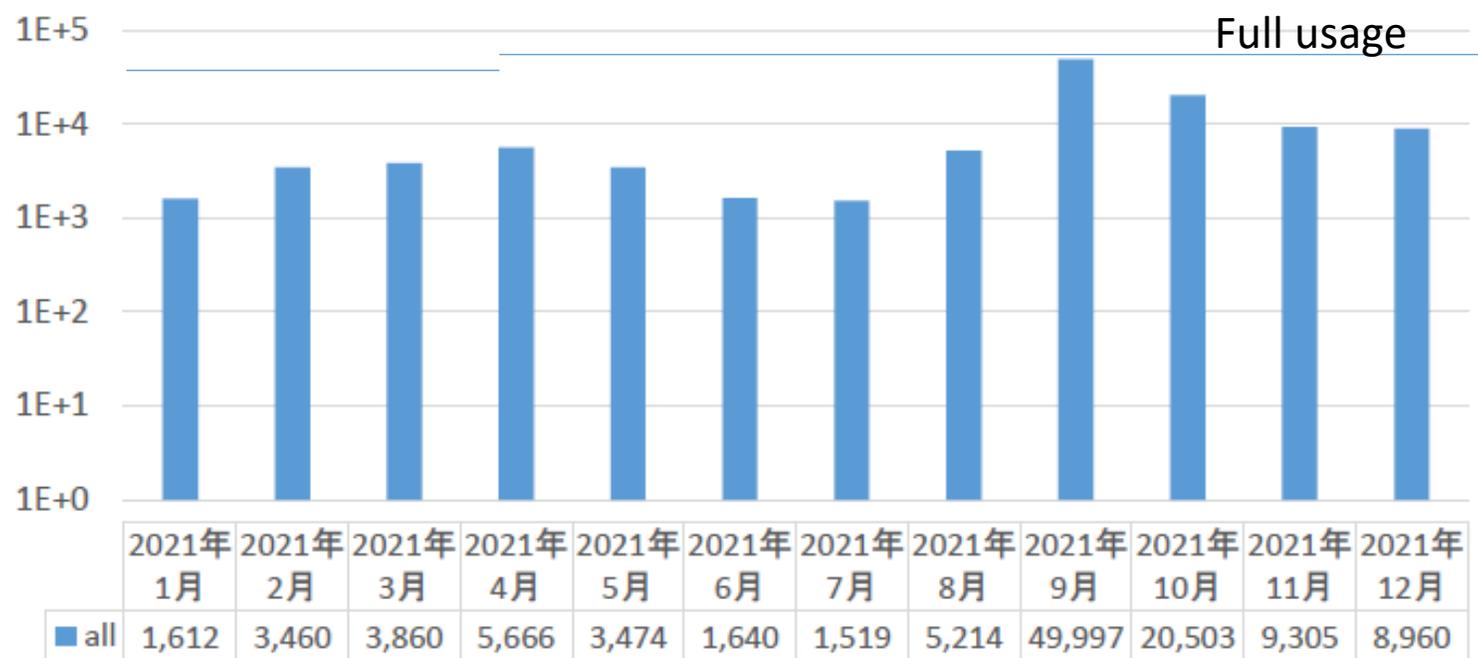
Who can use LSC

1. **HSC Observers** in the current semester
 - 1 year priority access
2. **General users**
 - Per registration every year (condition ~ MDAS)
 - HSC data processing
 - Soon be extended to general observing data
3. **ADC/Subaru Telescope's Operation and R&D**
 - SSP (HSC and PFS in the near future, and beyond?)
 - Open-use support & strategic use (e.g., time-critical use)
 - Database/pipeline developments
 - Other commitments for external projects (e.g., Euclid UNIONS, possibly Roman-J and Rubin etc.)

