

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

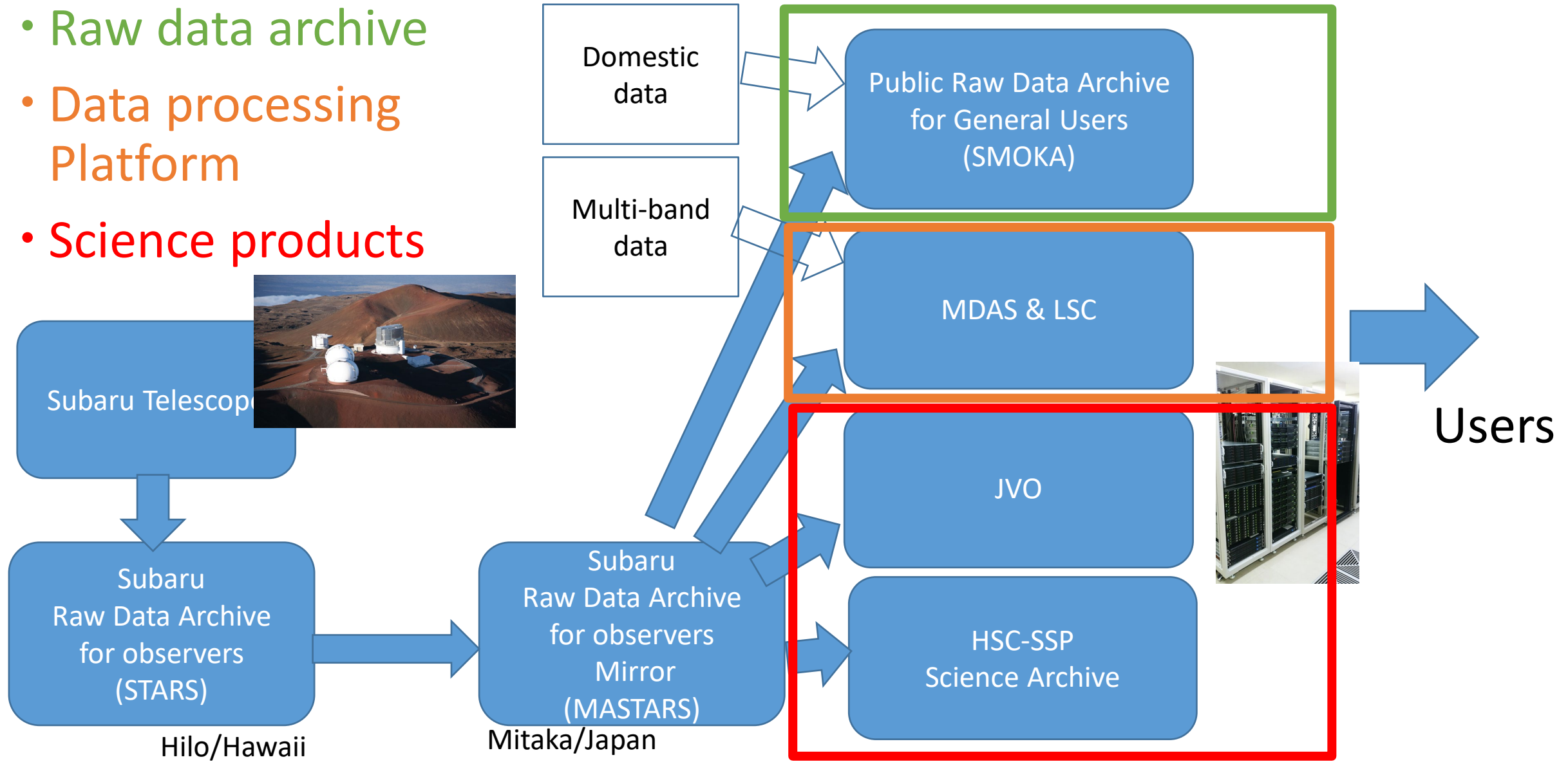
Hisanori Furusawa

for ADC/Subaru Telescope open-use support team

1/31/2023 Subaru Users Meeting FY2022

Data Flow and Services for Subaru Data Sciences

- Raw data archive
- Data processing Platform
- Science products



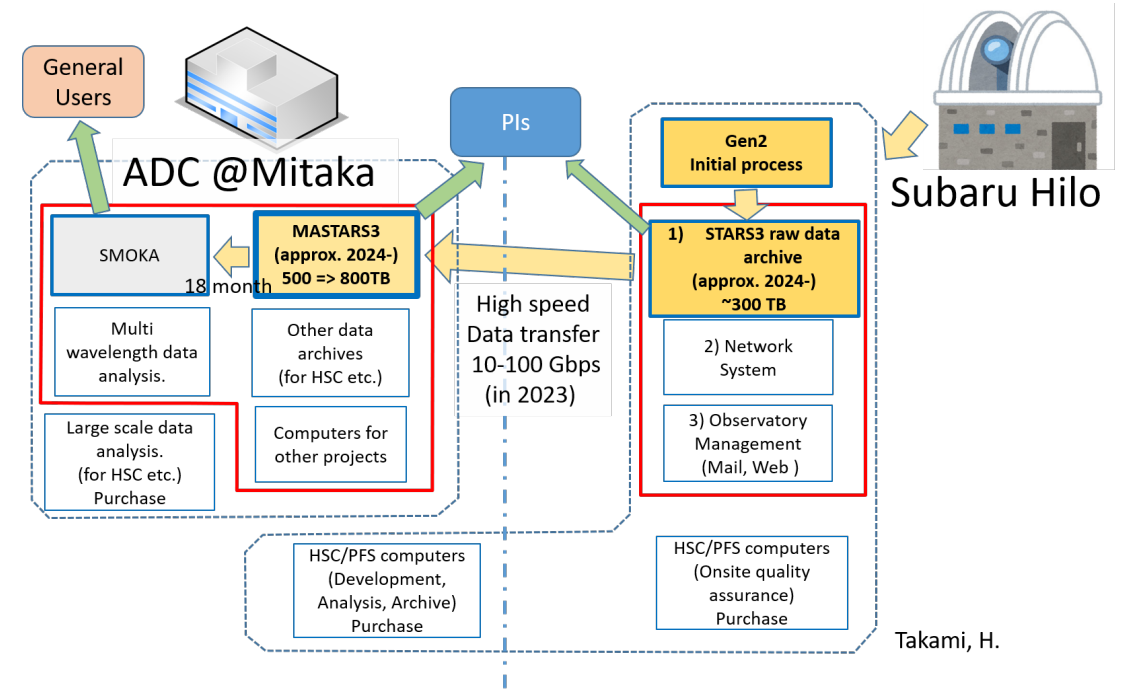
1. Raw Data Archive – This fiscal year

- STARS/MASTARS (for observers)
 - Procurement and preparation for the next system (in Mitaka, Hilo) [to maintain STARS2](#) – continued
 - Discussion of the [next-generation Subaru archive system](#) across Subaru-to-ADC (STARS3)
- SMOKA (for public users)
 - See Onozato-san's talk on Day3 PM
 - [REACH](#) on IRD data newly released
 - [SCExAO](#) data ([MEC](#), [FastPDI](#)) and [PFS](#) data under discussion
 - Serving Seimei, Tomo-e-Gozen data
 - Preparation for the next rental contract of the ADC computers

The image displays two web-based interfaces for astronomical data archiving. The left interface is the 'STARS 2.0 Subaru Telescope Archive System'. It features a header with the system name and a navigation bar with links like 'MySTARS', 'Search by', 'Utilities', and 'Exit'. Below this, there's a search section with a 'Search by: Proposal_ID' dropdown and a search box. A table lists search results with columns for Proposal ID, Instrument, and other details. The right interface is the 'SMOKA Archive Advanced Search'. It has a header with the title and two links: 'Click here for SUP Search (Suprime-Cam data Search)' and 'Click here to know how to search'. The main section is titled 'Search Conditions' and includes fields for 'Object Name (for name resolve)', 'Coordinate System' (Equatorial, Equinox, J2000), 'Field of View (arcmin)' (Circular, Rectangle), 'Observation Date', 'Exp Time (sec)', and 'Observer'. It also includes a 'Resolver' section with radio buttons for 'SIMBAD', 'NED', and 'Don't Resolve', and a 'Radius(arcmin)' field.

Subaru-ADC Next-generation Raw data archive

- Subaru prioritizes observing and proprietary data operation for observers
- Long-term archive is primarily operated by ADC, in coop. with Subaru
- Timeline (TBD)
 - ADC runs MASTARS on next rental system with STARS on STN6 (some time FY2024?-)
 - Integrated operation across Subaru and ADC – schedule TBD)

