FY2024 – Science User Support for Subaru Telescope Observing Data at Astronomy Data Center (ADC)

Hisanori Furusawa
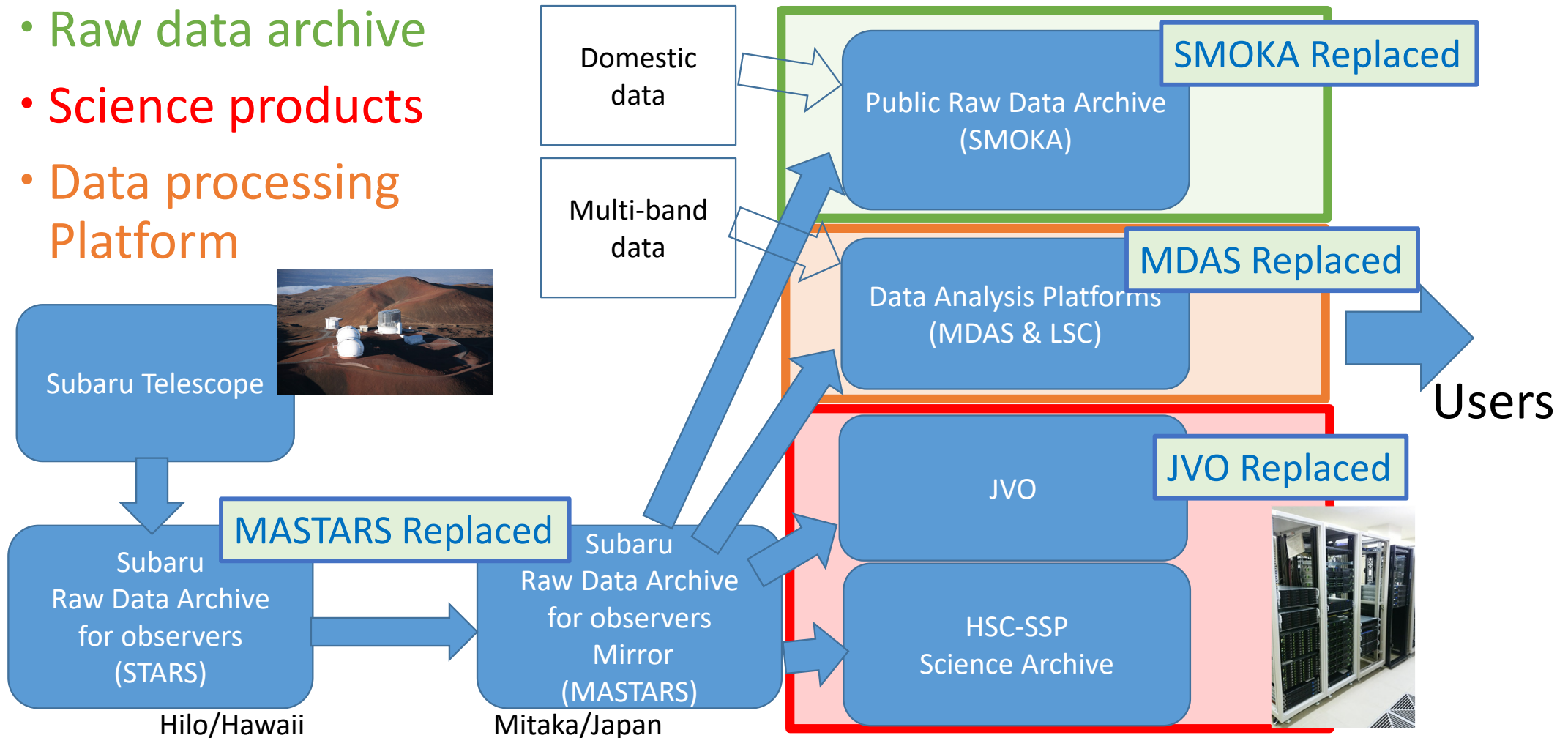
for ADC/Subaru open-use support team
1/28/2025 Subaru Users Meeting FY2024

ADC : Data Flow and Services for Subaru Data Sciences

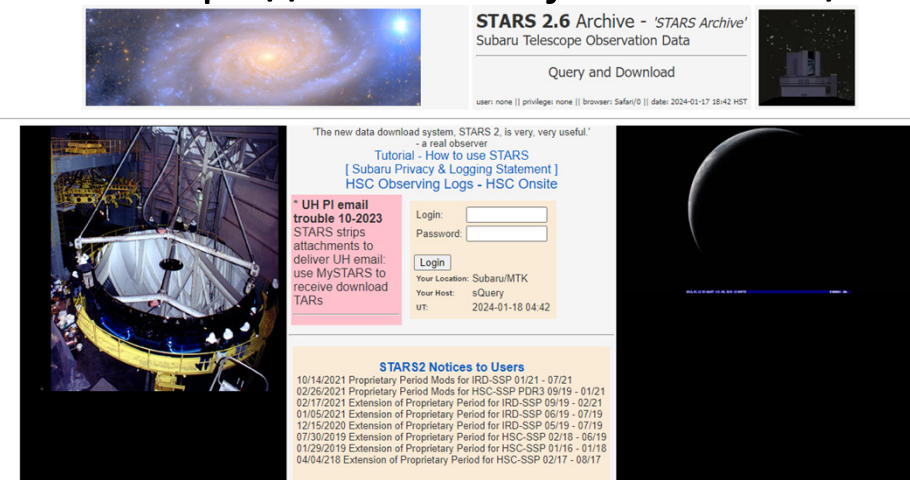Renewal of rental computer contract (Replacement) in ADC systems (2024.7)