LSC System Usage

- CPU occupation
 - <~20% CPU time on average
 - 70-80% in peak production phase
 - Room for additional processes
- Storage usage
 - ~4.0PB / 5PB
 - Has 4 gen. of HSC-SSP files

Total CPU time [core*days] in each year-month



Feedback at ADC Users Meeting (MDAS-LSC Connection)

#1 May.19.2021, #2 June.24.2021

- Various versions of HSC data analysis pipeline (hscPipe)
 - 4 (PDR1), 5(S17A), 6(PDR 2), 7(S19A), 8(PDR3)
- More general use
 - soon extending the system usage to general observing data
- Better file sharing and transfer w/ LSC to MDAS
 - NFS mounts from lsc (login node) to MDAS-NFS
 - For more efficient interactive work
- Transporting files to users' sites
 - discussing a feasibility

FY2022 onward

- Optimization of the system usage & func. for open-use & strategic activities
- Replacing/Updating the system components

Science Archives for Science-ready Products

- HSC-SSP Data Release

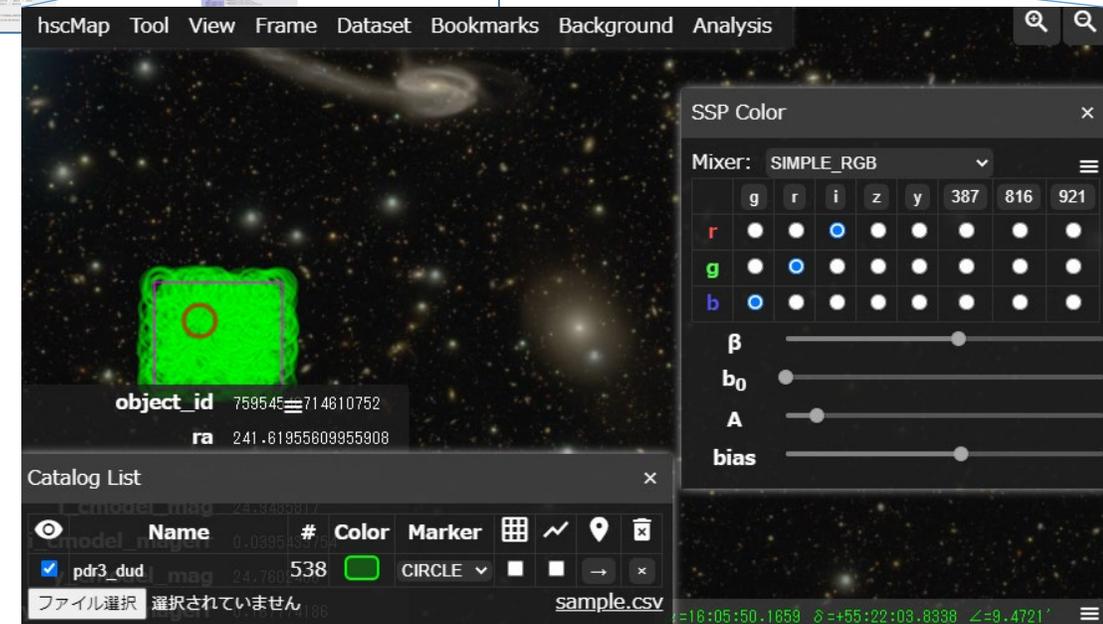
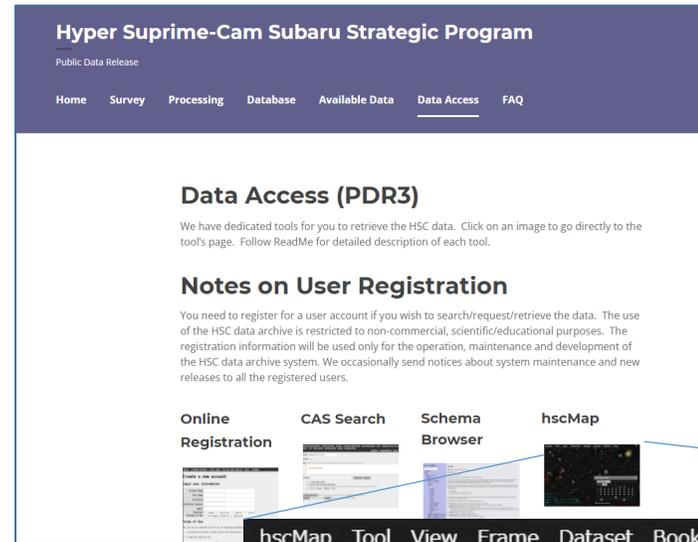
- Run with Subaru Telescope