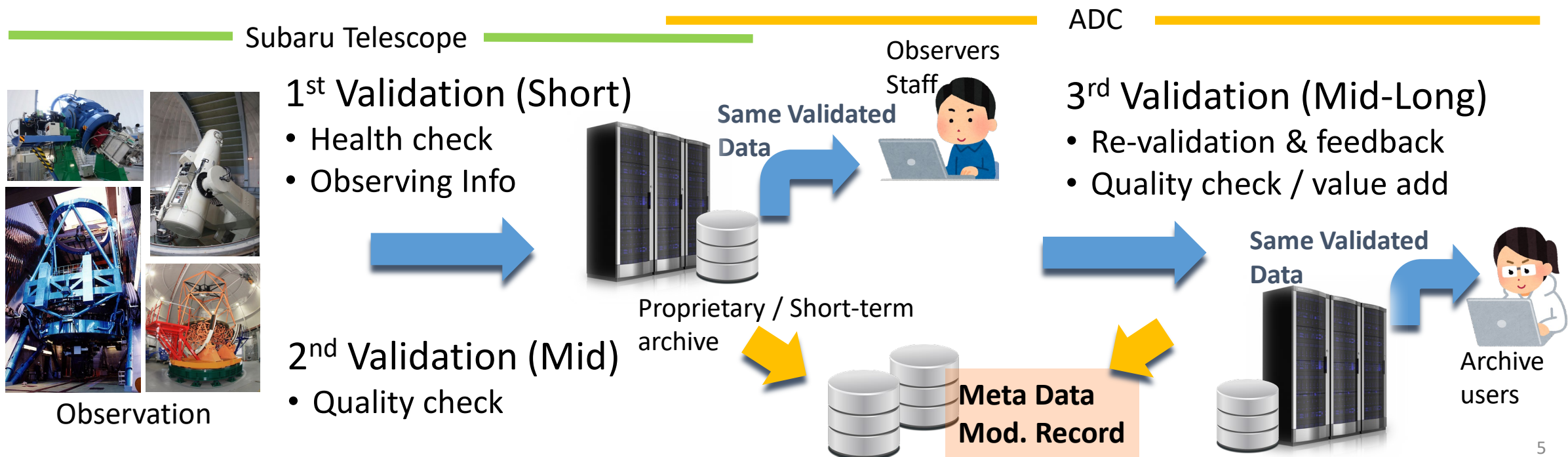


Caution: Transition Timing is still TBD

Takami, H.

Data Management Concept for Better Usability - TBD

- **Integrated single workflow** from Summit to Mitaka to the public release throughout the data lifetime
- **Proper validation and** health/quality check in each node (Pyo+)
- Proper data compression – SCEXAO, HSC, PFS, ... (Pyo, Takami, +)
- Sharing a **single metadata** and modification database

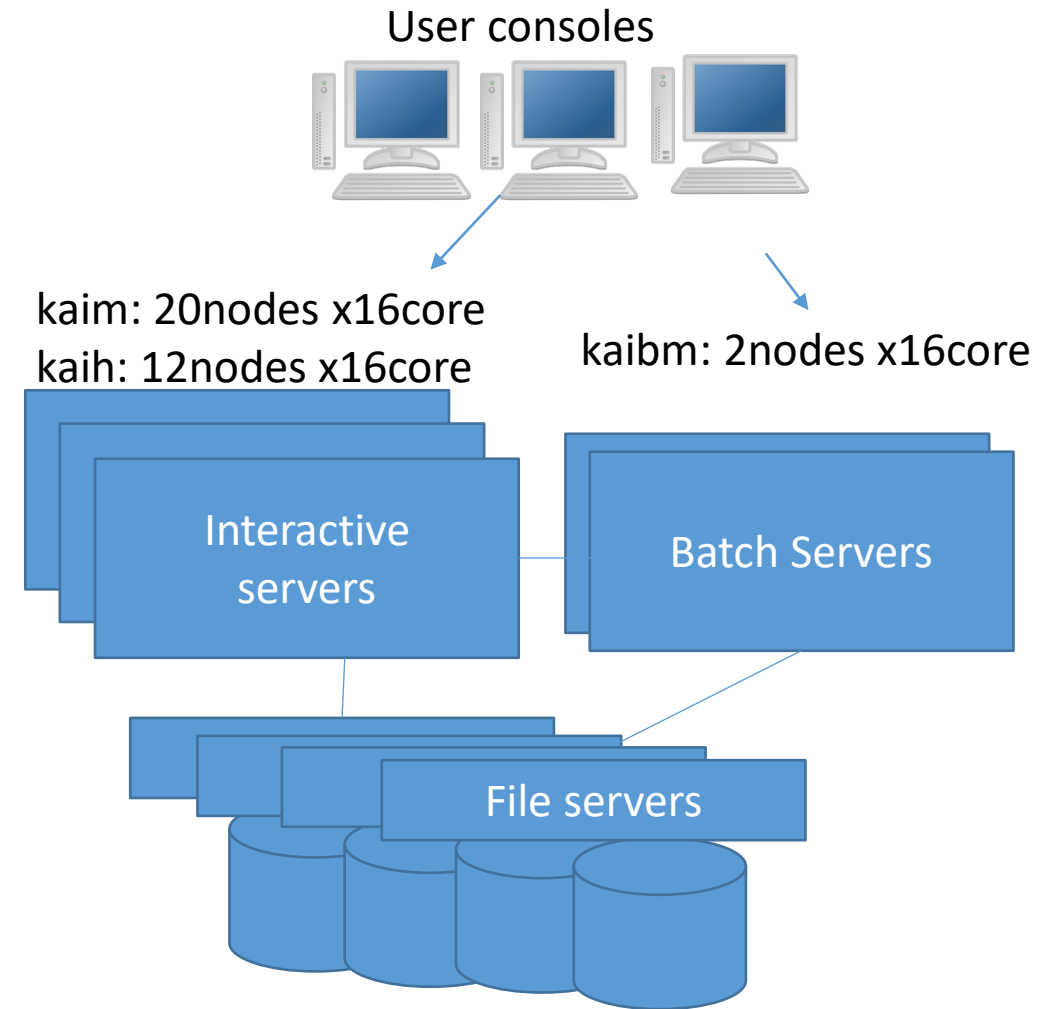


2. Data Processing Platform – This fiscal year

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

For general **interactive data analysis** work
+ a small set of batch jobs

- **New 1.2PB NFS area** appended for data transition & backup
- Preparing for the **computer update** with the contract renewal
 - A considerable fraction of CPUs are going to in-house support
 - Your comments will be welcome in optimized system design