- Raw data archive
- Science products
- Data processing Platform

Subaru Telescope

Subaru Raw Data Archive for observers (STARS)

Hilo/Hawaii

MASTARS Replaced

Subaru Raw Data Archive for observers Mirror (MASTARS)

Mitaka/Japan

Domestic data

Multi-band data

Public Raw Data Archive (SMOKA)

SMOKA Replaced

Data Analysis Platforms (MDAS & LSC)

MDAS Replaced

JVO

JVO Replaced

HSC-SSP Science Archive

Users

STARS/MASTARS

- STARS (Hilo; Tom Winegar et al.)
    - STN5->6: DB reconfigure. Efficient handling units of HSC exposures, etc. – done
    - PFS multi-propid handling - ongoing

- MASTARS (Mitaka; system being operated by ADC)
    - Migration to a new server completed. In normal operation.
    - System was greatly reduced, due to severe budget cut.
    - Planning for data compression.

- Ongoing / future work
    - The next-generation archive system across STARS-SMOKA – just kicked off

https://stars2.naoj.hawaii.edu/

SMOKA (public archive component) rental computer system replacement

HDD+LTO volume increases but (reliable) HDD volume decreased.

	before	after
Rental system Supported by Fujitsu		
HDD	**2.7 PB**	**1 PB**
LTO tape	**1.5 PB**	**3 PB** (+ 10 PB, just arrived)

LTO as a warp drive (Tomo-e, HSC)

non-rental, not Supported		
HDD		**+ 1 PB** (+ 1 PB, by March)

Major fraction of system into in-house

https://smoka.nao.ac.jp/

SMOKA Archive Advanced Search

Click here for SUP Search (Suprime-Cam data Search).
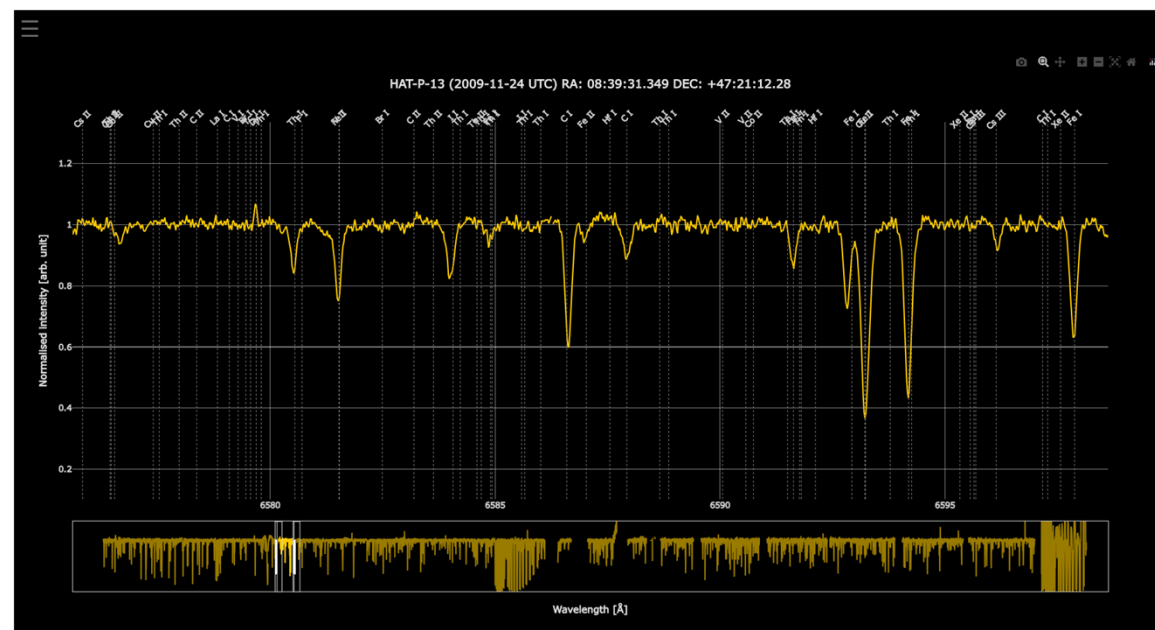*Click here* to know how to search.

Search Conditions

Object Name (for name resolve)	Resolver
Object Name	● SIMBAD ○ NED Resolve
	○ Don't Resolve

Coordinate System
Equatorial ▾   ● Circular

center (RA)   copy/paste to   Radius(arcmin)
Center Coordinate   From To   10.0

Equinox
J2000 ▾

center (DEC) copy/paste to
Center Coordinate   From To

Field of View
(arcmin)   ○ Rectangle
auto ▾

From (RA)   To (RA)
Corner Coordinate   ->   Opposite Corner Coordinate
From (DEC)   To (DEC)
Corner Coordinate   Opposite Corner Coordinate

Observation Date	Exp Time (sec)	Observer
Observation Date	Exp Time	Observer

Please see SMOKA report later

HDS Spectrum on JVO FITS WebQL

- HDS data were reduced using HDS reduction pipeline developed by the JVO team.

- The number of processed data is 41,968.

- They are available at http://jvo.nao.ac.jp/portal/subaru/hds.do for Quick Look purpose.

- The data search interface has also been updated.

