FY2025 – Astronomy Data Center Updates from Open-use data services for Subaru Telescope

Hisanori Furusawa (ADC)

10/29/2025 Subaru Users Meeting FY2025

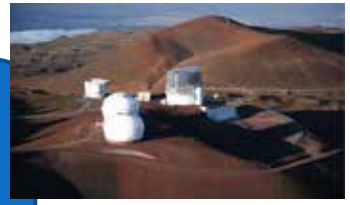
ADC : Data Flow and Services for Subaru Data Sciences

- New systems operating from 2024.7 à Stabilizing the services
- ADC Users Meeting will be held on November 18.

u Raw data archive

u Science products

u Data processing Platform



Subaru Telescope

Domestic data

Multi-band data

Public Raw Data Archive (SMOKA)

SMOKA updated

Data Analysis Platforms (MDAS & LSC)

MDAS updated

Users

MASTARS updated

Subaru Raw Data Archive for observers (STARS)

Hilo/Hawaii

Subaru Raw Data Archive for observers Mirror (MASTARS)

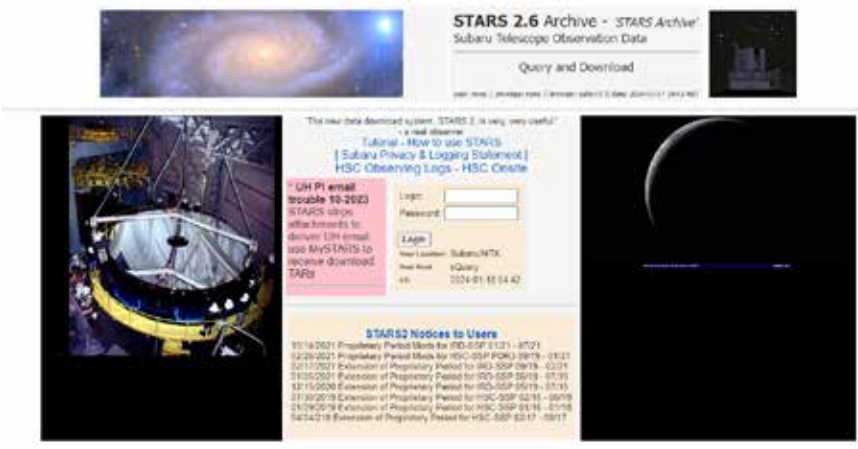
Mitaka/Japan

JVO

JVO updated (SCam, HDS data etc)

HSC-SSP Science Archive

STARS/MASTARS



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

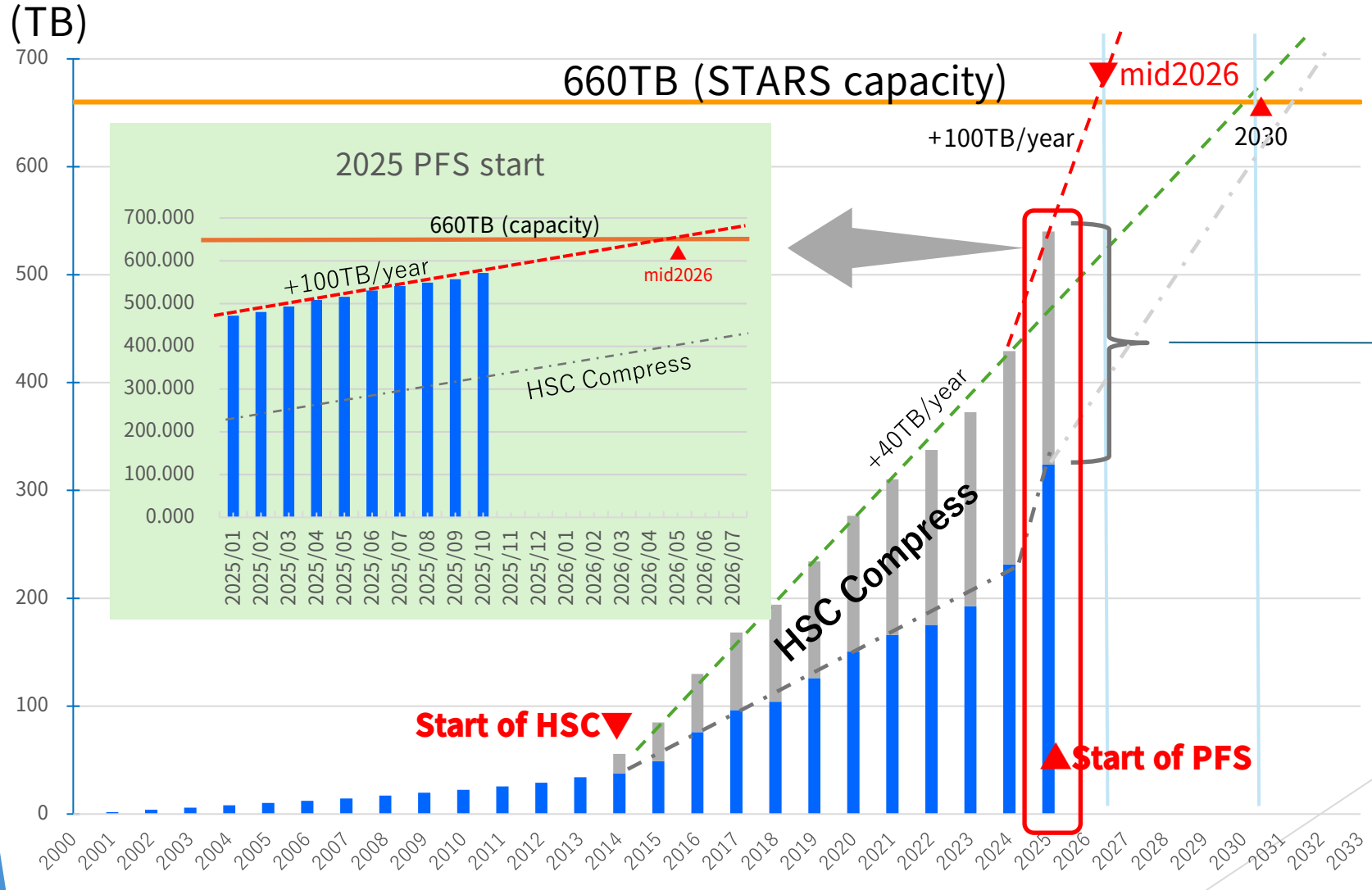
- STARS(Hilo; search UI and download)
  - PFS queue data release started.
  - Data rate increased after PFS starts (next slide)
  - Storage added +110TB to mitigate the increased data rate.
- MASTARS(Mitaka; for download only)
  - Recently, download from MASTARS was roughly twice of that from STARS.
  - Storage added ~1PB (it includes the development system)
  - Data backed up to LTO8.
  - System trouble (hardware failure) in mid August.

    MASTARS stopped for ~3weeks. We are sorry for inconvenience.

    (STARS+MASTARS redundancy worked well for such a hardware failure)
- Ongoing / future work
  - Data compression (fpacking) for making disk spaces (mainly for STARS).
  - The next-generation archive system across STARS-SMOKA.
  - Conceptual design ongoing.

Cumulative Data Volume in STARS

- Data rate (dominated by HSC) was 40TB/yr à is 100TB/yr w/ PFS
- Lossless data compression will generate ~200TB to mitigate the disk shortage



- We are now working on a lossless compression of old (>1.5yr) HSC data.
- Upon completion of the compression,approximately 200TB is expected to become available.
- The compressed FITS files are provided directly to the users.

