## Web-based Hyper Suprime-Cam Data Providing System

Michitaro Koike,<sup>1</sup> Hisanori Furusawa,<sup>1</sup> Tadafumi Takata,<sup>1</sup> Paul Price,<sup>3</sup> Yuki Okura,<sup>1</sup> Yoshihiko Yamada,<sup>1</sup> Hitomi Yamanoi,<sup>1</sup> Naoki Yasuda,<sup>2</sup> Steven Bickerton,<sup>2</sup> Nobu Katayama,<sup>2</sup> Sogo Mineo,<sup>2</sup> Robert Lupton,<sup>3</sup> Jim Bosch,<sup>3</sup> and Craig Loomis<sup>3</sup>

<sup>1</sup>National Astronomical Observatory of Japan, Osawa 1-12-35 Dai 3 Sakura Heights 325, Mitaka, Tokyo 1810015, Japan

<sup>2</sup>*Kavli Institute for the Physics and Mathematics of the Universe, Kashiwa* 277-8583, Japan

<sup>3</sup>Princeton University, Princeton, NJ 08544, USA

**Abstract.** We describe a web-based user interface to retrieve Hyper Suprime-Cam data products, including images and catalogs. Users can access data directly from a graphical user interface or by writing a database SQL query. The system provides raw images, reduced images and stacked images (from multiple individual exposures), with previews available. Catalog queries can be executed in preview or queue mode, allowing for both exploratory and comprehensive investigations.

## 1. HSC and HSC Survey

Hyper Suprime-Cam (HSC) is a new, gigantic camera for the prime focus of Subaru telescope on Mauna Kea, Hawaii. It covers a 1.5 degree diameter field of view with 104 scientific grade Hamamatsu CCDs (Miyazaki 2012). HSC saw first light in January 2013, and is now in the phase of scientific verification. With the combination of wide field coverage  $(1.77 \text{ deg}^2)$ , large aperture (8.2 meter) and excellent image quality of the Subaru telescope, HSC will provide a large, high-quality dataset for various scientific programs. An HSC Collaboration team (consisting of astronomers in Japan, Taiwan and Princeton University) successfully proposed a Strategic Survey Program (SSP) to use 300 nights over 5 years to survey wide (~ 1500 deg<sup>2</sup>), deep ( $4 \times 7$  deg<sup>2</sup>) and ultradeep  $(2 \times 1.77 \text{ deg}^2)$  layers in both broad (grizy) bands and some narrow band filters. The expected data volume is about  $\sim 3$  million FITS image data, which corresponds to 300 TB after reduction by our data pipeline. The final catalog is expected to contain  $\sim$ 1 billion celestial objects, which are multiply measured on each exposure, and a total of  $\sim 6$  billion individual detections/measurements. As the output data is so large, it is important to develop an effective data archiving and retrieval system, to accelerate the scientific activities using HSC SSP data.

# 2. Implementation Details

As our target is to develop the data providing system for the SSP collaborators in Japan, Taiwan and Princeton, we focused on developing a web-based system. The system has to provide the resultant FITS data and some metadata of data pipeline, various measurements (e.g., magnitudes, coordinates, shapes and their time dependencies) for all celestial objects in our catalog, and information on data quality and observational enviroments, etc. We designed a Data Archive System (DAS) for providing the data files, and Catalog Archive System (CAS) for enabling various types of catalog queries. The information on data quality metadata, etc., is designed to be linked to the relevant image and catalog data.

# 2.1. Data Archive System (DAS)

The DAS provides images taken by HSC along with source catalogs from the same images. The image data includes raw images, reduced images, coadded images (i.e., stacks of multiple exposures) and their corresponding source catalogs. The catalogs consist of various measurements, such as magnitudes, coordinates, and shapes of the sources. We can select and retrieve the data by following way:

- 1. Query for frames (Figure 1)
- 2. Select frames to download (Figure 2)
- 3. Wait for the retrieval job to finish (Figure 3)
- 4. Download the provided files (Figure 4)

				SOP BAL						
							_	_		
	Enablemed	Shinna houses	bia.inch	Genalon, Bearth	Aplikiers Lepinsersed;	<b>CALIBOR</b>	<b>Source</b>	<b>BAR</b>	punt	k
oata :	Search									* 10
Protect	Resald Dis	weight(Frame) 00	rendqL(Reals)							
Filte										
Seco	- Dean	Caso Case	Case-							
Obsex	vation D	ate								
(Same	0000									
Page 2	013-01-01 0	0.00								
800 3	013 03 06 0	0.00								
Spatz	e1									
West	0.000									
-	140	degre	ani: 30		degree					
Petitore	10	anna								
Ordea										
		IN COMPANIE								
	us inder									
Search										

Figure 1. Query for frames: specify filter, observation date or spatial area in the sky.



Figure 2. Select frames: on the preview window in the right bottom in the page, the regions corresponding to the selected frame are highlighted.