- Development
 - Data Production
 - Data services
 - Helpdesk

(see posters by
Morishima+: DB develop.
Harasawa+: Public data analysis)

- JVO portal

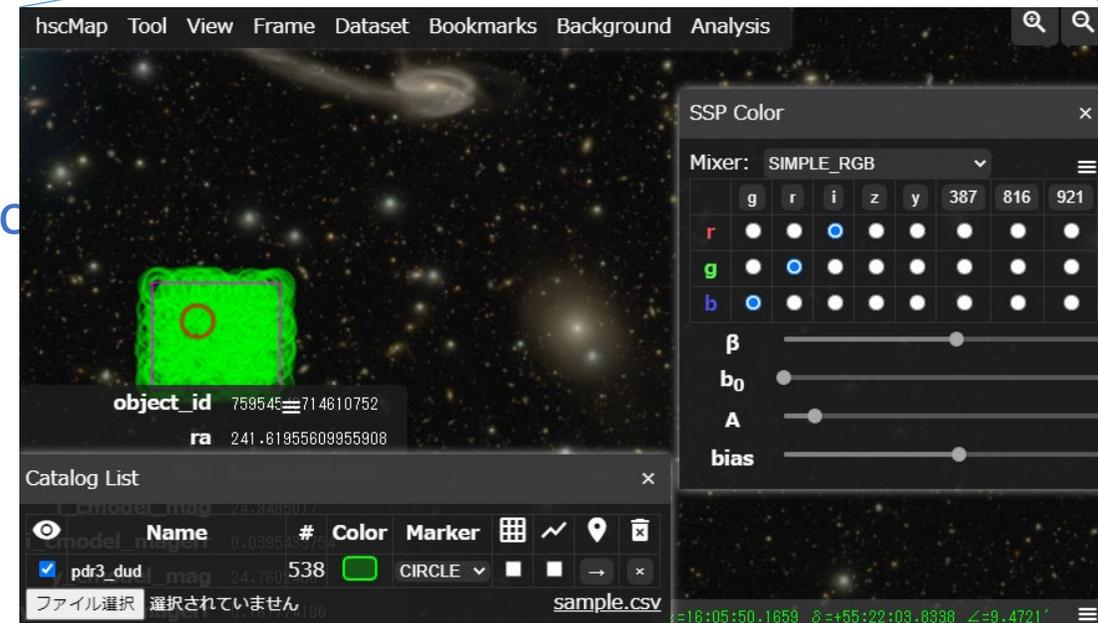
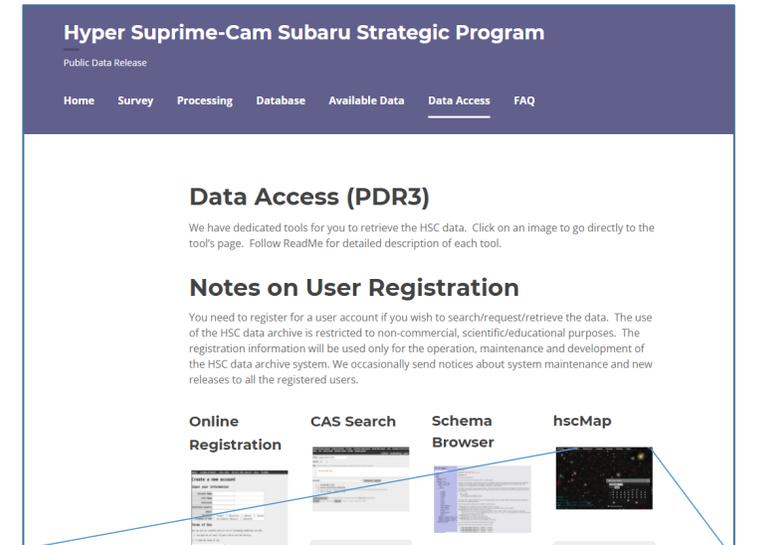
- ALMA
 - Gaia
 - etc.