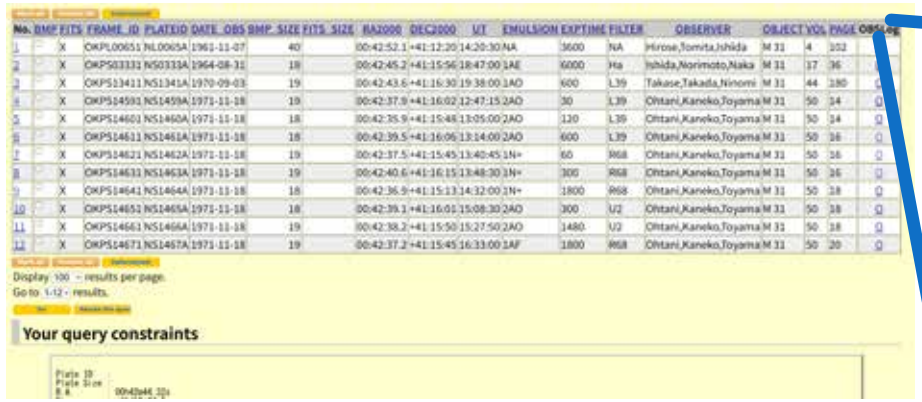
https://smoka.nao.ac.jp/

SMOKA – Public data archive (Subaru-Mitaka-Okayama-Kiso Archive System)

- Releasing public raw data from 36 instruments including 19 instruments of Subaru
- A total of 305 papers using SMOKA data have been published in major refereed journals. (6 papers come out from last meeting)
- Update since the last Subaru UM
  - Experimental feature: CNN-based Clear/Cloudy Classification for Subaru All-sky Monitor Images (March 2025 ~)
  - New release : GAOES-RV of Seimei telescope (March 2025 ~)
  - New release : Tomo-e Gozen Raw Data (partially) of Kiso Schmidt (March 2025 ~)
  - New release : Okayama 188cm/74inch Photographic Plate Digitized data (October 2025 ~)

See report by Koyama, S. after me

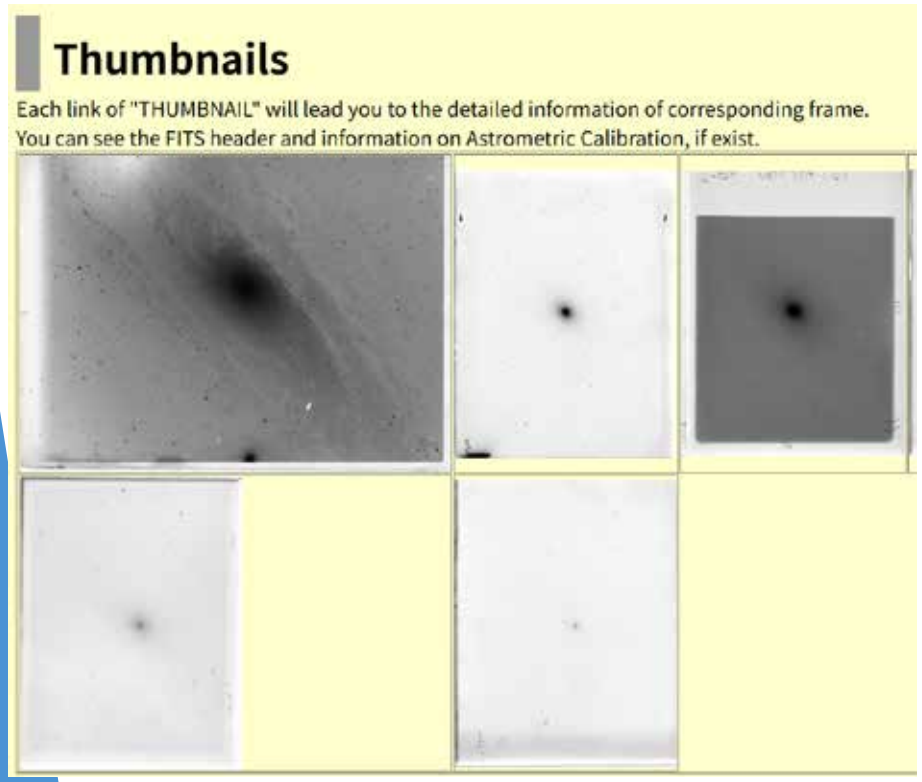
Okayama Photographic Plate Archive (Oct 2025~)



No.	DMP	FITS	FRAME_ID	PLATEID	DATE_OBS	DMP_SIZE	FITS_SIZE	RA2000	DEC2000	UT	EMULSION	EXPTIME	FILTER	OBSERVER	OBJECT	VOL	PAGE	OBSLog
1		X	OKPL00651	NL0065A	1961-11-07	40		00:42:52.1	+41:12:20	14:20:30	NA	3600	NA	Hirose,Tomita,Ishida	M 31	4	102	
2		X	OKPS03331	NS0333A	1964-08-31	18		00:42:45.2	+41:15:56	18:47:00	1AE	6000	Ha	Ishida,Norimoto,Naka	M 31	17	36	
3		X	OKPS13411	NS1341A	1970-09-03	19		00:42:43.6	+41:16:30	19:38:00	1AO	600	L39	Takase,Takada,Ninomi	M 31	44	180	
4		X	OKPS14591	NS1459A	1971-11-18	19		00:42:37.9	+41:16:02	12:47:15	2AO	30	L39	Ohtani,Kaneko,Toyama	M 31	50	14	
5		X	OKPS14601	NS1460A	1971-11-18	18		00:42:35.9	+41:15:48	13:05:00	2AO	120	L39	Ohtani,Kaneko,Toyama	M 31	50	14	
6		X	OKPS14611	NS1461A	1971-11-18	18		00:42:39.5	+41:16:06	13:14:00	2AO	600	L39	Ohtani,Kaneko,Toyama	M 31	50	16	
7		X	OKPS14621	NS1462A	1971-11-18	19		00:42:37.5	+41:15:45	13:40:45	1N+	60	RG8	Ohtani,Kaneko,Toyama	M 31	50	16	
8		X	OKPS14631	NS1463A	1971-11-18	19		00:42:40.6	+41:16:15	13:48:30	1N+	300	RG8	Ohtani,Kaneko,Toyama	M 31	50	16	
9		X	OKPS14641	NS1464A	1971-11-18	18		00:42:36.9	+41:15:13	14:32:00	1N+	1800	RG8	Ohtani,Kaneko,Toyama	M 31	50	18	
10		X	OKPS14651	NS1465A	1971-11-18	18		00:42:39.1	+41:16:01	15:08:30	2AO	300	U2	Ohtani,Kaneko,Toyama	M 31	50	18	
11		X	OKPS14661	NS1466A	1971-11-18	19		00:42:38.2	+41:15:50	15:27:50	2AO	1480	U2	Ohtani,Kaneko,Toyama	M 31	50	18	
12		X	OKPS14671	NS1467A	1971-11-18	19		00:42:37.2	+41:15:45	16:33:00	1AF	1800	RG8	Ohtani,Kaneko,Toyama	M 31	50	20	

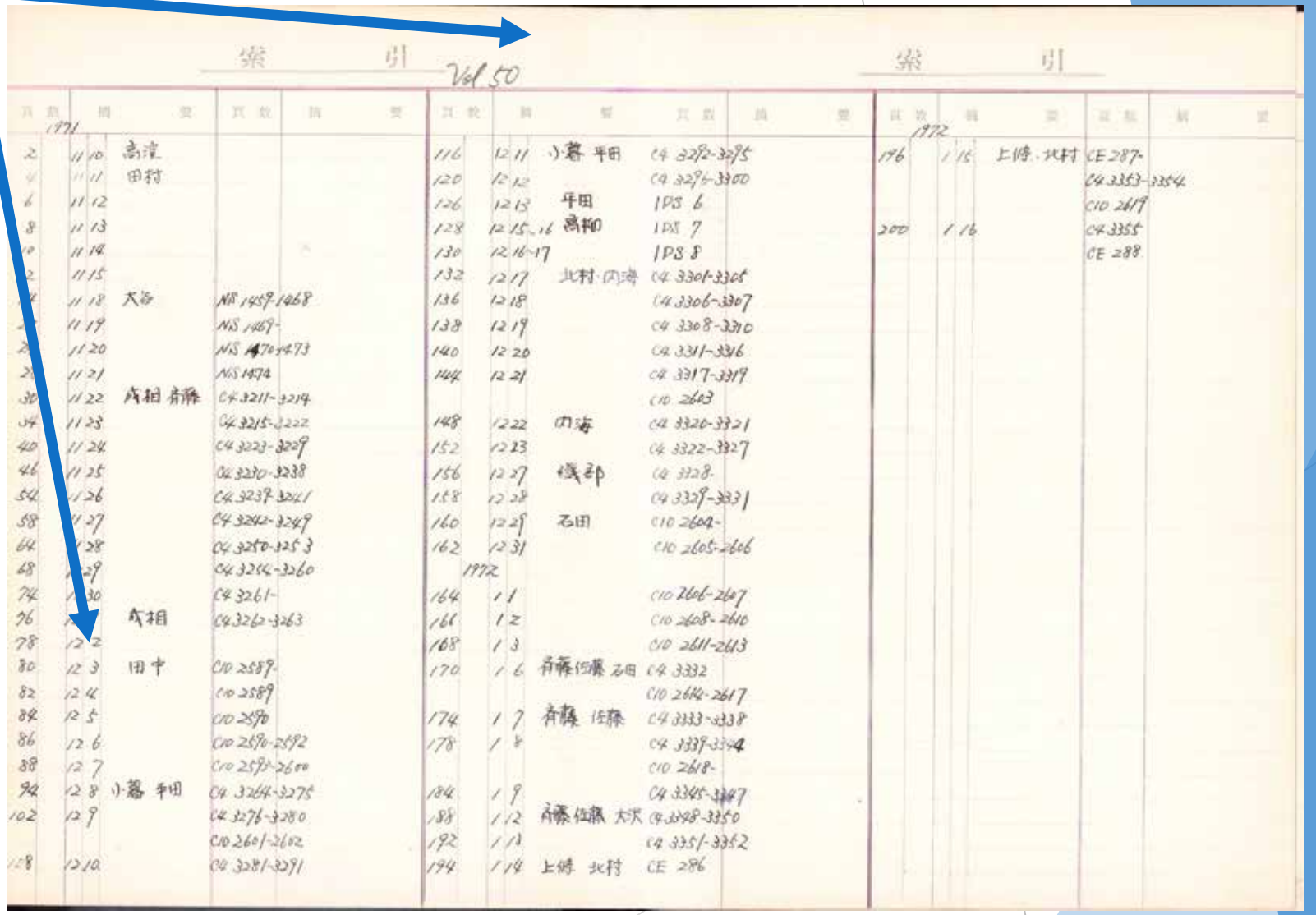
Display 100 ~ results per page.
Go to 1-12 ~ results.