- An interactive spectrum viewer for HDS has been developed and is available by clicking the "WebQL v5" button on the search result page.
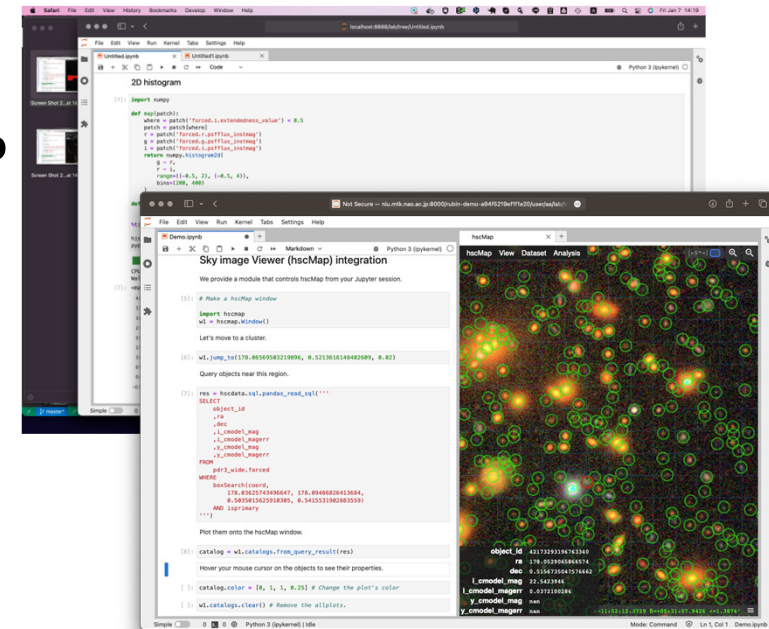
HSC SSP Data Release (all in-house)

- SSP internal data release S23B (for the entire SSP dataset) (2024.7)
  - Reprocessing catalog generation (improving deblend) underway
- Prototyping SP on the HSC data release PDR3 as extension func.
- ADC Team Tentative target plan
  - Internal review by a few experts - continued
  - Preview by SSP collaboration by 2025 Summer?
  - Design for PDR (and public data) this year
- PFS SP will start open-use operation soon

MDAS (for interactive tasks)

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

System updated in July 2024

User consoles

Before

After update

VPN

VPN

Switch

In-house maintenance

20 x16core
12 x16core
(512core Xeon)

Rental components

Switch

Switch

8 x 64core
(512core EPYC)

1 x16core

2 x16core

Interactive servers

Batch job servers

update

Interactive server

Interactive servers

Rocky 8

RHEL7

Lustre

NFS

1.7PB

1.2PB

NFS

1.2PB

NFS

4PB

LSC (different system)
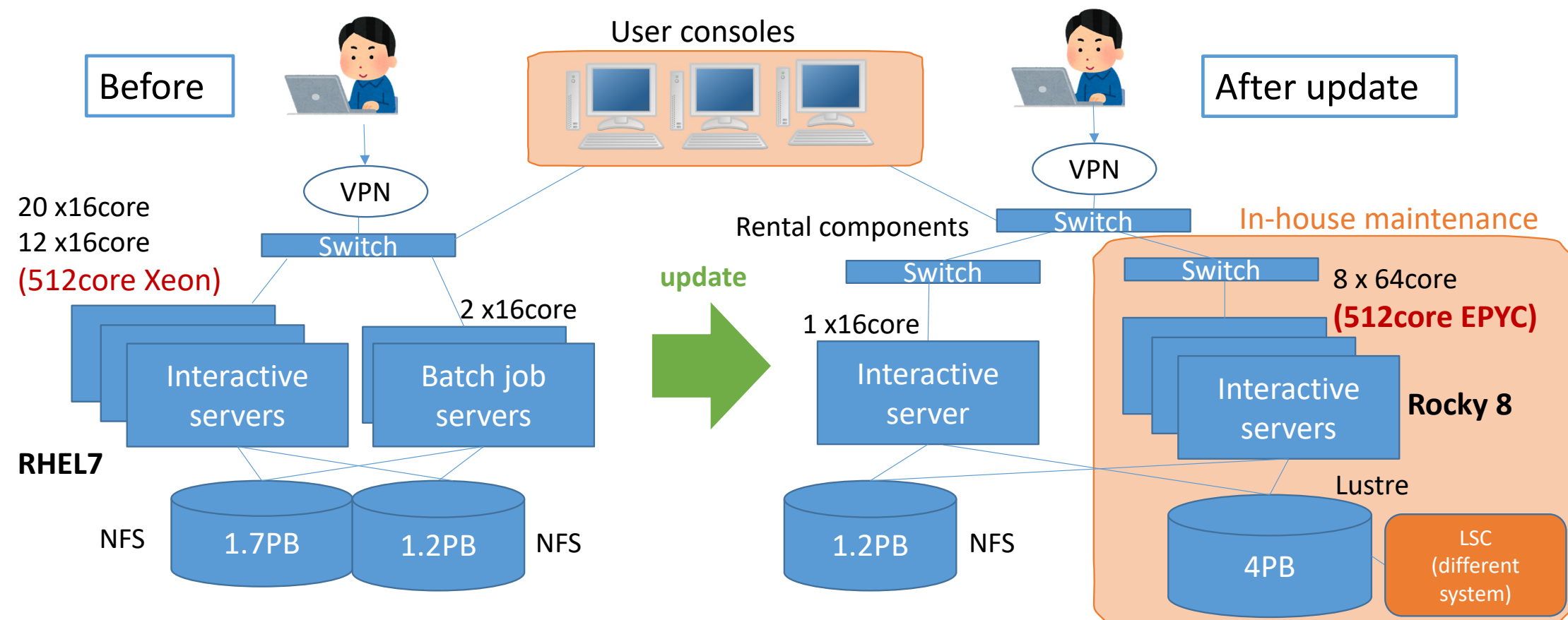
- Maintains the **same scale of CPU and memory size** with **larger storage area on Lustre**
- **No batch job system**

MDAS

- Status
  - The new system is maintained through a hybrid of rental and in-house management.
  - The system was launched in **July 2024** and has achieved stable operation.
  - Subaru and ALMA users make up the majority.