NFS: total 1.7PB (30TB/user) + new NFS backup area 1.2PB

2) LSC

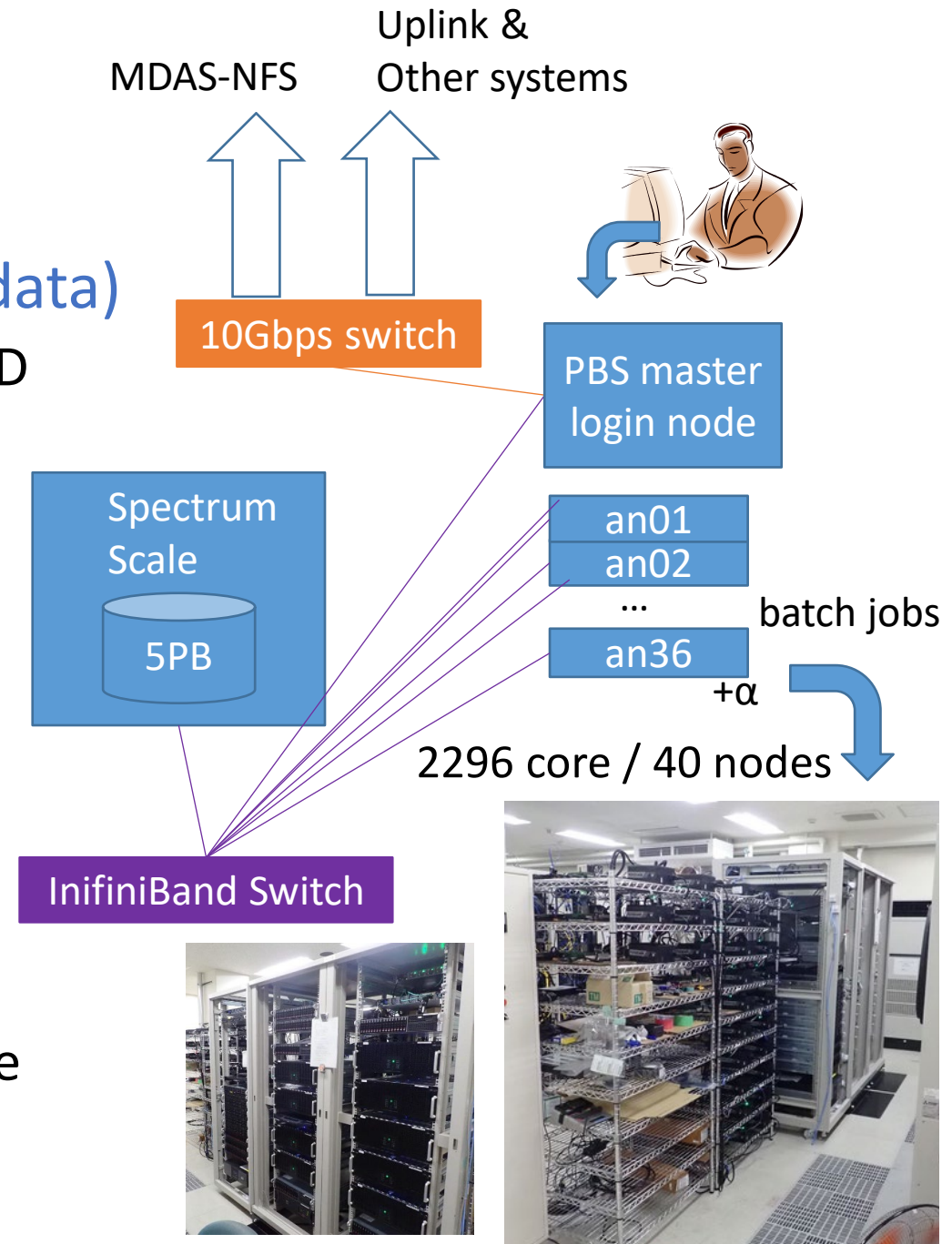
Intensive processing for huge data sets (HSC data)