HSC-SSP Public Data Release

<https://hsc-release.mtk.nao.ac.jp/doc/>

- Joint operation across ADC and Subaru Telescope
- 3 Public Releases (PDR3 2021.8)
 - ~1800 registered users
 - Next PDR (TBD) will be the final SSP-PDR
- Services
 - Processed and calibrated images and catalogs
 - DAS for image downloads
 - CAS with SQL database query
 - Python API for query and downloads
- Will be extended to combined data service with PFS data



ADC Future Prospects and Ongoing Preparation

- Caution: Under developing within ADC and not yet established
- In cooperation with Subaru Telescope

Make Raw and Science Data Products Legacy to the Community

- Establish and Operate a next-generation raw data archive
 - Permanent usable raw data archive
- Continue efforts of QA and Science data production
 - Update HSC-SSP data products and calibration, to be extended to other public data
 - Cooperate Suprime-Cam science data archive with CADC
- Offer data services for best science use of archived raw data and products
 - Develop Science Platform and Fast Databases
 - Develop Pipeline algorithms for Subaru Telescope community
 - Increase data coverage with multi-waveband data products



ADC as a base facility of data sciences for Subaru & Japan community

Contribution for Rubin collaboration

- Rubin Independent Data Archive Center (Lite-IDAC) for LSST catalog products
 - A regional data archive as a partner of Rubin collaboration
- NAOJ: preparing **WFSI group** (Subaru+ADC+others) for HSC+Rubin+Euclid+..
- Trying to request ~13 LSST PIs for Japanese community (c.f. Day 2 discussion)