Your query constraints



Thumbnails

Each link of "THUMBNAIL" will lead you to the detailed information of corresponding frame.
You can see the FITS header and information on Astrometric Calibration, if exist.



索引 Vol 50 索引

1971

頁数	月	日	摘要	頁数	摘要
2	11	10	高濱		
	11	11	田村		
6	11	12			
8	11	13			
10	11	14			
12	11	15			
14	11	18	大谷	NS 1459-1468	
16	11	19		NS 1469-	
18	11	20		NS 1470-1473	
20	11	21		NS 1474	
30	11	22	成相 斉藤	C4 3211-3214	
34	11	23		C4 3215-3222	
40	11	24		C4 3223-3229	
46	11	25		C4 3230-3238	
54	11	26		C4 3239-3241	
58	11	27		C4 3242-3249	
64	11	28		C4 3250-3253	
68	11	29		C4 3254-3260	
74	11	30		C4 3261-	
76	12	1	成相	C4 3262-3263	
78	12	2			
80	12	3	田中	C10 2587-	
82	12	4		C10 2589	
84	12	5		C10 2590	
86	12	6		C10 2591-2592	
88	12	7		C10 2593-2600	
94	12	8	小暮 平田	C4 3264-3275	
102	12	9		C4 3276-3280	
				C10 2601-2602	
128	12	10		C4 3281-3291	

頁数	月	日	摘要		頁数	摘要
116	12	11	小暮 平田	C4 3292-3295		
120	12	12		C4 3296-3300		
126	12	13	平田	IPS 6		
128	12	15-16	高柳	IPS 7		
130	12	16-17		IPS 8		
132	12	17	北村 田淵	C4 3301-3305		
136	12	18		C4 3306-3307		
138	12	19		C4 3308-3310		
140	12	20		C4 3311-3316		
144	12	21		C4 3317-3319		
				C10 2603		
148	12	22	田海	C4 3320-3321		
152	12	23		C4 3322-3327		
156	12	27	成都	C4 3328-		
158	12	28		C4 3329-3331		
160	12	29	石田	C10 2604-		
162	12	31		C10 2605-2606		
1972						
164	1	1		C10 2606-2607		
166	1	2		C10 2608-2610		
168	1	3		C10 2611-2613		
170	1	6	斉藤佐藤 石田	C4 3332		
				C10 2642-2617		
174	1	7	斉藤 佐藤	C4 3333-3338		
178	1	8		C4 3339-3344		
				C10 2618-		
184	1	9		C4 3345-3347		
188	1	12	斉藤佐藤 大沢	C4 3348-3350		
192	1	13		C4 3351-3352		
194	1	14	上條 北村	CE 286		

1972

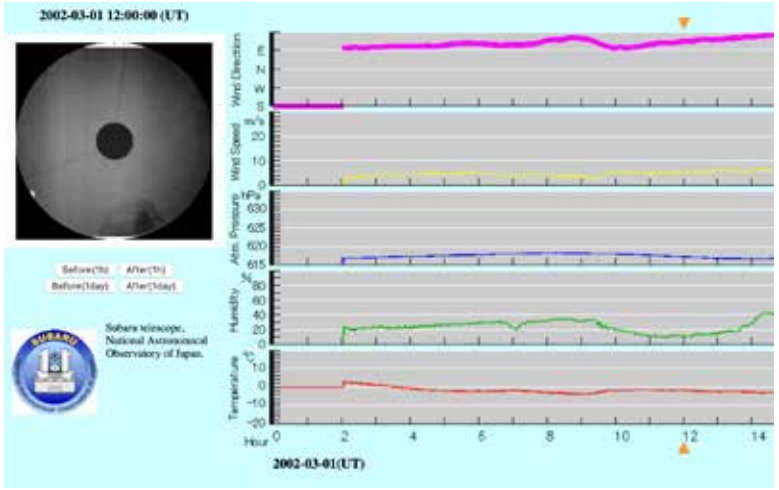
頁数	月	日	摘要		頁数	摘要
196	1	15	上條・北村	CE 287-		
				C4 3353-3354		
				C10 2619		
200	1	16		C4 3355		
				CE 288		

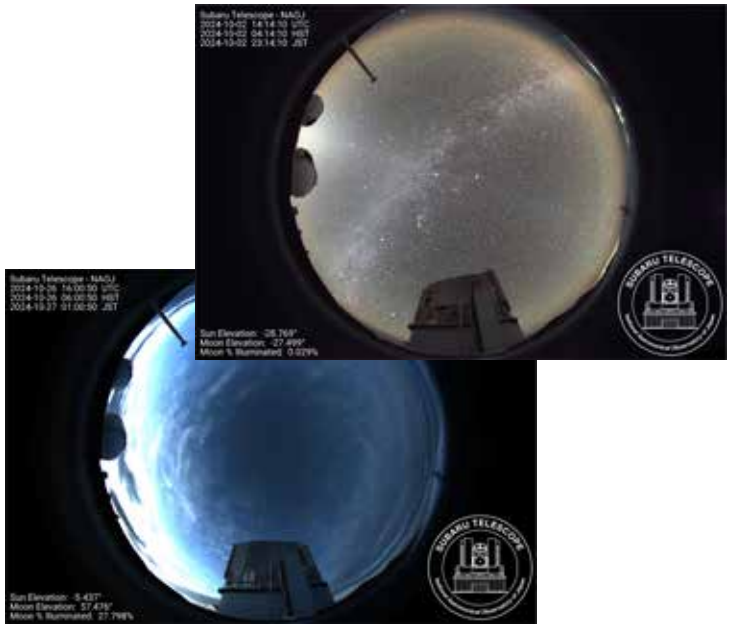
- Collaboration with Yanagisawa+

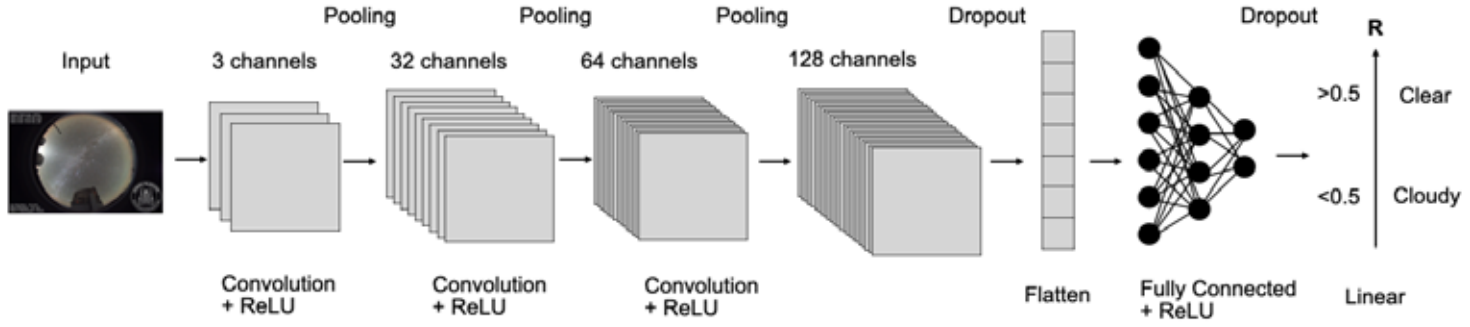
Weather Tagging of Subaru SkyMonior Images