- Both for Open-use and NAOJ's production + R&D

Who) HSC observers & General users

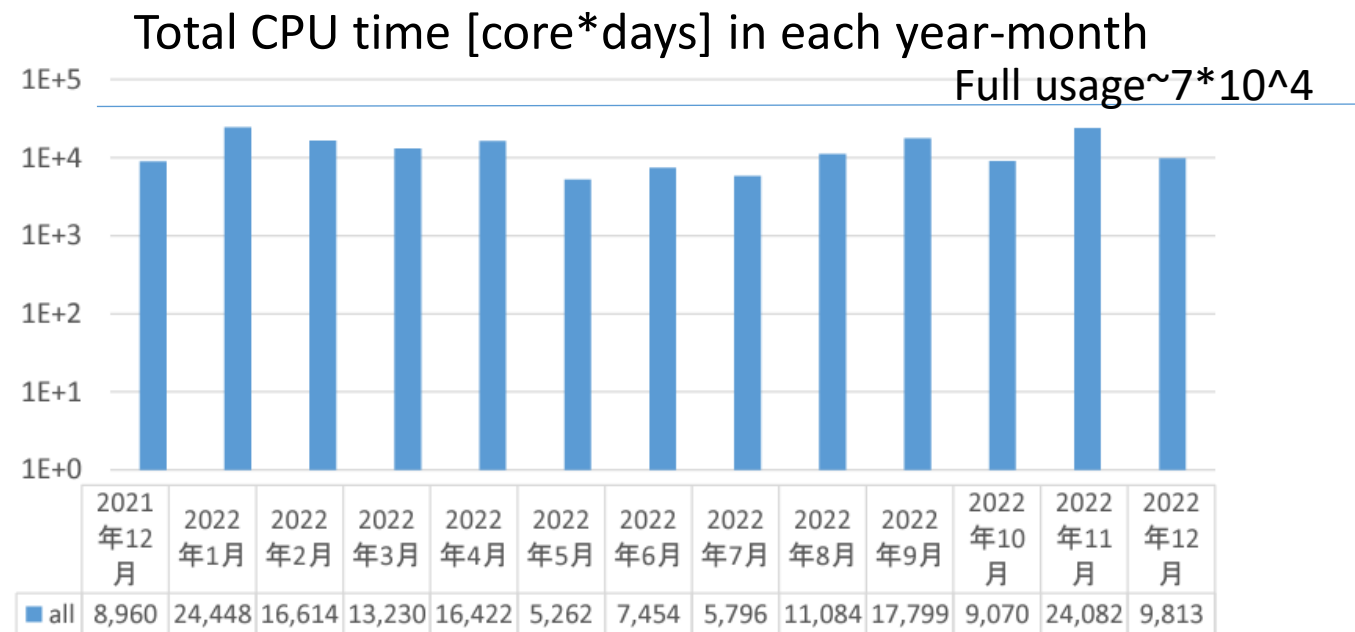
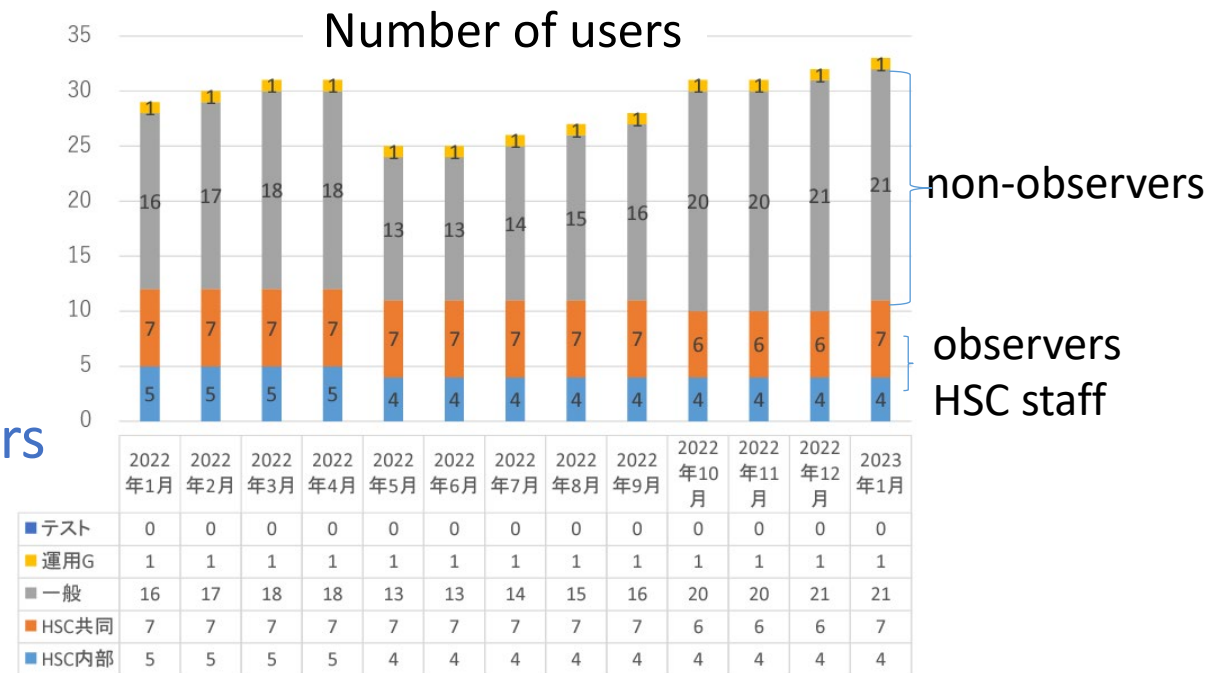
How) Prioritized access to resources by user status through batch job system (PBS)

- **Goup ID is available** to share data within your collaboration
- Approaching the end of support contract
 - We may have to prioritize some components to reduce running cost – your comment is welcome



LSC System Usage

- User
 - > 30 users
 - 6-7 open-use observers & 20 general users
- CPU occupation
 - ~20% CPU time on average
 - 60-70% in peak
 - going to have another intensive processing run soon
- Storage usage
 - ~4.2PB / 4.5PB (94%)
 - Has 4 generations of HSC-SSP products (s18a..21a)



