*Astronomical Data Analysis Software and Systems XXIII*
*ASP Conference Series, Vol. 485*
*N. Manset and P. Forshay, eds.*


# Web-based Hyper Suprime-Cam Data Providing System

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

[1]*National Astronomical Observatory of Japan, Osawa 1-12-35 Dai 3 Sakura Heights 325, Mitaka, Tokyo 1810015, Japan*

[2]*Kavli Institute for the Physics and Mathematics of the Universe, Kashiwa 277-8583, Japan*

[3]*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 $\deg^2$), 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 ($\sim$ 1500 $\deg^2$), deep ($4 \times 7$ $\deg^2$) and ultra-deep ($2 \times 1.77$ $\deg^2$) 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)

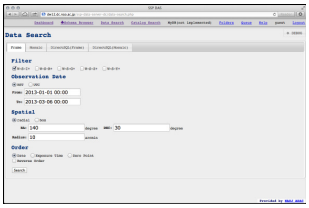


Figure 1. Query for frames: specify filter, observation date or spatial area in the sky.

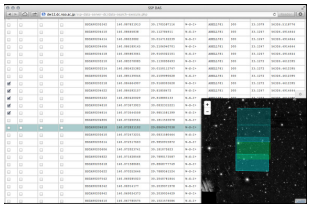

Figure 2. Select frames: on the preview window in the right bottom in the page, the regions corresponding to the selected frame are highlighted.

### 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:

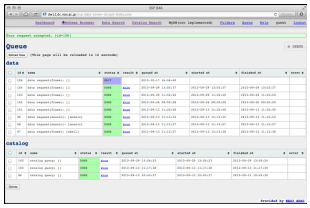

Figure 3. Wait for retrieval: when the job finishes, it's status is shown as "DONE".

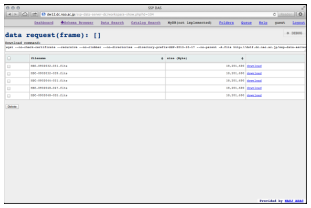

Figure 4. Download: files can be downloaded separately by clicking the appropriate link or all at once by executing the `wget` command shown at the top of the page.

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.

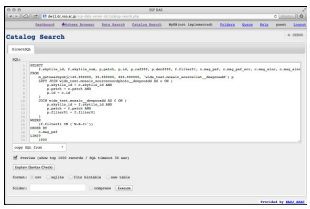

Figure 5. Select objects: specify filter, observation date or spatial area in the sky.

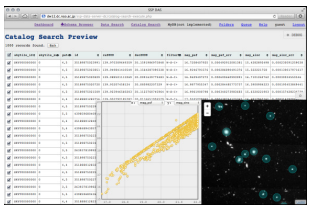

Figure 6. Preview: On the preview window in the right bottom in the page, the regions corresponding selected frame are highlighted.

### 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)

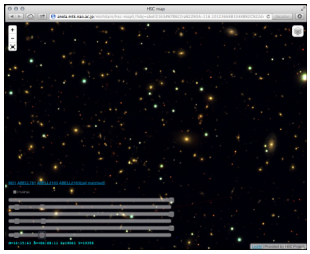

Figure 7. Image preview before calibration

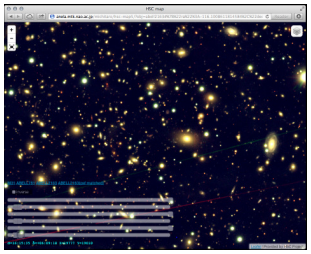

Figure 8. Image preview after adjusting sliders on the bottom-left of the page.

### 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