328 effective users

Number of users per observation Data used

Subaru 83, Okayama 7, ALMA 109, ASTE 9, NRO45 22, VERA 11, Solar 26, Archive 26, Others 24

Number of users per field of use (total)

Opt/IR 115, Radio 135, Solar 21, Others 15
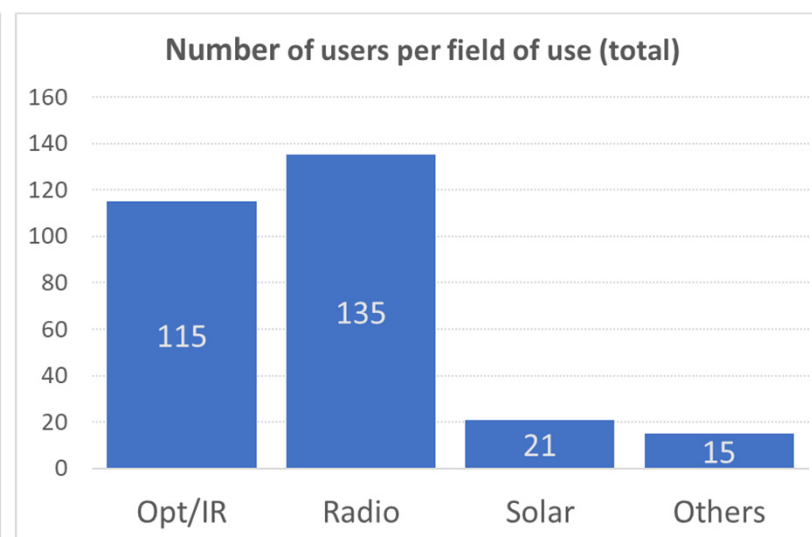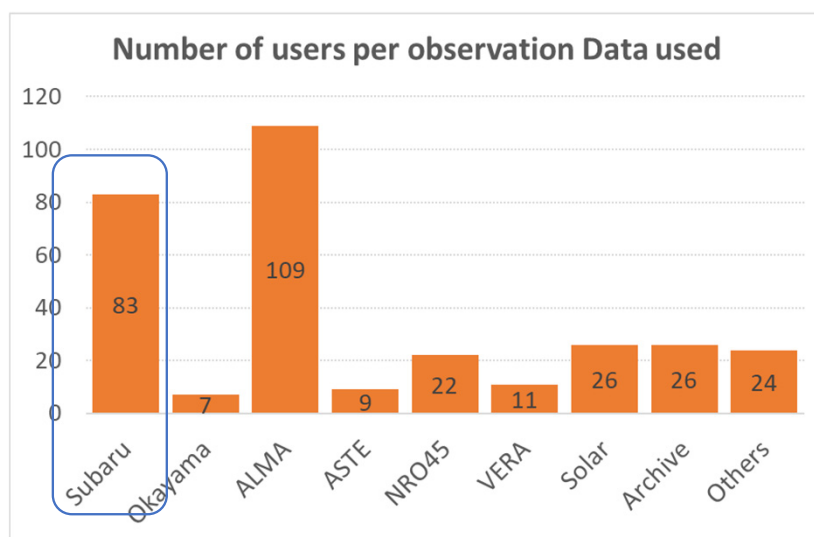
LSC (large scale processing)

- Continue to be cooperated by ADC-Subaru
- Discussion for maintaining the LSC function
  - GPFS 5PB in the 6th year, support contract expired
  - Facing severe shortage of GPFS capacity, with cooperation by users
  - Preparing external disks (Lustre), 2PB (last year)
  - Maintaining existing computing nodes with OS being updated
  - Necessity of Introducing additional resource limit, e.g., disk quota per user
  - Limiting the number of jobs that can be executed simultaneously
- Adjustment of resource usages & preparing for PFS operation
  - Separation of computing nodes between HSC-SSP (~1500cores)
  and「general users + observers」-- continued
  - Will assign a fraction of resources for PSF
  - Revisit the resource limit issue in the ADC-UM (Feb 6)

  Your cooperation & comments appreciated

2296 core / 40 nodes on 5PB GPFS

LSC System Usage

※See the uploaded slides later

- User (increased)
  - 40 users
  - 9 current observers & 26 general users
- CPU occupation (same as FY23)
  - PBS jobs only (without HTCondor) included in this calculation
  - ~ 10-20% CPU time on average
  - ~ 25% in peak
  - Partly depends on HSC-SSP tasks
- Storage usage
  - 4.5PB / 4.5PB (~100%)
  - Removed old HSC-SSP products
  - New SSP processing underway
  - Cooperation of users needed