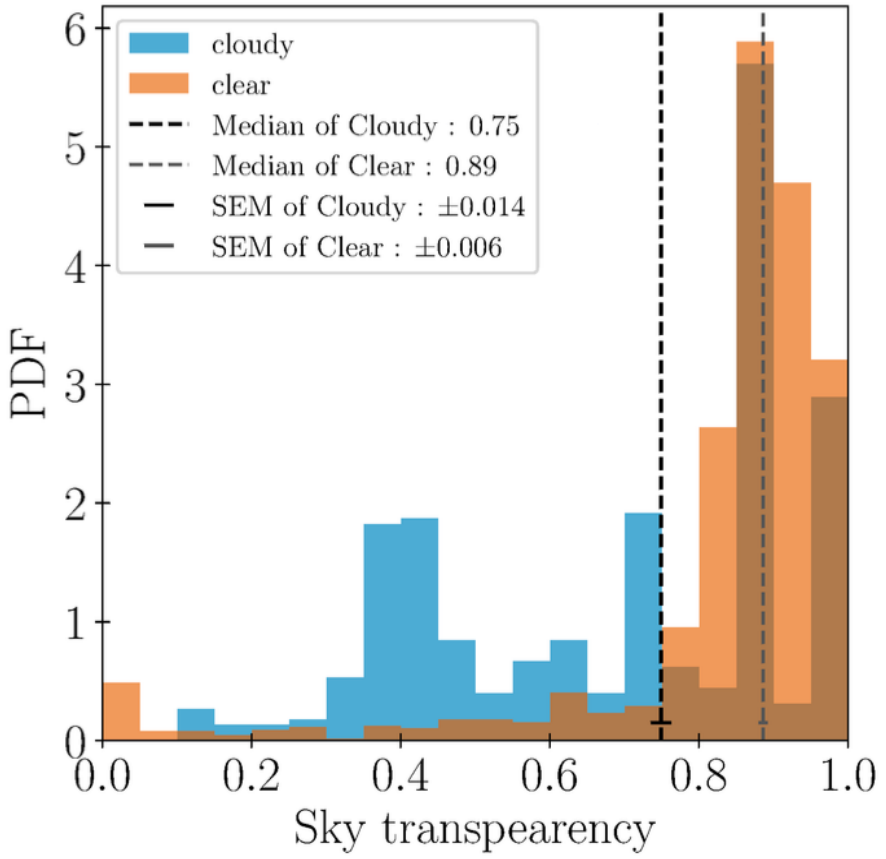
March 2025~

- Discriminating Fair or Hazy with Supervised ML (CNN) on past data (HSC QA System)
- Provide environment data for more reliable scientific analysis









New Opt/IR data archive system: Purposes/Goals

- Developing a next-generation optical & infrared data archive system to secure the Subaru data archive in the next decades
- Develop the next STARS/MASTARS archives **– for observers**
- Also will coordinate with SMOKA (or its successor system) **– for public** for comprehensive operation, with a particular focus on designing a data operation model.
- The redundancy of data backup will be decreased due to increasing data production rate and costs.

- We want to start operation well before the end of the current computer systems at Mitaka (2029/06), and start services with new systems, parallel operation with current STARS/MASTARS for ~1 year in mind for a smooth migration.

New Opt/IR data archive system: Current Status

In the phase of conceptual designing

- Mainly handled by ADC members
- Computer system for the development/testing/initial operation have been set-up and in operation from March 2025. (based on Open Access Promotion Budget by MEXT)
- Starting from designing the next generation STARS/MASTARS – current focus
- Confirming Basic functions of current STARS/MASTARS
- Iterating over questions/issues for operation model between ADC and Subaru (not complete)
- Discussions about FITS file correction policies w/ Subaru staffs
- Common development of some functions with current STARS/MASTARS (FITS file compression of HSC data, cold data back up to LTOs, MEF data handling etc.)
- Drafted development guidelines – coding rules, implementation guidelines, and the memo for implementation and execution environment.
- The review process is expected to take place within six months.

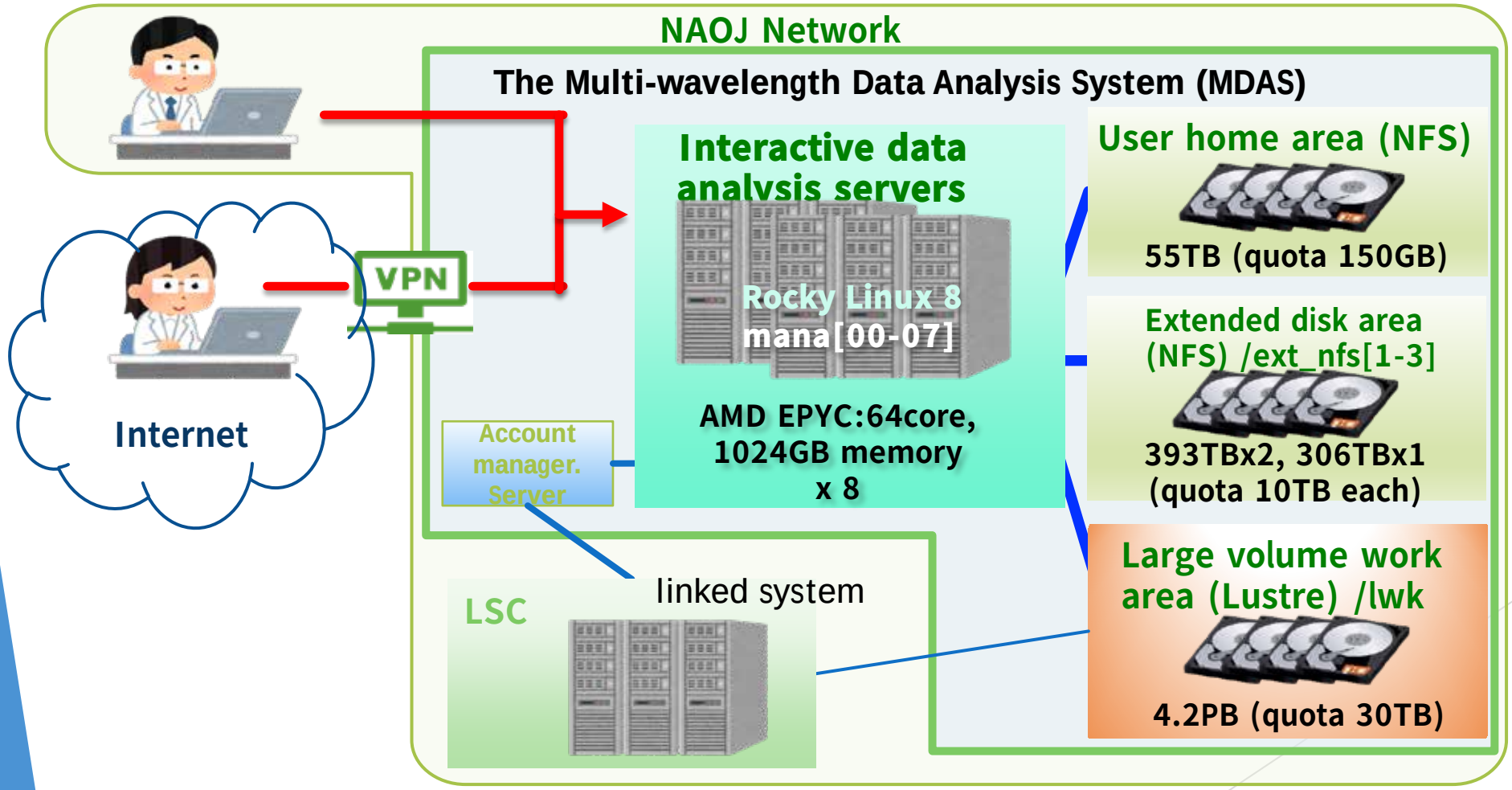
Will begin detailed designing after fixing issues and reviewing the conceptual design

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

MDAS – interactive platform

- Has largely moved to purchased system and inhouse maintenance in the last-year's update
- Operating with no big problems (there are a set of technical issues to be addressed)

Current system configuration (since July 2024)



