HiPSgen: The Hierarchical Progressive Survey Generation Tool

Caroline Bot, Pierre Fernique, Anais Oberto, Daniel Durand, Thomas Boch, Mark Allen, Mihaela Buga, François Bonnarel

HAL Id: insu-03842964
https://insu.hal.science/insu-03842964

Submitted on 7 Nov 2022

HiPSgen: the Hierarchical Progressive Survey generation tool

Caroline Bot,[1] Pierre Fernique,[1] Anais Oberto,[1] Daniel Durand,[2] Thomas Boch,[1] Mark Allen,[1] Mihaela Buga,[1] and François Bonnarel,[1]

[1]*Centre de données de Strasbourg (CDS), Observatoire Astronomique de Strasbourg, UMR7550, Strasbourg, France;* `caroline.bot@astro.unistra.fr`

[2]*National Research Council Canada, CADC, Canada*

Abstract. Hierarchical Progressive Surveys (HiPS) have become a widely used standard for storing and exploring large image surveys in the last decade. This success is a combined effect of different tools and contexts (the possibility of tile by tile visualisation, the evolution of HiPS clients (e.g. Aladin Desktop, Aladin Lite, firefly, WWT), the IVOA context and standardisation, . . .). Among these, the HiPSgen tool has had a special role. HiPSgen enables anyone to easily transform a set of individual images into a HiPS and its usage has enabled the creation of a vast collection of surveys.

Hierarchical Progressive Surveys (HiPS; Fernique et al. 2015) have become a widely used standard (Fernique et al. 2017) for storing and exploring data sets. This is witnessed beyond usage statistics, as new clients like FireFly/IPAC, WWT/AAS, WWT/ChinaVO, Stellarium and DIGISTAR have now adopted this VO image standard. The success of HiPS is probably due to some converging facts such as the expansion of the expansion of hierarchical visualisation techniques based on multi-resolution tiles (e.g. cartesian methods like the one used by Sky-Map, cylindrical methods like Google Sky, and regular tesselation of the sphere like HTM or HEALPix), the adaptation of a well known tool (Aladin Desktop and then development of Aladin Lite), the simplicity of the HiPS standard, the IVOA context, different partnerships between institutes (e.g. ESAC, JAXA, IRAP), etc. Among these keys to success, the development and accessibility of HiPSgen has probably played a major part.

HiPSgen[1] is a generation tool that produces HiPS out of a set of individual images. HiPSgen enables the building of a large collection of surveys, in a relatively short timescale. These surveys span a large variety of characteristics including resolution, wavelength, depth, coverage, . . . Being able to generate surveys as diverse as Planck, the Digital Sky Survey or the PanSTARRS surveys has most probably been a decisive point to demonstrate the usefulness of the HiPS format.

1. What is HiPSgen?

HiPSgen is the tool to project a set of individual images to create a HiPS. These input images can be FITS files, JPEG files with associated WCS headers, cartesian allsky maps or HEALPix maps. The reverse tool (projecting a HiPS onto a standard frame

---

[1] `https://aladin.u-strasbg.fr/hips/HipsIn10Steps.gml`

defined by a standard FITS header) has been developed as a service called HiPS2FITS (Boch et al. 2020). HiPSgen has been developed in the Java programming language and it has been integrated into the Aladin Desktop application. It is therefore compatible with Linux, Mac or Windows and available to anyone using this application (e.g. computer engineer, astronomer, project manager, amateur astronomer, students). HiPSgen can also be run using the batch mode (command line). One of the advantages of using HiPSgen inside Aladin Desktop is the possibility to instantaneously visualize the resulting HiPS.

The development of HiPSgen started in 2009 with a prototype called Skygen which was then further developed. The first official release was in 2012. The initial HEALPix library was replaced by a new and independent version in 2019. Usage and interaction with CDS partners including CADC, ESAC or IRAP further consolidated and improved the tool according to its usage and the user needs.