Number of users

general users

observers

HSC staff

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

Full usage~7*10^4

	2023年12月	2024年1月	2024年2月	2024年3月	2024年4月	2024年5月	2024年6月	2024年7月	2024年8月	2024年9月	2024年10月	2024年11月	2024年12月
all	6,623	14,102	18,350	4,105	15,837	12,987	18,549	11,005	7,167	7,246	10,670	12,547	18,274
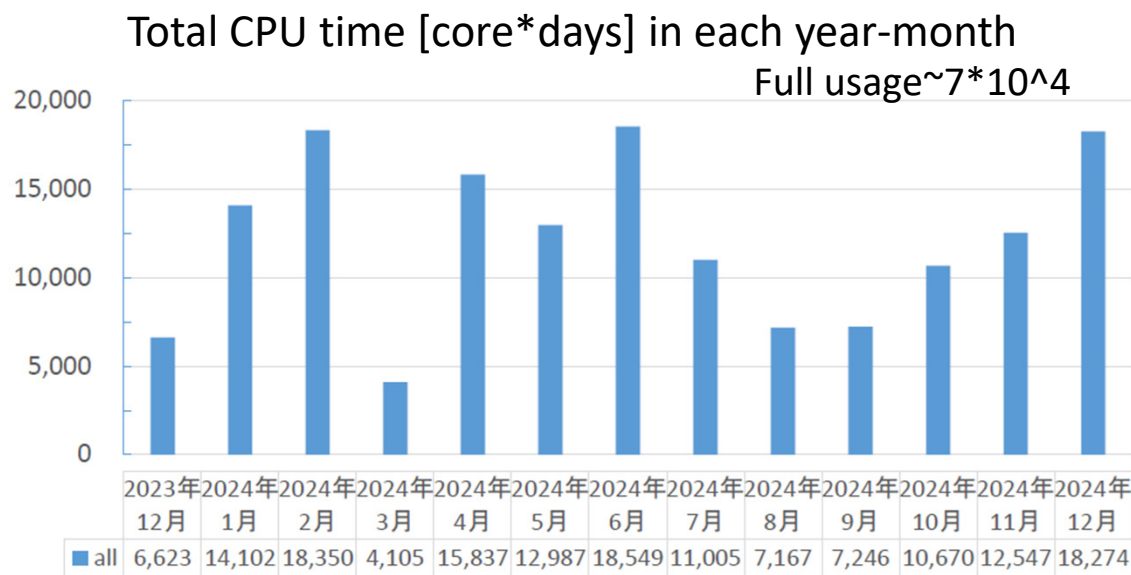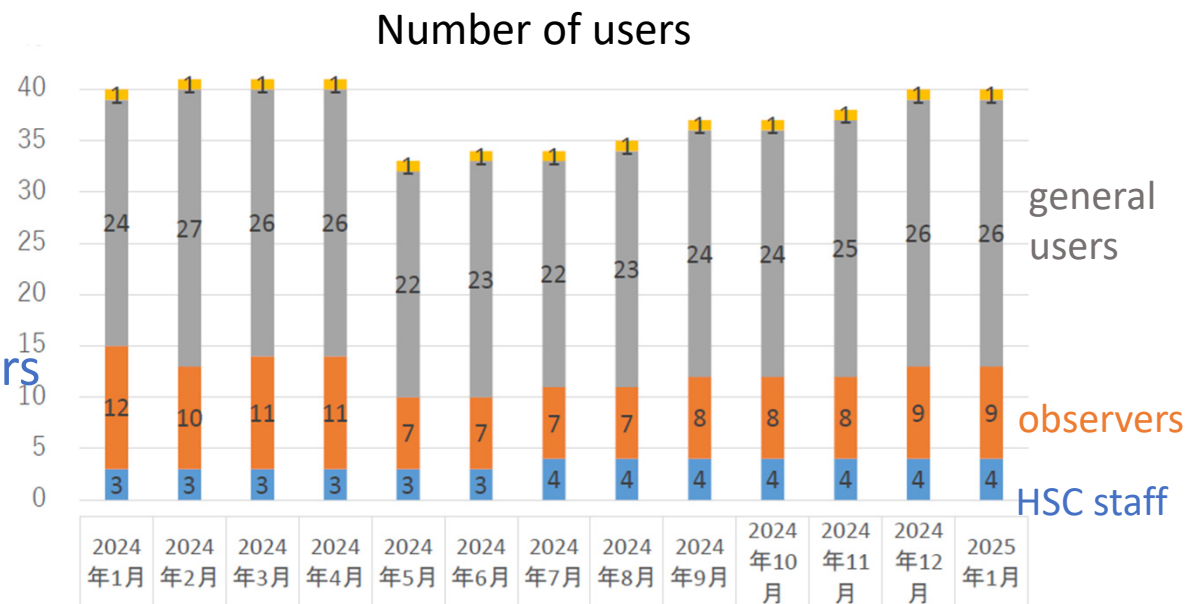
Data compression guideline

- Lossless compression is now formally allowed by the observatory
  - Subaru: <u>Archival Data Management Committee</u> will take care of the operation
- Will be recommended for large data sets (in archive and future instruments)
- Application of compression
  - HSC will be the first target for relaxing the issue for archival storage shortage for the expanding data rate
  - Fpack RICE compression is assumed
    - CFITSIO and Astropy can decode it without special action
    - Checksum before & after the compression to be recorded to ensure the info is not modified
  - Timeline is TBD but before long
- On both STARS/MASTARS and then on SMOKA
  - After the transition, users will get compressed files by default

Concept of Reallocating Subaru-bldg. Open-use Room
(on behalf of Subaru Office: Koyama Y.)

- Subaru building runs <u>short of staff room</u> space

- Subaru is investigating to re-arrange the open-use room in the Subaru building to bring the visitor room (currently 2nd floor) down into it

- Tentative idea
  - Keep ~1/3..1/2 for openuse RM
  - Assign ~1/2..2/3 for the Subaru visitor RM
  - Free partitions to divide the space
    -> could use the whole room for
       Subaru or ADC events (Koshu-kai etc)
  - Timeline TBD,
    but they hope to make it early next FY?