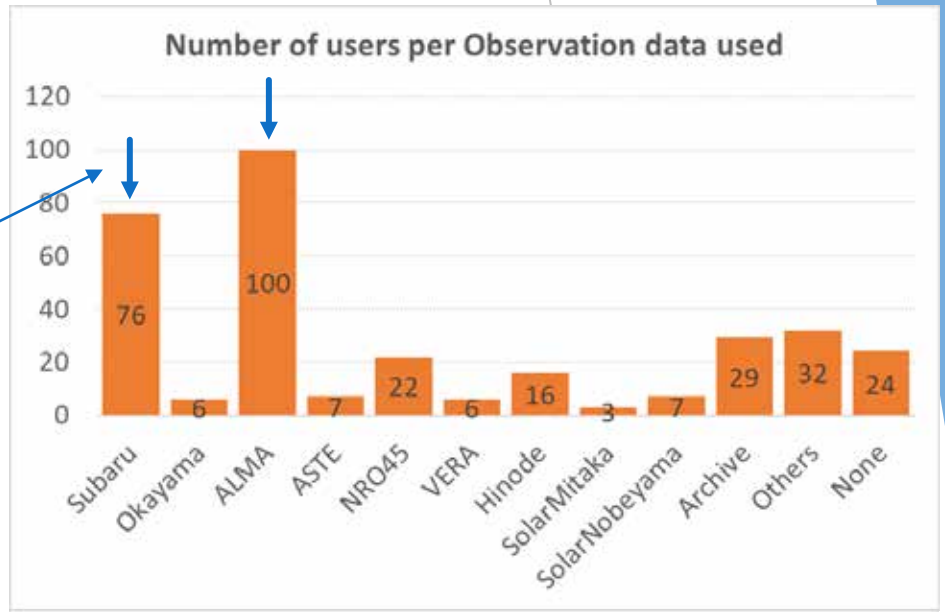
System total resources

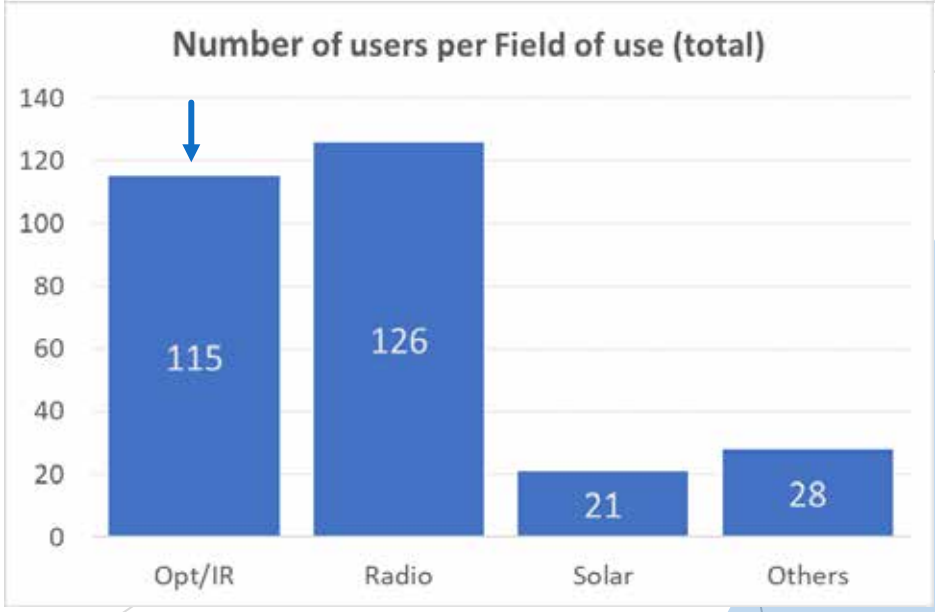
★ CPU cores: **512**
★ Memory: **8TB**
★ Storages: **5.2PB**
(home 55TB, NFS 1PB, Lustre 4.2PB)

MDAS

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

- Status and Usage
  - **Number of active users: 338** (as of the end of FY2024)
  - **Subaru and ALMA users** make up the **majority** of the active users.
  - After system migration to an in-house management system for the major components, system downtime due to maintenance and troubleshooting has slightly increased, but **overall stable operation has been maintained**.





LSC – PC cluster for large-scale processing (HSC, PFS, general use)

☆ **Since the last Users' Meeting:**
- **Increase in the number of users: 40 → 57**
- **Introduction of Quota on the GPFS area: Area usage from 100% → 93%**
- **Start of PFS business use (Compute resources are shared with HSC operations)**
- **Congestion of the general/observational user nodes**

☆ **Next Steps / Future Challenges**
- **Start of operations for the new work area with Lustre filesystem**
- **OS update (RHEL 8 based? Or should be 9?)**
- **Adjustment of compute resource allocation and queue priority**
- **System replace (The current system is 7 years old)**
- **Strengthening affinity with MDAS**

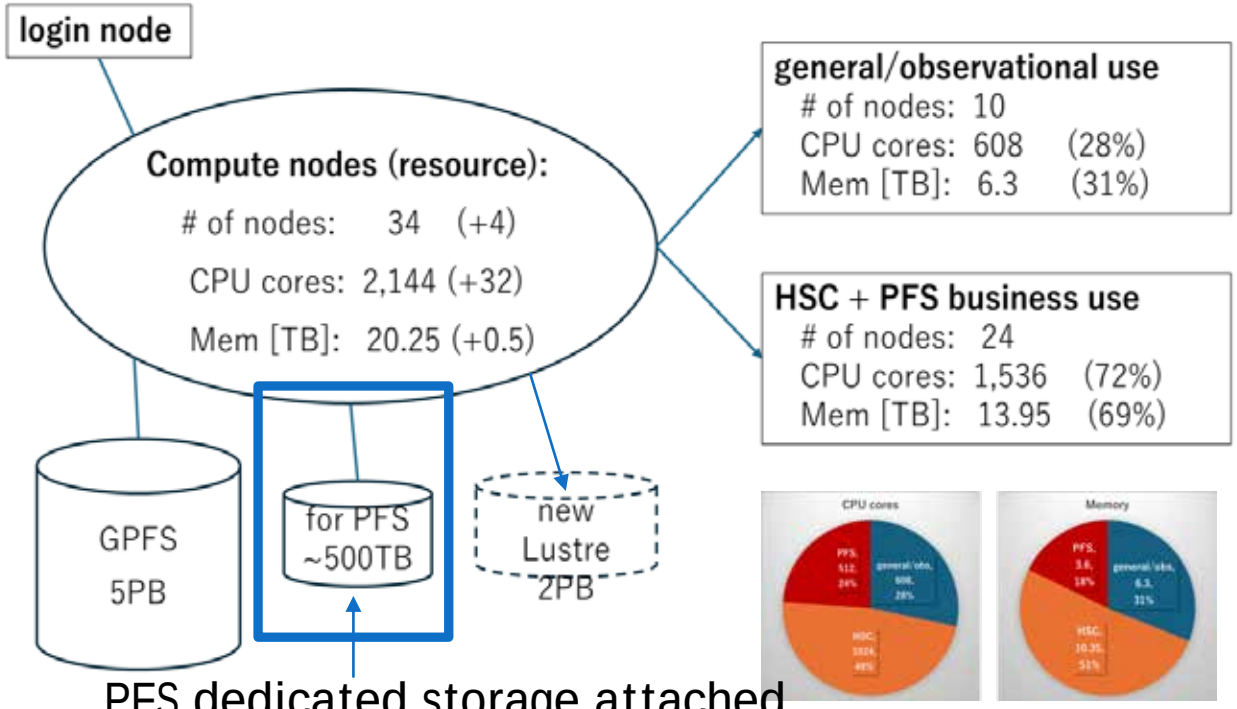
Details will be reported at the ADC Users' Meeting.

LSC

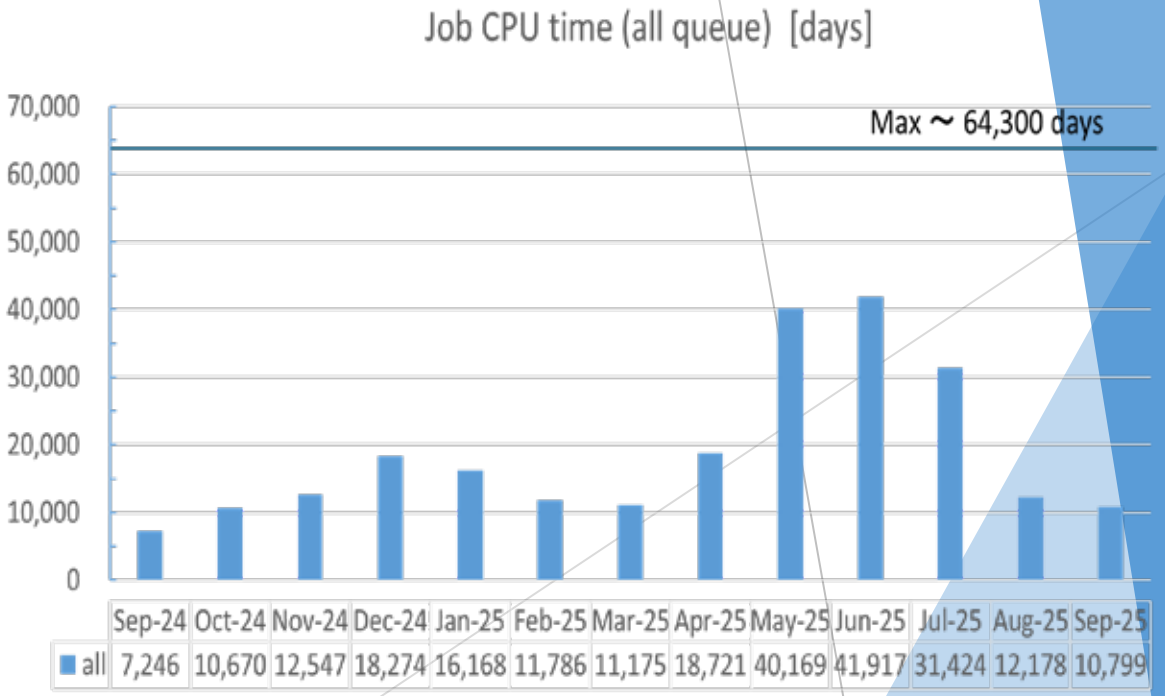
☆ Data analysis system optimized for large-scale data (HSC, PFS, …)