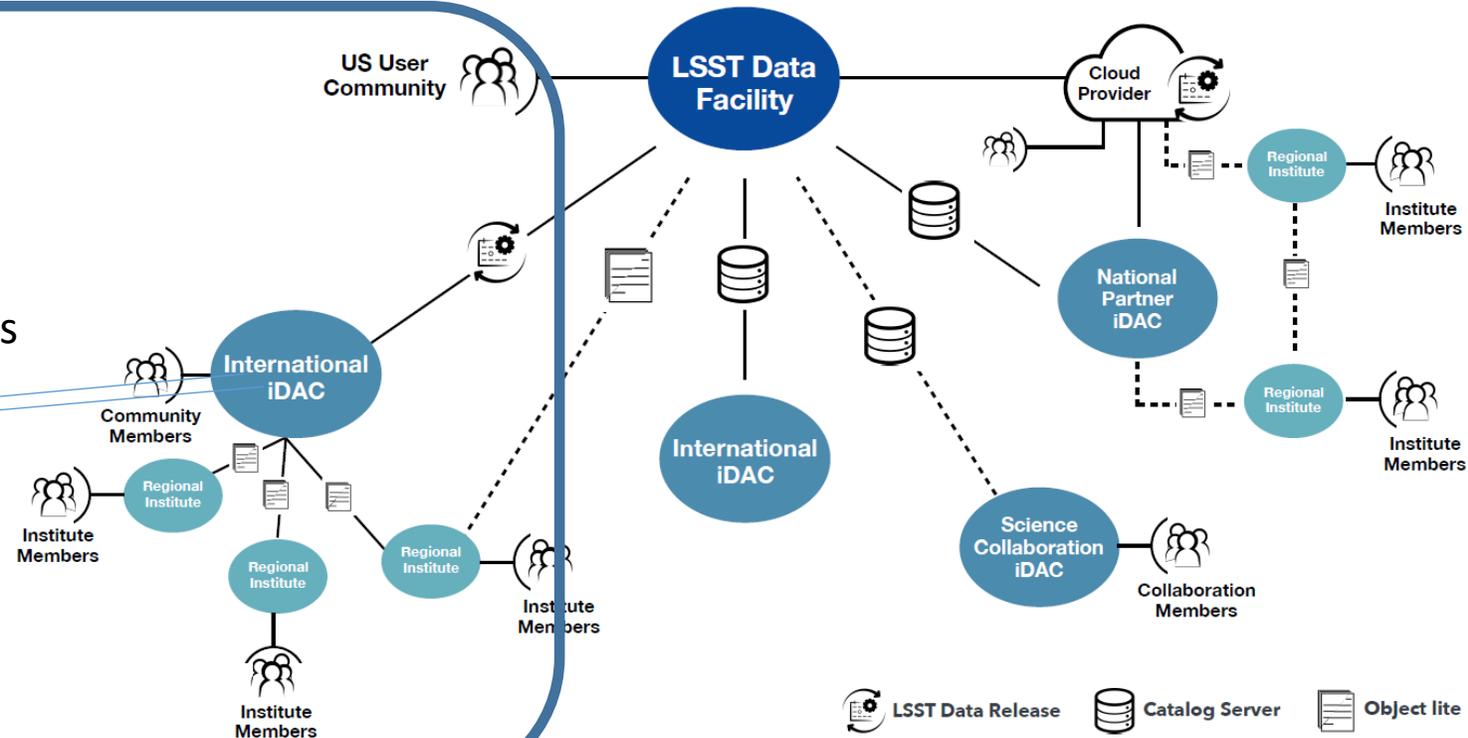
A proposal of Japanese Regional IDAC

Kavli-IPMU



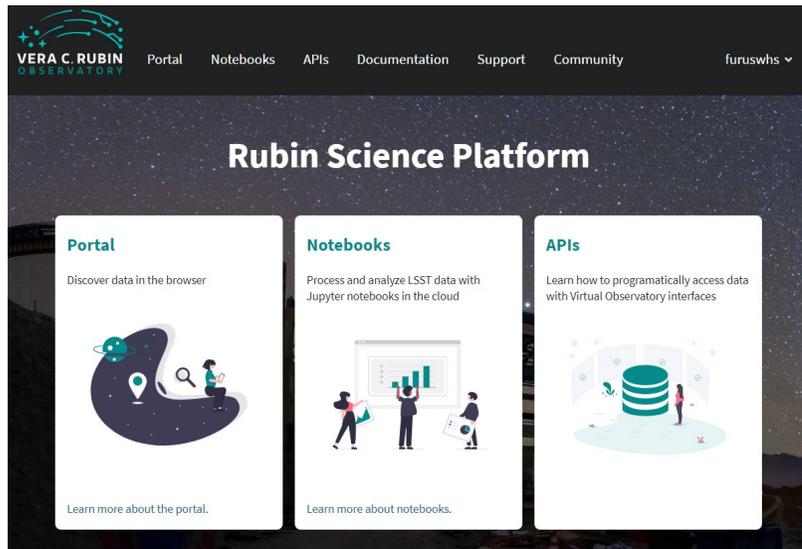
LSST
Subaru Telescope
Other joint programs