# 2.2. Catalog Archive System (CAS)

CAS provides source catalogs of the data from HSC, either for individual CCDs or coadded images. The catalogs consist of various measurements, such as magnitudes, coordinates, shapes and their time dependencies of the sources. The data retrieval flow is as follows:

#### Web-based Hyper Suprime-Cam Data Providing System

							345 422							
	1 12 0 dellocasa	K.Br										0.00		ä
	Instituted this	head	Jons		bia.im	ah sea	iles. Annesh 1840	Bjert i	Legionerteil 🛤	803	form finite	panels	Le	5
	et anosphel, 110-010													
													+ 00	
	othis name will be	nie.	ated in		and the second s									
					-	result a	questi at		started at.		distant of			
	data angent (free) < ()				MATT									
	ten representationers (1)				20MR	-	312-04-04 12-00-2				2443-49-24 32+43			
	Arts, 10(101) (1048) (1)					-					2412.00.04 14.00	4.0		
	data angent (franc) < ()				1000	644	811-18-(8 88.04-18 F		1013-03-04 03-04-09		1913-09-34 10-38-09			
	inis represidentes ( )				10HB	alasa .	313-0113 12-0114		2012/09/12 20:2014		2443-00-13 16+24	-0		
	Anton and participant (	-	-			-						44		
	data angent (monta) - (	hint	14		1048	diam.	883-88-13 13-18-1		001-00-01 30-01-07		(813-05-13 16+13			
	in mention (	8411			iona -	-					1010-00-03 10:03	- 40		
Log														
	-		status		-	panet et		-	e 10		advent of		-	
	setting percent 11		1048		-		11.06.11							
	satalog gamps 11		and the second		ihan -	ana-as a 11.07.24				24	2012/00/12 10:17:00			
	saturate gamps []		1044		Child I	*****	10.00.01				11-01-12 10-12-09			
		Berner Standard (1444)      Berner Standard (1444)	montainer, sementiener, i.i.e.Schrift montainer, i.i.e.s	even prevente. Life/BKI (************************************	event semplefiel. (147511)  (147511)									

Figure 3. Wait for retrieval: when the job finishes, it's status is shown as "DONE".

data         regret/final/s         1         1         1           U         Max         1         1         1           U         Max         1		Gaableoard	Coloran Desease	Inte describ	Crisics Search	MyDB Joost Laplanes	and failure &	ein feln	parents La
0         No.1000.05.0.04         No.100.05         No.100.05           0         MO.1000.05.0.04         No.100.05         No.100.05           0         MO.1000.06.01.04         No.100.05         No.100.05	Inclusion	COMPAND.			tes the tory good	1000-000-00-0		and a star rate of	
05         405         50,400		C.Lenser				ataa (898a)			
0         09,40000-001,00+         10,110,00         6maski           0         00,40000-001,00+         10,110,00         6maski           0         00,40000-001,00+         10,110,00         6maski           0         00,40000-001,00+         00,000         60,000           0         00,4000-001,00+         10,110,00         60,000		101.070250.01					10,001,000	for last	
0         000-0010-011.0xv         10x10.0x0         00x0000           0         000-0010-011.0xv         00x000         00x000           0         000-0010-011.0xv         10x10.0x0         00x000           0         000-0010-011.0xv         10x10.0x0         00x000		100-001010-00	1.014				10, 221, 484	download .	
Image:		800-000000-00	11.634				10,311,414	tion load	
0 98-0010-01.034 (9.17.49) doklad		101.01010.00	1.5.1				10,000,000	EDG.AME	

Figure 4. Download: files can be downloaded separately by clicking the appropriate link or all at once by executing the wget command shown at the top of the page.

- 1. Select objects (Figure 5)
- 2. Preview the result (Figure 6)
- 3. Submit the query job. The user can select the output format. Available formats are: CSV, SQLite, Fits-bintable.
- 4. Wait for the query job to finish.
- 5. Download the result.

		CHARGE TO THE P		SIP BAL						10
					NyOD (cost Lepimented)	biles	Summ	Bola .	parents	Lines
Catalo	g Sear	ch								
Tirect22										
0000 III	1000 1077 2000 0.007 2000 0.007 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.00000 0.0000 0.0000 0.0000 0.0000 0.00000 0.0000	<pre>https://www.sensity.com build compared and build compared and bui</pre>	s. conserved global c. in Add	unnun, kulu par	HIT, C.C.Institutto, c.ma, p a preside protocological desp de t		et, err		10, d. <b>.</b> 00	1
	Sentas Check									
	See Ced	Ine C Eller birneb	in Convisio							
Solders			oprass team							
								trans		

Figure 5. Select objects: specify filter, observation date or spatial area in the sky.



Figure 6. Preview: On the preview window in the right bottom in the page, the regions corresponding selected frame are highlighted.

### 2.3. Web User Interface

Currently, the web user interface (UI) is implemented mainly in PHP and JavaScript. The web UI has an image preview window (bottom-right in Figure 2), implemented using Leaflet, a modern open-source JavaScript library for interactive maps. This shows an image of the sky, and the user can pan and zoom, similar to Google Maps. In the DAS preview page, regions corresponding to the selected frames are highlighted in the preview window. We made a extension of Leaflet that can handles 16bit/pixel images and interactively calibrate color tone. (Figures 7 and 8)

Koike et al.



Figure 7. Image preview before calibration



Figure 8. Image preview after adjusting sliders on the bottom-left of the page.

# 2.4. Authentication

We plan to apply LDAP authentication for user access control on a user-by-user basis. By sharing the LDAP database with the Hilo base of Subaru Telescope, access rights to the both DAS and CAS contents will be limited to only the registered Co-I members.

Each user is assigned their own temporary disk space (capacity typically  $\sim 20$  GB) as a staging area to store any files generated by queries and to be downloaded by a user, including symbolic links to the products, combined coadd images over a requested sky patches on the fly, or catalog CSV files etc.

### 2.5. Job control and Backend Processes

Job control is done by Python scripts running on each of processing and database servers, with a shared job control table. The maximum number of simultaneous processes is limited on each server, for load balancing and stable operation of the system.

The web server updates queued jobs in the job control table, while the database server processes the query jobs. The processing server makes symbolic links to images or combines the coadd images from multiple sky patches into a single image per request.

Acknowledgments. We would like to thank to all members of the HSC development team for continuous support of our development work.

### References

Miyazaki, S. 2012, in Ground-based and Airborne Instrumentation for Astronomy IV., edited by S. Miyazaki, vol. 8446 of SPIE Conf. Ser., 84460