☆ Compute resources managed by job management system (PBS)

system configuration and resources



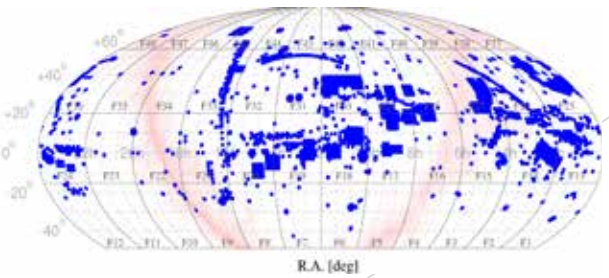
PFS dedicated storage attached
Share the CPU resources



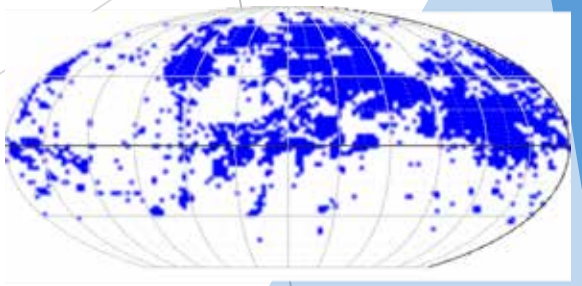
HSC/PFS/Science Platform

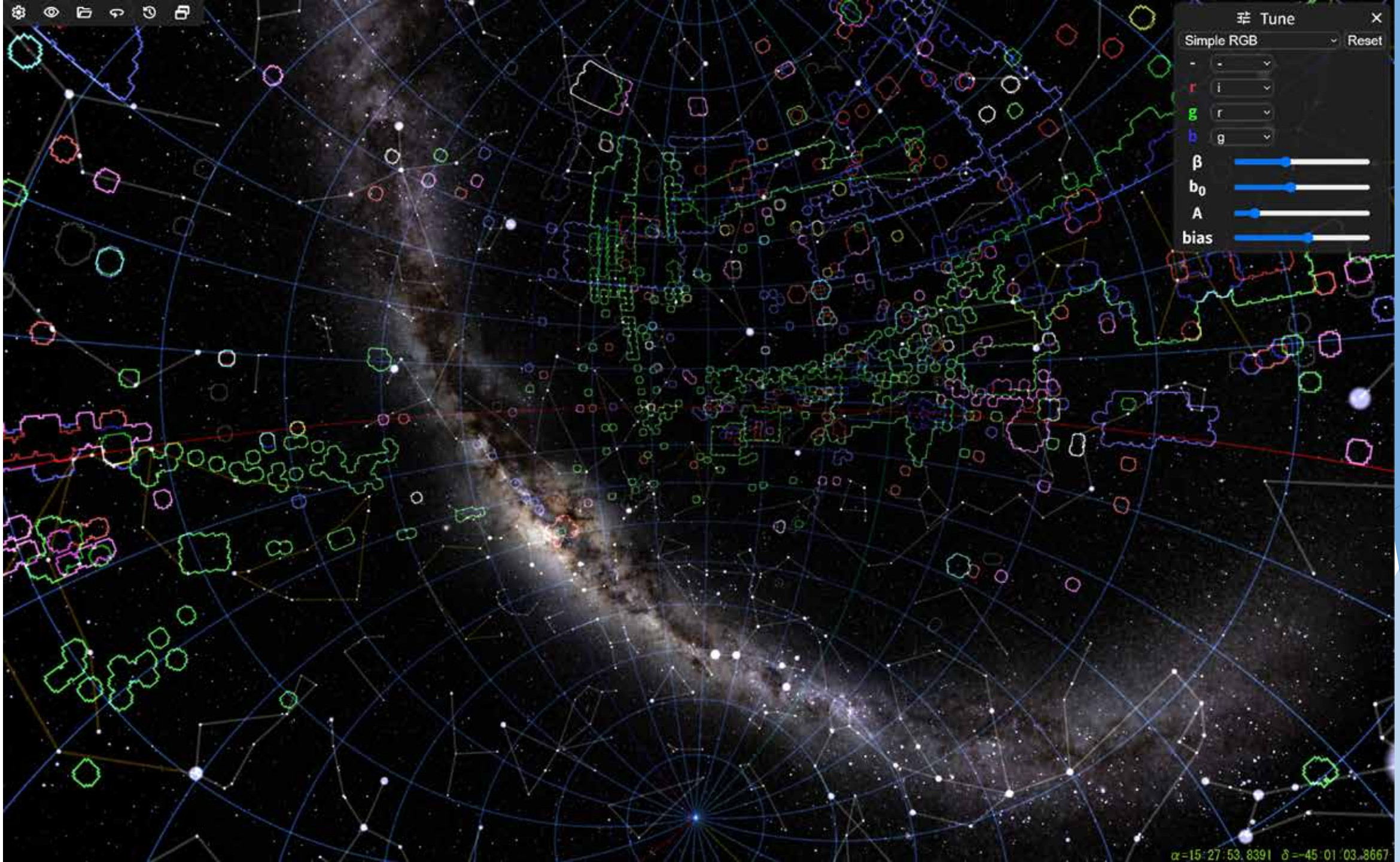
- HSC-SSP data releases
  - The latest internal release S23B (Feb, 2025) being stabilized
  - PDR4 is going to target ~August 2026
    - => long-term public archive will be operated primarily by ADC in coop. with Subaru
- HSC Legacy Archive
  - The 3rd version of Data Release (2014-2020) coming Soon
  - Products 800TB over 7315 sq.deg (17.7% of the entire sky)
  - Catalog Database will contain 1.5 billion sources in forced photometry
  - See Harasawa-san's poster
- Science Platform
  - PFS open-use operation started
  - Working on HSC-SSP version

LA2016: 3400 sq.deg



LA2020: 7315 sq.deg

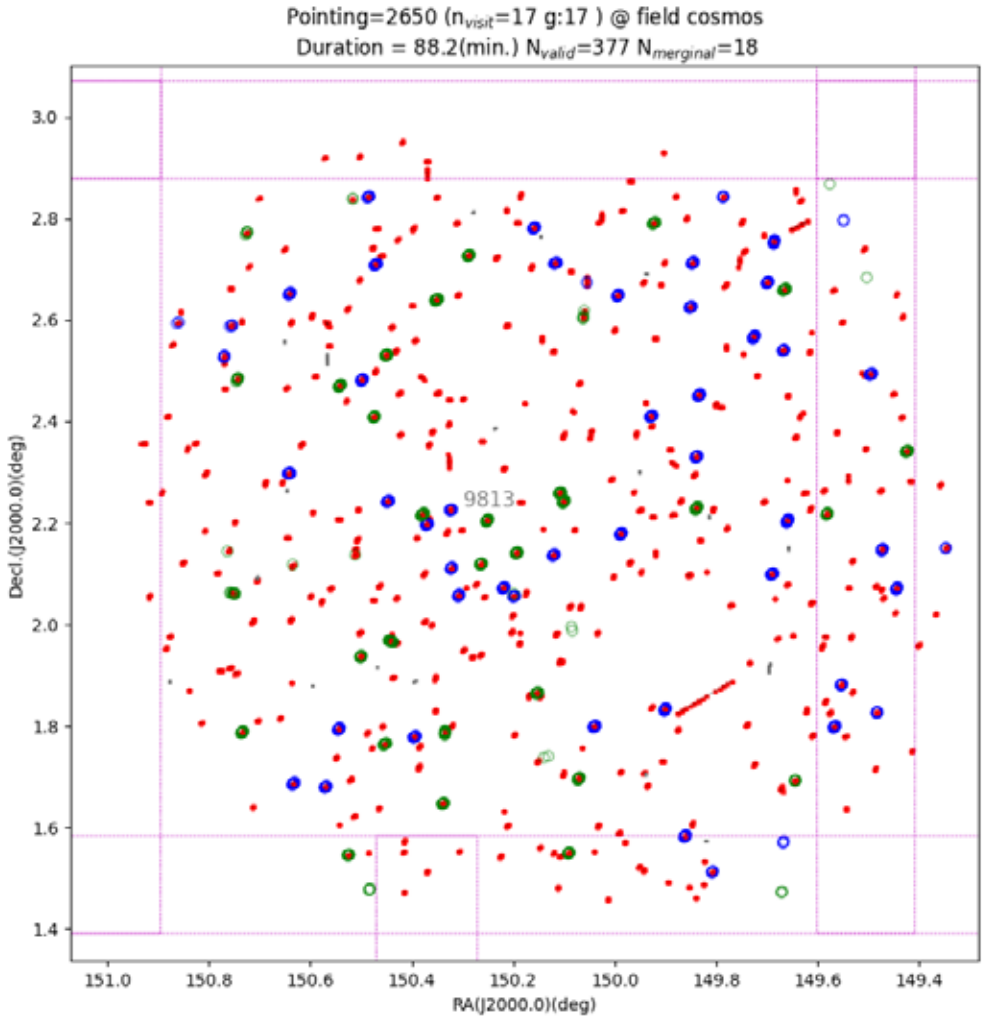



Tune
Simple RGB
Reset
-
r
i
g
r
b
g
β
b_0
A
bias
α=15 27 53.8391 δ=-45 01 03.8667

