

# Creating and managing very large HiPS: the Pan-STARRS case



## Abstract

HiPS (Hierarchical Progressive Surveys) is a proven Virtual Observatory standard which enables an efficient way to deliver potentially huge images collection and allows for fast visualisation, exploration and science applications. CDS has recently published the HiPS for Pan-STARRS g and z-bands images, covering three quarter of the sky at a resolution of 250mas per pixel. We will describe in our poster the challenges we faced and the lessons learnt in generating and distributing these HiPS made of 47 million FITS tiles, amounting to 10 trillion pixels and more than 20TB per band. In particular, we will detail the methods we developed to optimize the generation, the storage and the transfer of the HiPS. In addition, a color HiPS, based on the two already available HiPS, has been made available and can be visualised from HiPS clients, like Aladin Desktop or Aladin Lite.

## Links

- Pan-STARRS data in Aladin Lite [aladin.unistra.fr/AladinLite/showcase/PanSTARRS-DR1/](http://aladin.unistra.fr/AladinLite/showcase/PanSTARRS-DR1/)
- Available HiPS images surveys [aladin.u-strasbg.fr/hips/list](http://aladin.u-strasbg.fr/hips/list)

## 1 - Original files download



- Downloading of images with a single `wget` was not sufficient to take full advantage of the network link between STScI and CDS: transfer rate was only 12 MB/s.  
➔ using several **wget connections in parallel** increased the transfer rate to 46MB/s in average.
- **Total transfer time** for one band went down from **12 to 3.5 days**.