- We would like to hear comments
  from users (in ADC-UM on Feb 6, too)

Data Archive RM

Insei-Seminar RM

Keep OpenUse RM

Into Subaru Visitor RM
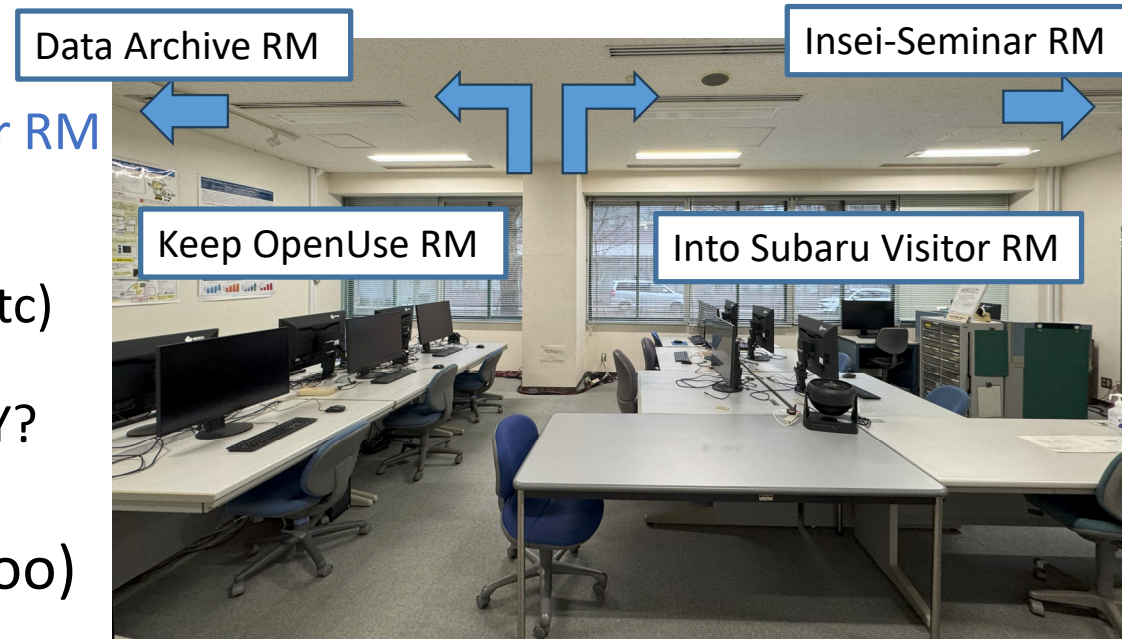
SMOKA
Current Status and Future Plans

Hisakazu Uchiyama
Yasushi Nakajima, Shuhei Koyama, Takeaki Ozawa, Junko Furusawa,
Hisanori Furusawa, Tadafumi Takata
(NAOJ Astronomy Data Center)

Subaru Users Meeting FY2024

What is SMOKA?

- **Subaru - Mitaka - Okayama - Kiso Archive System**
  - The astronomical data archive system for Japanese optical and infrared telescopes.
  - While STARS/MASTARS provides data access for Subaru observers,
    SMOKA serves as a public archive for all users after the proprietary period.
  - Mainly provides raw data for the purpose of creating new research results,
    verifying research findings, and promoting research and educational activities.

## Main data released in SMOKA

Subaru	S-Cam, FOCAS, HDS, IRCS, OHS/CISCO, MIRTOS, COMICS, CIAO, CAC, Kyoto-3DII, MOIRCS, HiCIAO, FMOS, HSC, CHARIS, IRD, SWIMS, MIMIZUKU, VAMPIRES
Okayama	SNG, OASIS, HIDES, ISLE, KOOLS, MuSCAT
Kiso	1kCCD, 2kCCD, KWFC
MiTSuME	MTA (AKENO), MTO (OKAYAMA)
Kanata	HOWPol, HONIR
Seimei	KOOLS-IFU, TriCCS
Nayuta	NIC

Note: Data released in a different system configuration

Tomo-e Gozen stacked data, Digitized photographic plate data, All-sky monitor images

The Trend of SMOKA Released Data Volume

Note: Excludes Tomo-e Gozen stacked data,
digitized photographic plate data,
and all-sky monitor image data.