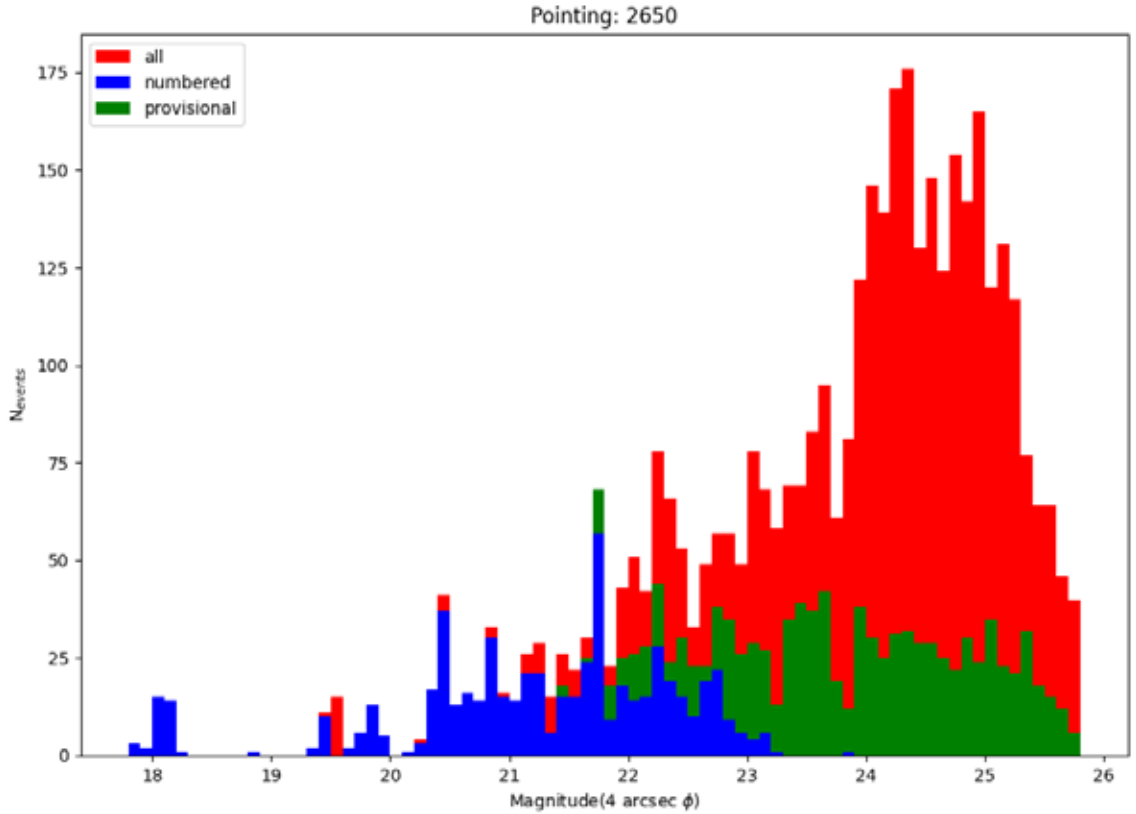
Hyper-Speed Database Development (Tsurugi)

- Application with HSC-SSP PDR3 to detect
  - Moving objects (based on CcdSrc + Coadd catalog data)
  - Variable objects (based on ForcedCcd + Coadd catalog)
- Over the past year, there has been significant progress in activities related to moving objects.
  - Identification of known moving objects
  - New detection of many faint moving objects (PDR3)
- Some initial results will be published in Ootsubo+2025 and Kasuga+2025
- The catalogs of identified/detected moving objects will be released together with the paper's publication hopefully within a year.

Sky Distribution of detected **Moving Object** candidates



Magnitudes of detected MO events (g band)
COSMOS @ 2019-04-04



Red: New detection
Blue: Numbered
Green: Provisional

- Looking forward to having further report & discussion at ADC-UM on November 18 (1 full day) at the same venue

- ADC services for the Subaru community include:
  - 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 (serving SCam, HDS data)
  - HSC https://hsc-release.mtk.nao.ac.jp/doc/
    https://hscla.mtk.nao.ac.jp/
- Developments, R&D:
  - Next-generation Opt/NIR archive
  - Science Platform
  - Database (Tsurugi)