NAOJ-ADC

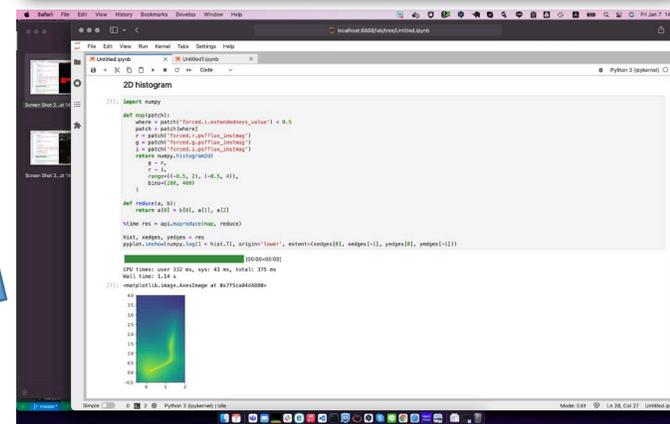
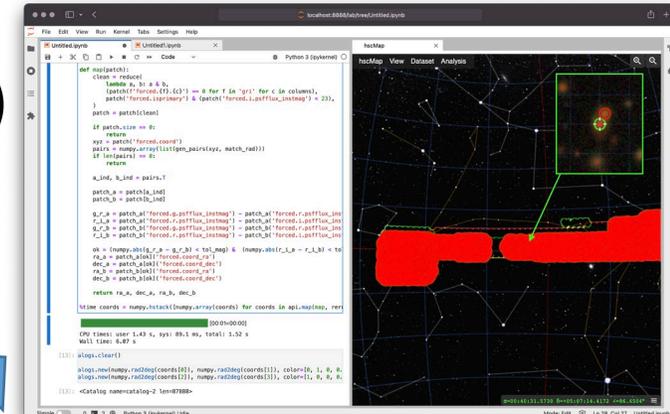
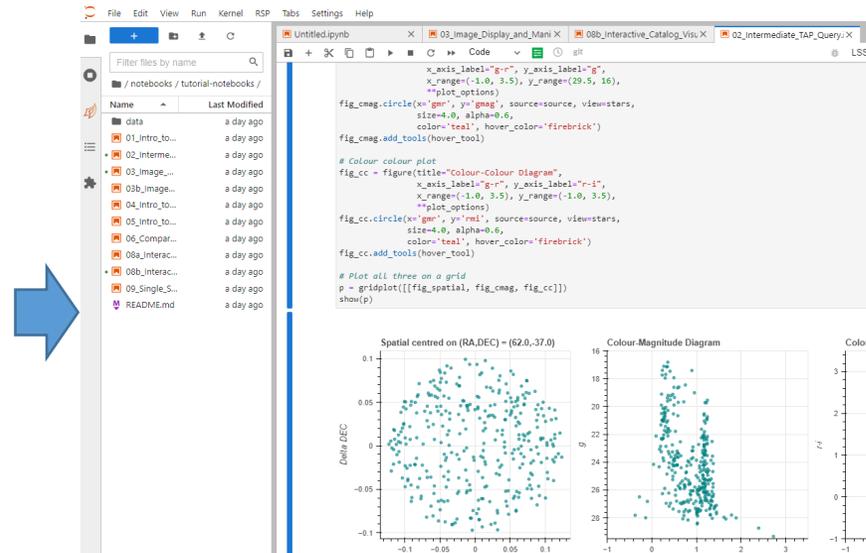


Rubin Science Platform (RSP) on IDAC + ADC/Subaru Value-added Datasets/Functions

- RSP: Allow efficient science processing of catalogs (and images) with Jupyter notebook + Python APIs + VO-compliant catalog query
 - Possibility of additions:
 - Other science data sets available at NAOJ (SSP, coadded and temporal catalogs, Euclid WISHES, etc..)
 - High-performance databases and interactive browsers
- addition of high-performance DB access



Snapshot from RSP DP0.1



Summary

- ADC offers/cooperates services of **raw data archive**, **data processing platform**, and **science products** for the Subaru community
 - STARS <https://stars.naoj.org/>
 - SMOKA <https://smoka.nao.ac.jp/index.jsp>
 - MDAS https://www.adc.nao.ac.jp/MDAS/mdas_e.html
 - LSC https://www.adc.nao.ac.jp/LSC/lsc_e.html
 - JVO <http://jvo.nao.ac.jp/index-e.html>
 - HSC <https://hsc-release.mtk.nao.ac.jp/doc/>
- ADC future roadmap being discussed in cooperation with Subaru Telescope includes:
 - Develop and operate **new-generation raw data archive**
 - Continue **data production for public data** and **update QA and calibration**
 - Develop and operate data services for **science products**, incl. high-performance DBs
 - **Host Rubin RSP+IDAC w/ various multi-wavelength data sets and software assets**