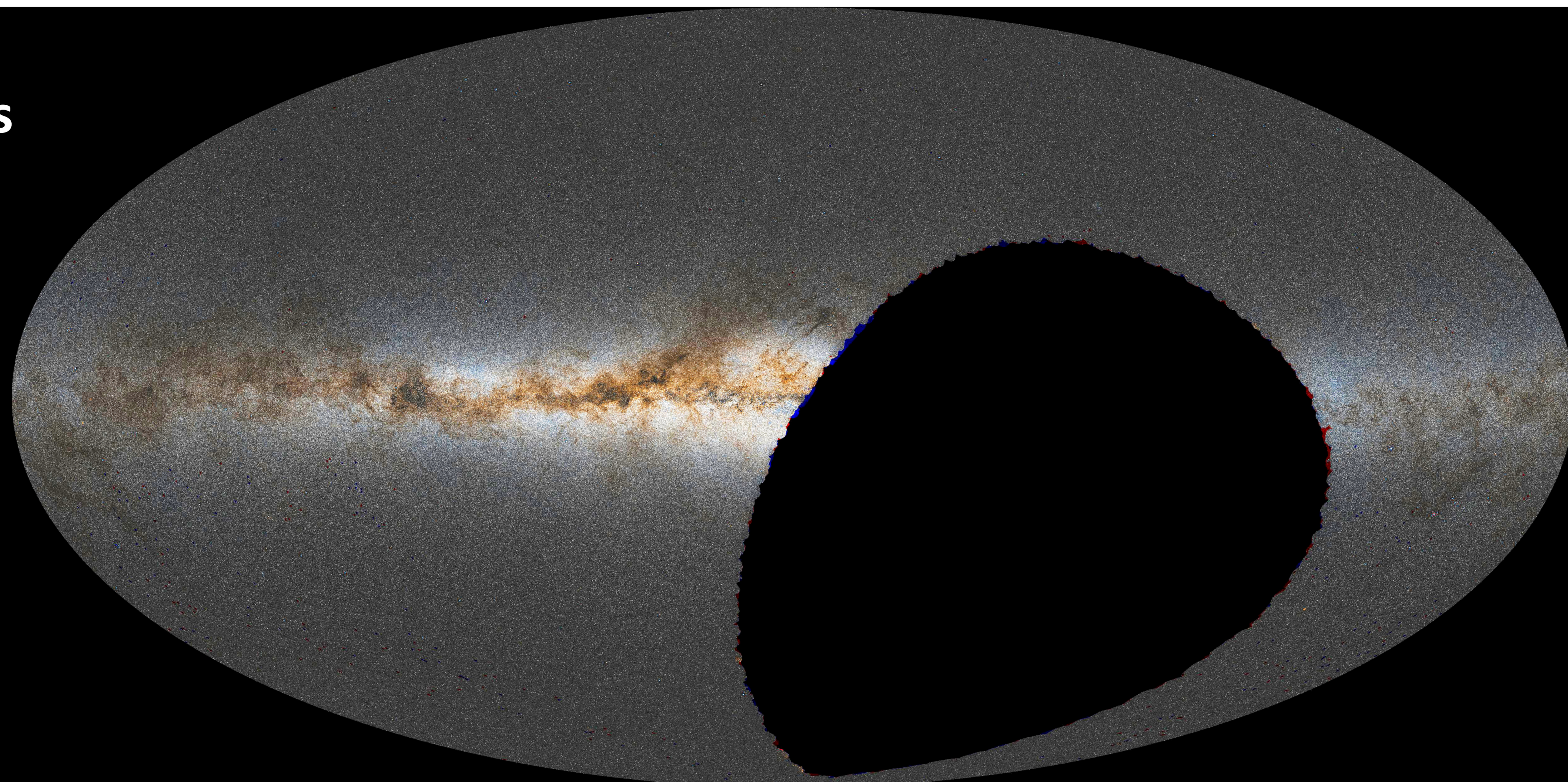
## 2 - FITS tiles generation



- Pan-STARRS images come as RICE\_1 tile-compressed FITS files.
- In our initial tests, FITS images had first to be uncompressed launching parallel instances of `funpack`.  
➔ **Hipsgen tool upgrade** to support RICE\_1 images as input.
- This improvement **saved us 4.5 days** and has brought down the **FITS tiles generation time** to 20 days (on a 5 years old server with 128 GB RAM and 32 hyper-threaded cores)

## PanSTARRS images

- ▶ **5 bands:** g, r, i, z, y
- ▶ **250 mas/pixel**
- ▶ **RICE compressed**
- ▶ **15 TB per band** (200k images)



## PanSTARRS HiPS

- ▶ **2 bands available:** g, z (+ color HiPS)
- ▶ **10 trillion pixels per band**
  - ▶ HEALPix order 20 (200 mas resolution)
  - ▶ 47 million tiles

## 3 - On-the-fly JPEG tiles generation

- Generation of JPEG tiles is a 20 days long process and result to generate 6TB of additional HiPS data.  
➔ We created a **Python service**, based on the Falcon framework and using `astropy.visualization` library, which **generates on-the-fly JPEG tiles** from the existing FITS tiles.
- Thanks to Apache rewriting rules, access to those JPEG tiles is totally transparent for HiPS clients.

## 5 - RGB tiles generation



- **Red:** z band, **Green:**  $0.5 * (z+g)$ , **Blue:** g band.
- **Python script to generate color JPEG tiles** from FITS tiles, with **Lupton-like arcsinh stretch** to maximize contrast.
- The AITOFF view above has been generated from the color HiPS published in the HiPS network and displayable in HiPS clients (Aladin Lite, Aladin Desktop, etc).

## 4 - Transfer to production server



- Our initial transfer process, based on `rsync` solely, was slow, averaging 12MB/s (1TB/day) on an internal Gigabit link.  
➔ using **parsync** (a **parallel rsync wrapper**), we increased the transfer rate to 100 MB/s. The **transfer time** has been **reduced from 20 days to less than 3 days**.

## Conclusion

**Total HiPS generation time** (from original FITS download to release) of one Pan-STARRS band has been **reduced from 80 to 30 days**. This new streamlined process has been put into practice for the creation of the z band now available in the HiPS network and will be applied to the generation of the remaining r, i and y bands, but also to other large image surveys we will process in the future. It also provides us with additional flexibility regarding creation of JPEG tiles.



Thomas Boch, Pierre Fernique