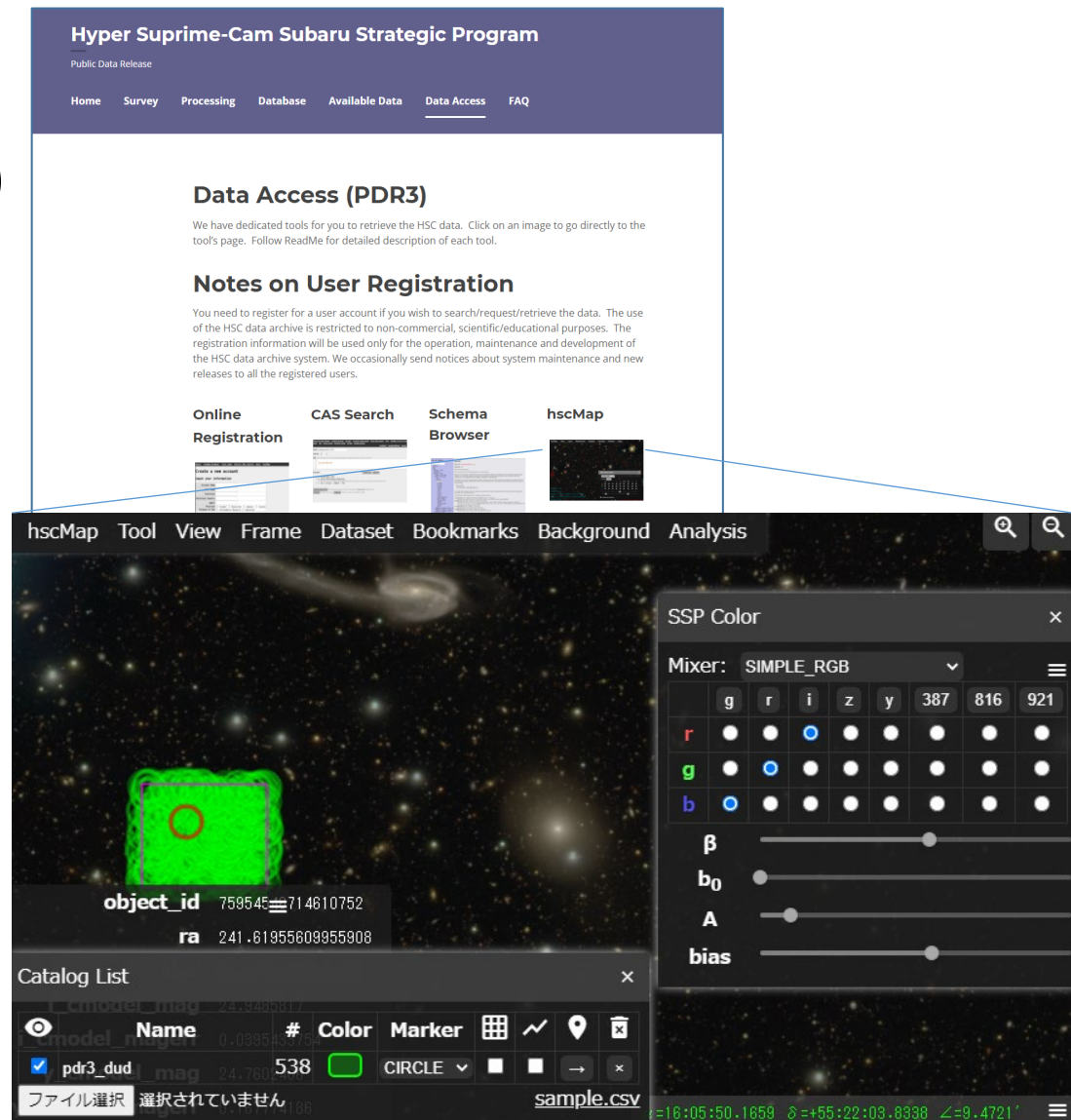
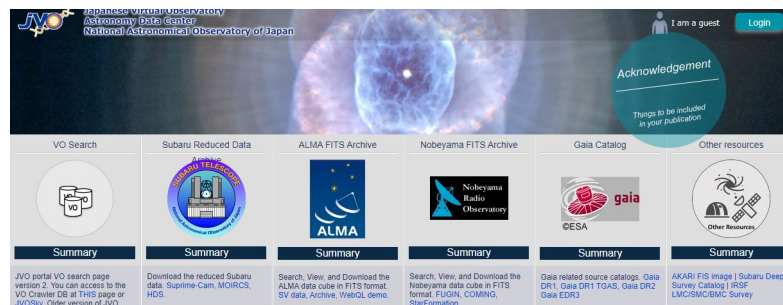
Feedback at ADC Users Meeting (MDAS-LSC Connection)

#1 May.19.2021, #2 June.24.2021

- Various versions of HSC data analysis pipeline (hscPipe) - addressed
- Availability for more general use - addressed
- Better file sharing and file transfer - addressed
 - NFS mounts from lsc (login node) to MDAS-NFS - addressed
 - For more efficient interactive node – continued
 - Transporting files to users' sites – continued