- Total data
  - 41 million frames
  - 556 TB

- Subaru data
  - 20 million frames
  - 346 TB
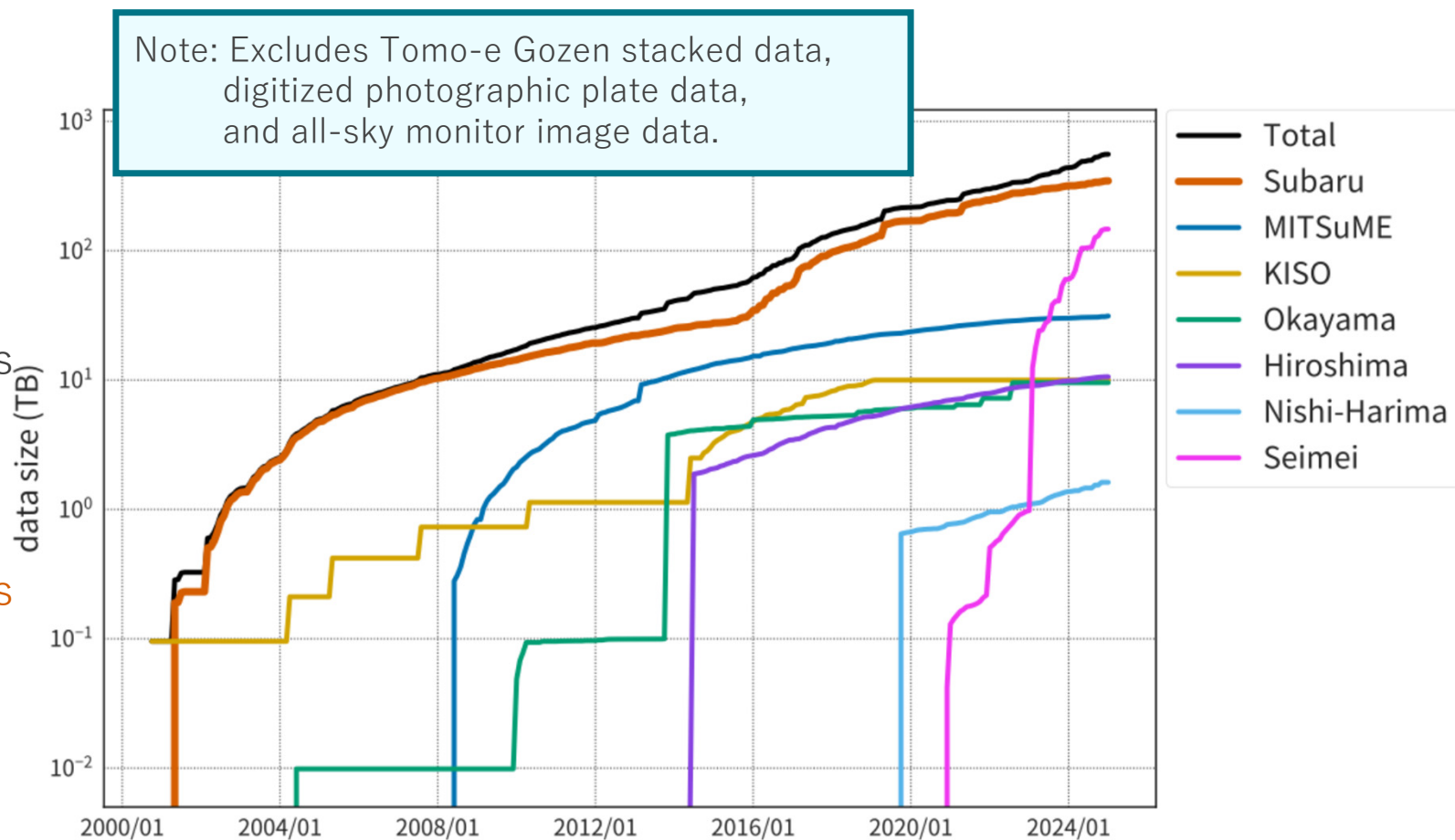
SMOKA Usage Status

- Annual registered users
  - ~250 (reset every year)

- Monthly data request volume
  - Number of frames:
    0.1 - 1 million
  - Data size: 1-10 TB