SMOKA

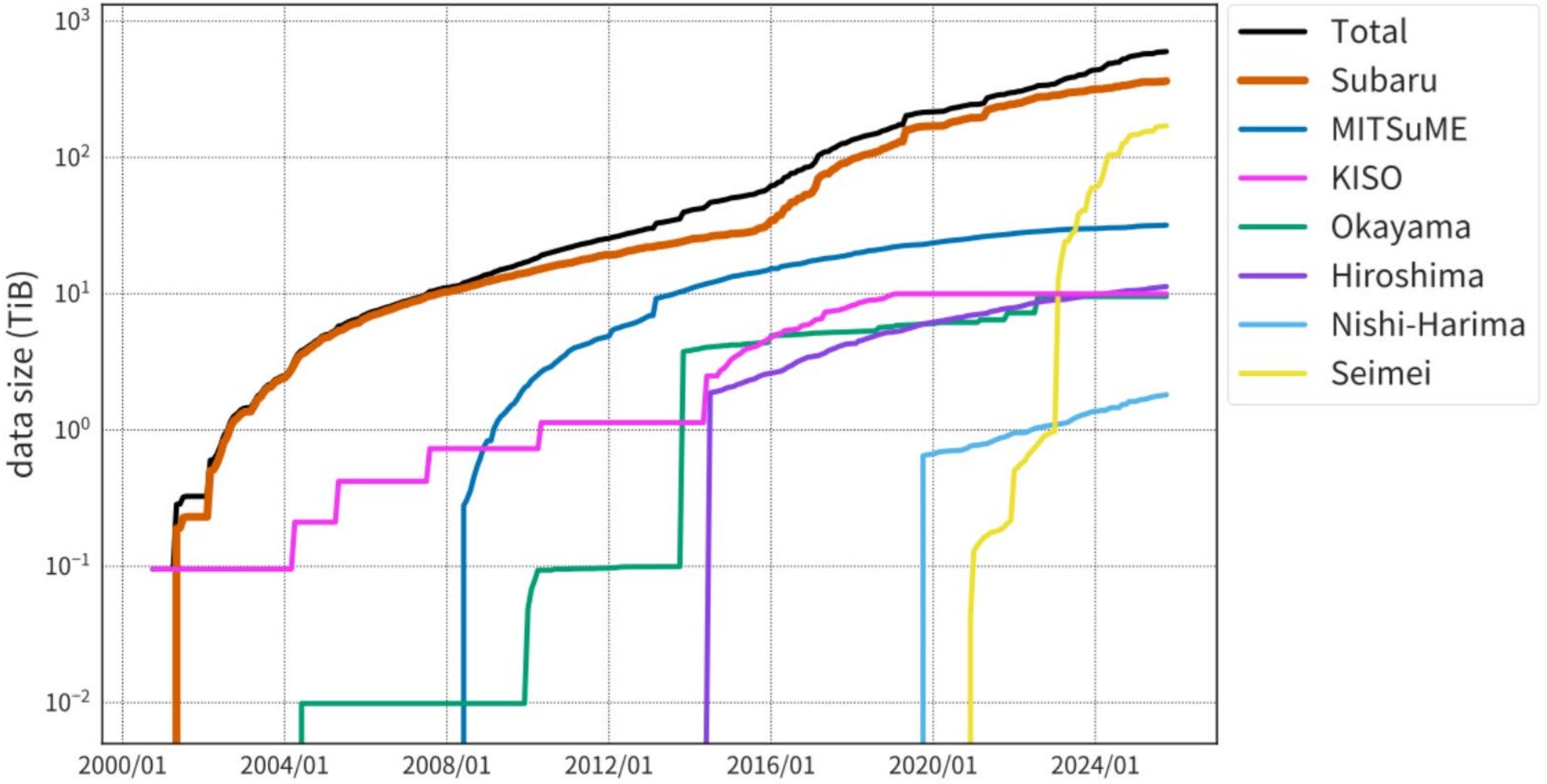
Current Status and Future Plans

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

Subaru Users Meeting FY2025

The Trend of SMOKA Released Data Volume

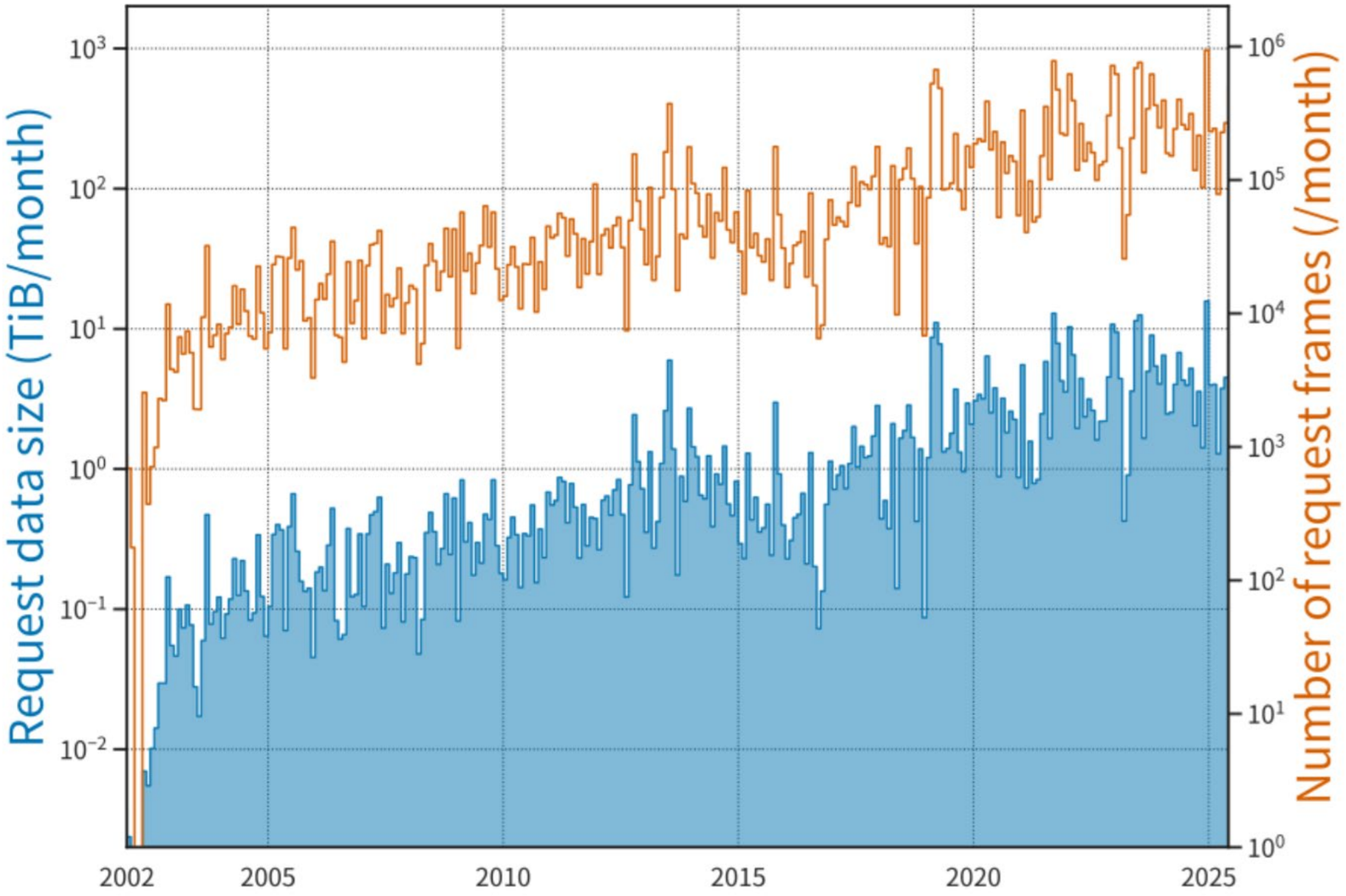
- Total data
  - 42 million frames
  - 599 TB
- Subaru data
  - 21 million frames
  - 364 TB



SMOKA Usage Status

- Annual registered users:
  - 219 (FY2024)

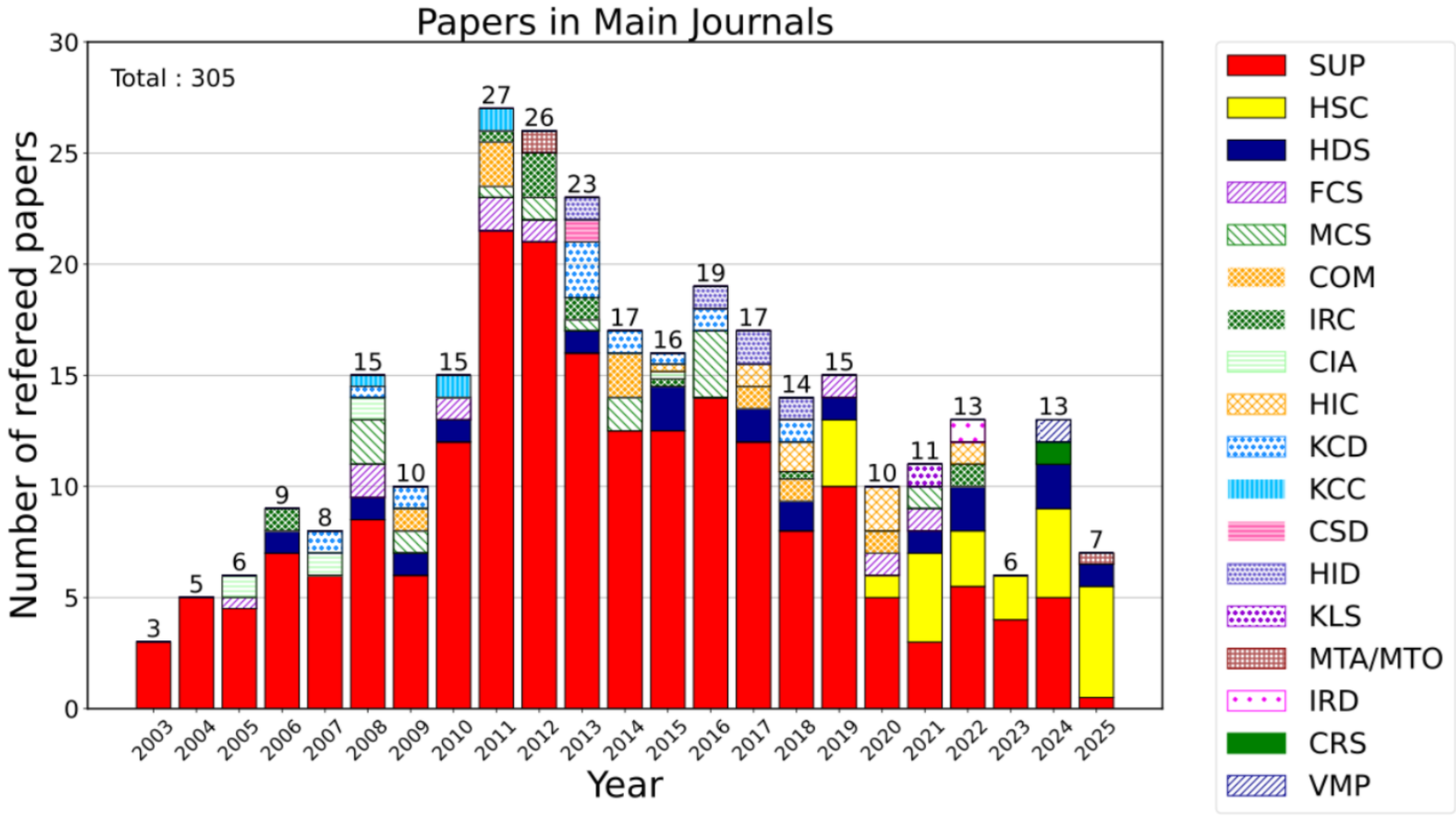
- Monthly data request volume
  - Number of frames: 0.1 - 1 million
  - Data size: 1-10 TB



Number of Papers Using SMOKA

- Toal: 305 (as of Oct. 23, 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

- New data release
  - March 2025
    - CNN-based Clear/Cloudy Classification for Subaru All-sky Monitor Images
    - GAOES-RV of Seimei telescope
    - Tomo-e Gozen Raw Data (partially) of Kiso Schmidt
  - October 2025
    - Okayama 188cm Photographic Plate Digitized data (BMP only)

- Future data release
  - Subaru
    - Fast PDI, MEC, PFS
  - Seimei
    - MIDSSAR
  - TAO
    - NICE, MIMIZUKU, SWIMS, ...