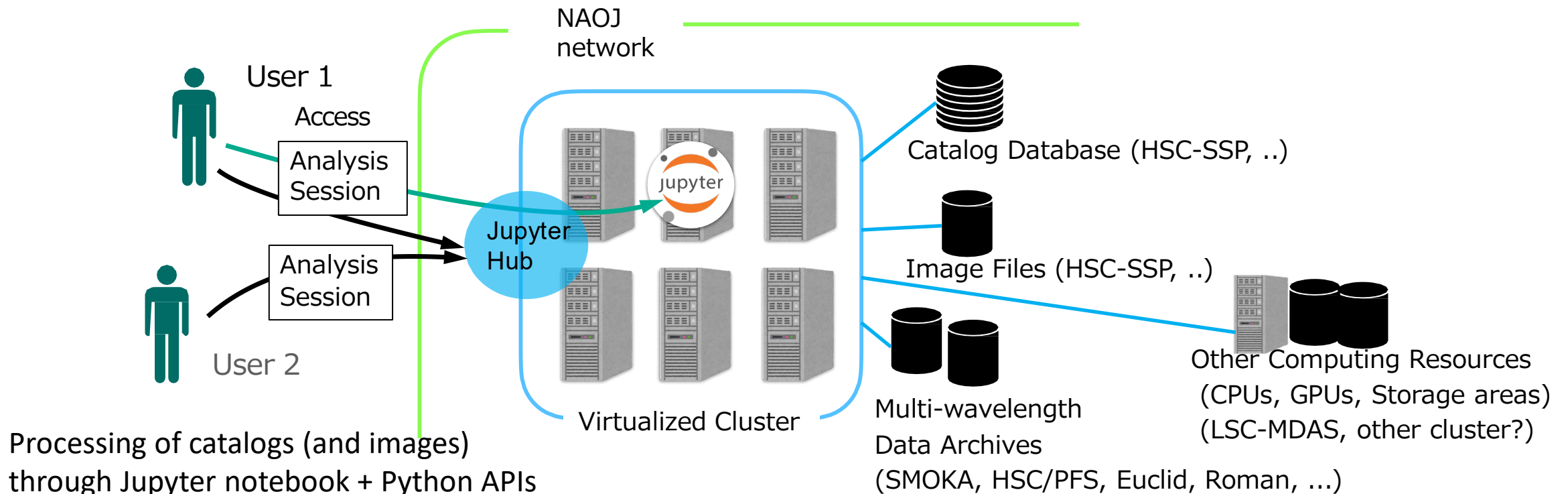
3. Science Products

- HSC-SSP Data Release (Subaru+ADC)
 - Preparing next data release (~2023 autumn)
 - Including all SSP data for the entire period
 - Testing Gen3 pipeline
 - ADC preparing for long-term operation including public data service (poster by Harasawa+)
- JVO portal
 - ALMA
 - Gaia
 - etc.



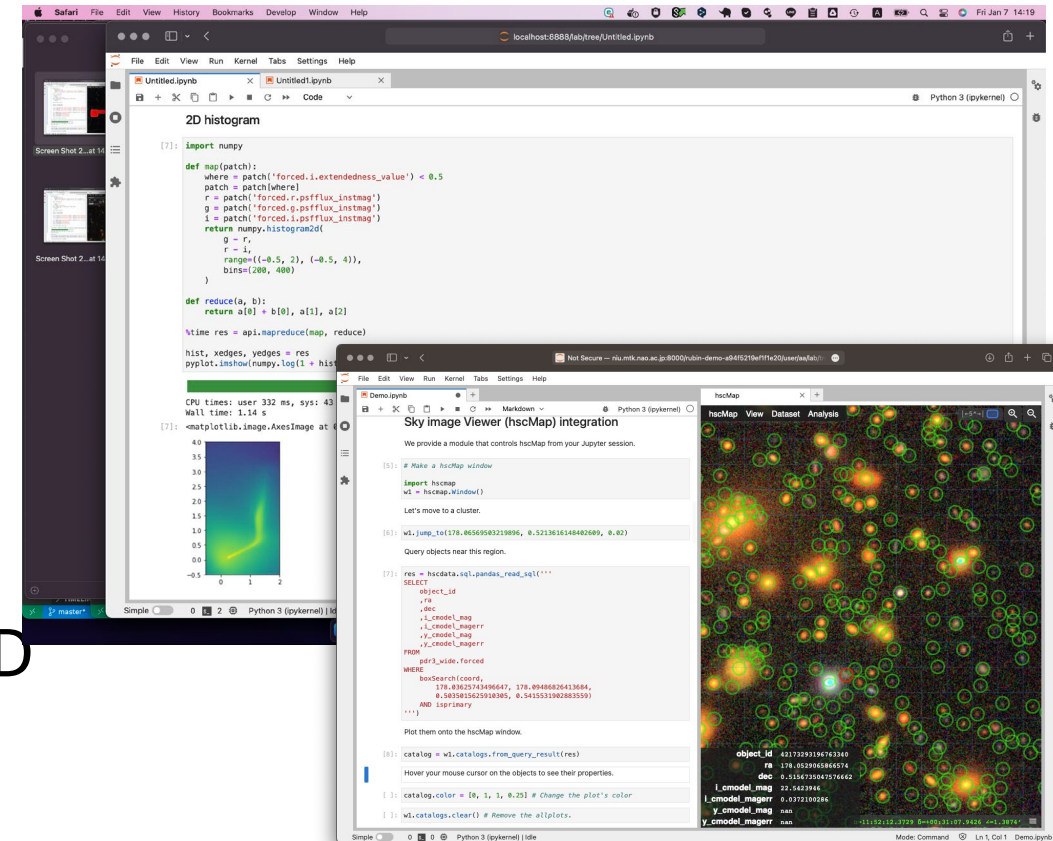
Development of Science Platform

- Designing a JupyterHub-based platform
 - to perform **efficient analysis over the existing products** from **remote**
 - to make efficient use of **computing resources**
- PFS involves the platform within a science archive framework