Trend of requested data size and number of frames from 2002 to Jan. 2025
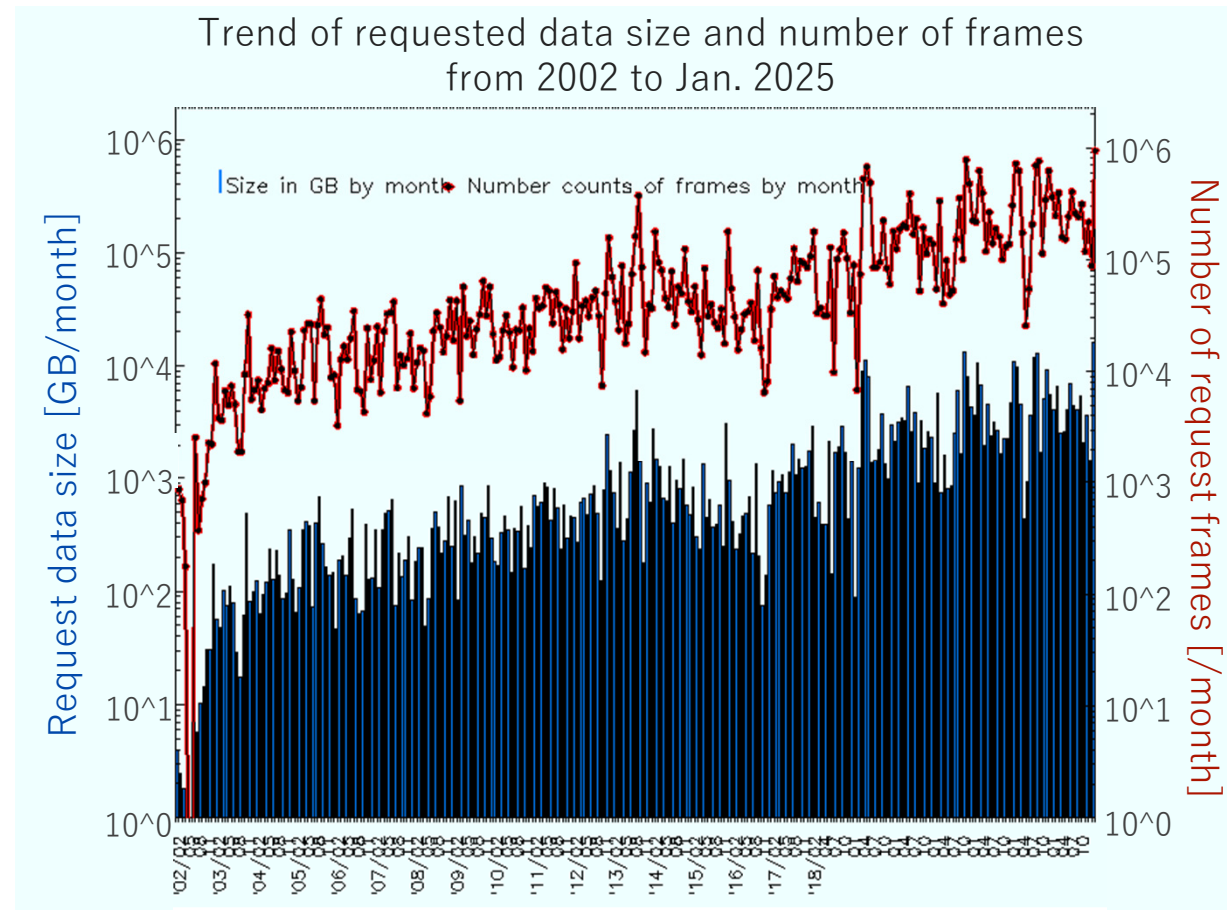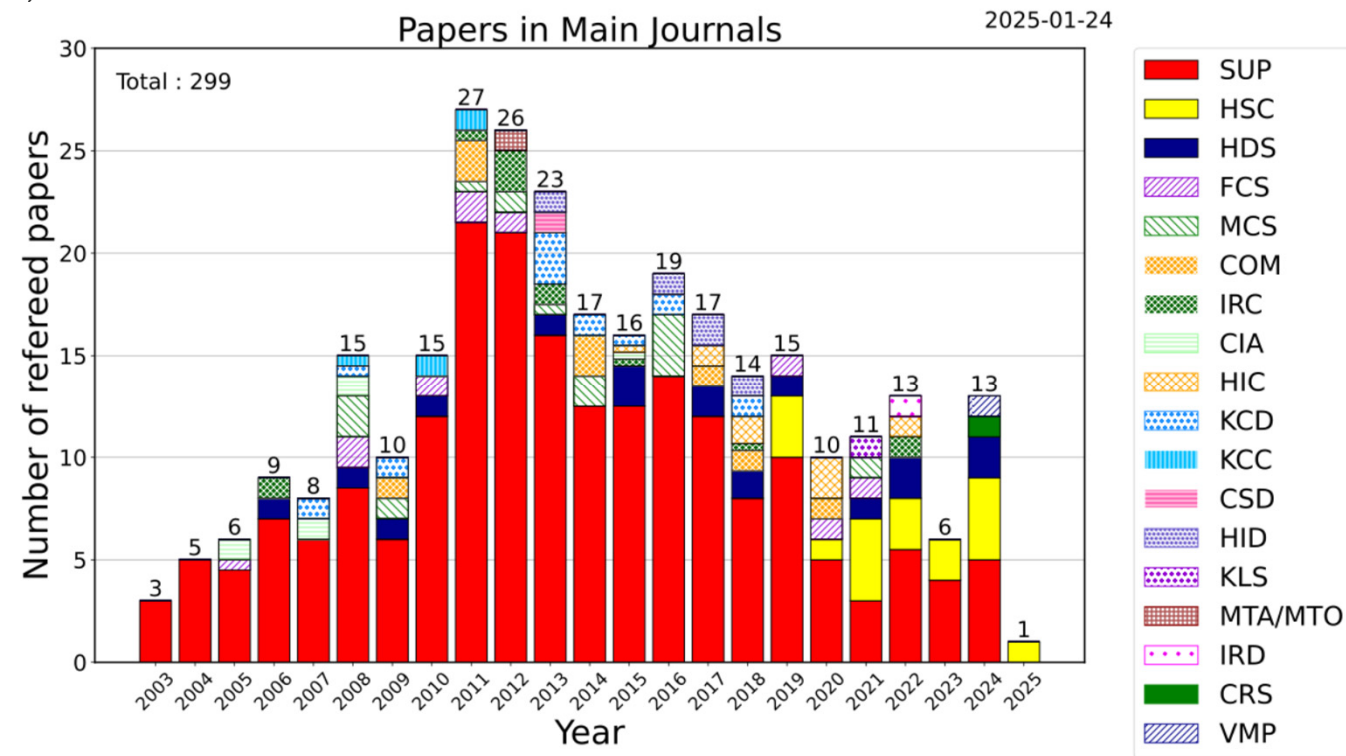
Number of Papers Using SMOKA

- Toal: 299 (as of Jan. 24, 2025)
  - ~14 per year

- Journals
  - ApJ, ApJL, ApJS, A&A, PASJ, PASP, NRAS, ICARUS, Nature, Science, …

Note:

Excludes DATA AVAILABILITY

Current Status and Future Plans

- Released new data
  - Subaru
    - ‣ All-Sky monitor image (from Apr. 18, 2024)

- Computer system replacement (July 2024)
  - Replaced computer systems responsible for data transfer, data processing, and storage.
  - All SMOKA services, which were suspended from June 25 to July 9, 2024, have now been fully restored.
    (DRM-HSC service was suspended until Dec. 10, 2024.)

- Future Data Release
  - Subaru
    - ‣ Fast PDI, MEC, PFS
    - ‣ CNN-based Clear/Cloudy Classification for All-sky Monitor Images
  - Seimei
    - ‣ GAOES-RV, NIRpol
  - Tomo-e Gozen raw data
  - TAO
    - ‣ NICE, MIMIZUKU, SWIMS, …

Summary

- What is SMOKA?

  - The astronomical observation data archive system providing public access to data from Japanese optical and infrared telescopes.

  - While STARS/MASTARS provides proprietary data access for Subaru observers, SMOKA serves as a public archive for all users after the proprietary period.

  - Mainly providing raw data for the purpose of creating new research results, verifying research findings, and promoting research and educational activities.

- 41 million frames (556 TB) are archived in SMOKA.

  - Including 20 million frames (346 TB) of Subaru data.

- 299 peer-reviewed papers using SMOKA have been published (~14 /year).

- SMOKA plans to continue new data release.

End of Slides