Prototype of HSC Science Platform

- ADC+Subaru is discussing to construct a science platform service connecting to the HSC data release PDR3
(See a poster by Morishima et al., too)
- ADC Team Tentative target plan
 - First prototype -2023.4
 - Internal review 2023.5
 - Preview by SSP collaboration 2023.9-
 - Design for PDR (and public data)
 - Preview for PDR 2024.4?
 - Initial operation for PDR in 2-3 years? TBD



Extending to participation in Rubin collaboration

- Rubin Independent Data Archive Center (Lite-IDAC) for LSST catalog products
- NAOJ (ADC+Subaru) preparing to request ~13 LSST PI for Japan (Day1PM, Day 3AM)
- Hope to contribute in IDAC & RSP with quickdb/hscMap. to utilize existing and future survey data sets (HSC+Rubin+Euclid+ etc) & multi-wave data service

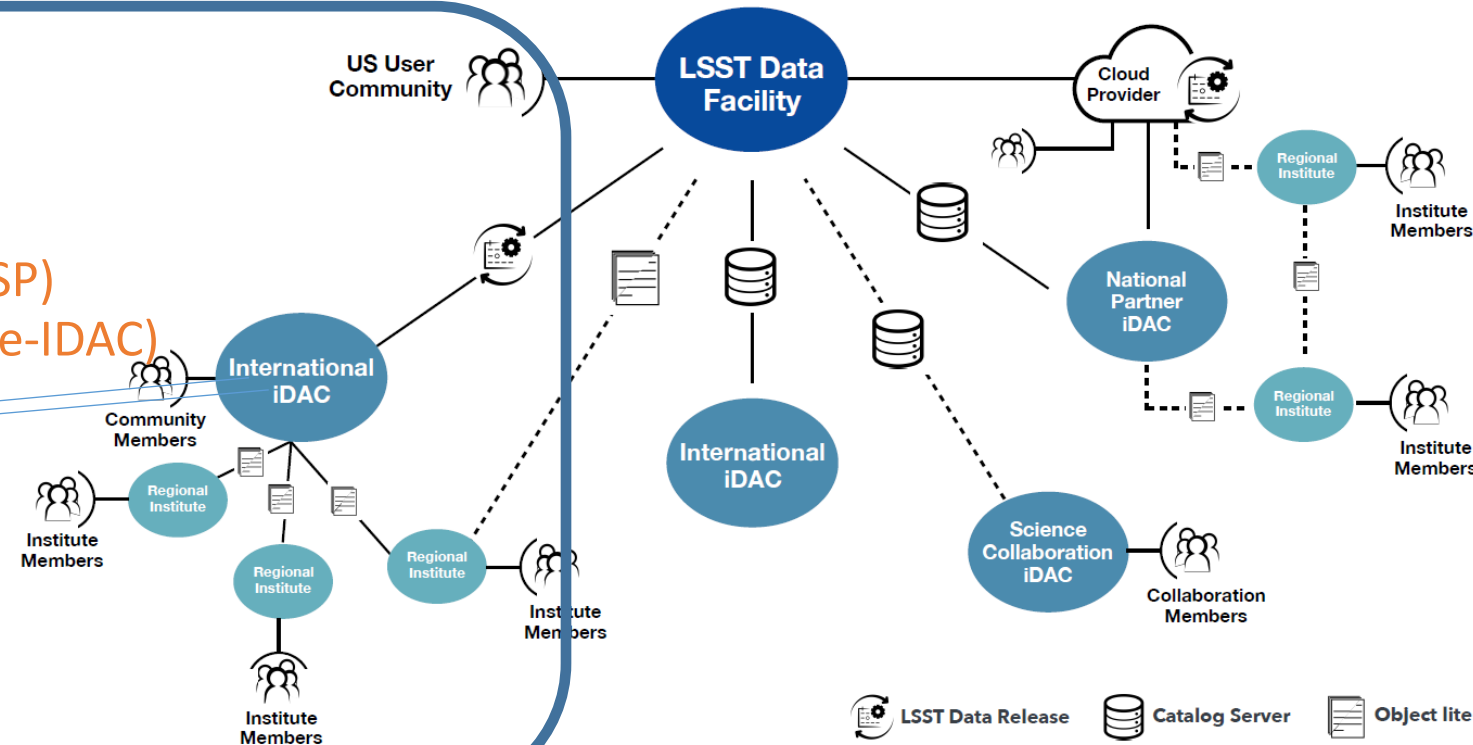
A proposal of Japanese Regional IDAC

Kavli-IPMU



Science Platform(RSP)
Catalog Archive (Lite-IDAC)

NAOJ-ADC



(Rubin RTN-003.pdf Figure 2.)

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/>
- Preparing computer update with rental contract renewal
- Discussing **new-generation raw data archive**
- Continuing **HSC data production and data service**, being transitioned to long-term operation at ADC
- Developing **science platform for HSC and Rubin**, aiming to **utilize existing and future data and computing resources**