2.   Projection onto a HiPS

HiPS are based on an HEALPix tesselation of the sky. The projection step from the individual images to the HiPS involves a resampling onto HEALPix cells. When necessary, a bilinear interpolation between individual surrounding pixels has been chosen. This method is simple, fast and robust. We note that compared to other projection methods, there is no absolute conservation of the signal. This is however usually a good-enough and widely used method for a lot of scientific purposes given the accuracy of the data and other sources of uncertainties. In cases where access to the actual original images is necessary, one can use the possibility to access progenitor images within the HiPS, as described in the IVOA note on progenitors[2].

Overlapping regions between individual original images are dealt with by doing a simple averaging between pixels in the overlap regions. Simple background subtraction can be performed. However, unlike tools like Montage, no advanced treatments of large scale drifts or sky subtraction are available. HiPSgen users are expected to process their images with their tools-of-choice before submitting to HiPSgen. By making this choice, HiPSgen offers less advanced treatment of the data but is very fast and offers a satisfying global visualization of a data set in a easy way.

Given the preservation of the pixel dynamic in HiPS, usage of photometrically based tools are possible. As examples of such usage, Aladin Desktop offers the possibility to display the histogram of HiPS pixel values falling within a given MOC. Another critical usage is the HiPS2FITS tool (Boch et al. 2020), with which one can project a HiPS onto a region of the sky, with scaling and projection defined by any given FITS header.

2.1.   Computation

The computation of HiPS from a data set with HiPSgen can be multi-threaded as it can run on several CPUs. This is preferred over "cluster" treatments (e.g. for cloud computing) which would imply the duplication of original data and then the fusion of the output results.

---

[2]https://www.ivoa.net/documents/Notes/HiPSProg/

One of the advantages of HiPSgen is that it is scalable. This means that there is almost no CPU or RAM limit to its usage. Surveys of any size can be processed, but of course depending on your machine performance, the process might be slow for very large surveys.

The multi-step treatment implies that the workflow can be fractionated (INDEX, FITS TILES, JPEG TILES). It also implies that one can easily continue an interrupted workflow, or merge intermediate results. These capacities are essential for long treatments of large data sets.

Performances depend on the type of surveys (e.g. how much overlay between images, whether images are compressed or not), the machine used (e.g. how much RAM or CPU, disk speed), as well as the HiPSgen parameter chosen (e.g. bitpix, partioning). Yet one can give a few examples of current performances. Building the DES (Dark Energy Survey) HiPS for the g band is a representative example as it is effectively 1 tera-pixel and the original images from this survey are ready to build (no compression) and have little overlays. Using a linux machine with 14 cores Intel E7 with 50 threads and 100GB RAM, it took 10 hours to project the DES g band 1 tera-pixel. Another example that speaks for itself is given in Boch & Fernique (2019), where they describe the processing of the PanSTARR HiPS. For 1 band, i.e. $7 \times 10^{1}2$ pixels covering 3/4 of the sky, it took 20 days (128 GB RAM, 32 hyper-threaded cores).

3. The resulting HiPS

The result of HiPSgen is a HiPS, i.e. a set of tiles ordered at all scales. Tiles can be in the different formats offered in the standard (FITS, JPEG or PNG). The resulting HiPS can be visualized directly in Aladin Desktop but also by an increasing numbers of HiPS aware clients. Once the HiPS is built with HiPSgen, it can be published via the HiPS network and become accessible to users.

Currently, HiPSgen has produced more than 800 000 billion pixels, including HiPS of a wide variety of surveys like DSS, WISE, PanSTARRS, DES, FERMI, Planck and many others. Different examples chosen to showcase the wide variety of coverage, resolution or wavelength are displayed in Figure 1. The full list of the currently existing HiPS on the HiPS network is available through the HiPS list aggregator[3].

4. Conclusion

HiPSgen is a powerful tool that enables the easy building of Hierarchical Progressive Surveys out of a set of individual images. Some of the assets of HiPSgen are its versatility, its computation speed, the simplicity of usage and the possibility to directly visualize the resulting HiPS with a progressive view in the same application in which it is available (Aladin Desktop). We believe that HiPSgen has played a significant role in the success of the HiPS image standard and will continue to be used extensively for the image surveys to come.
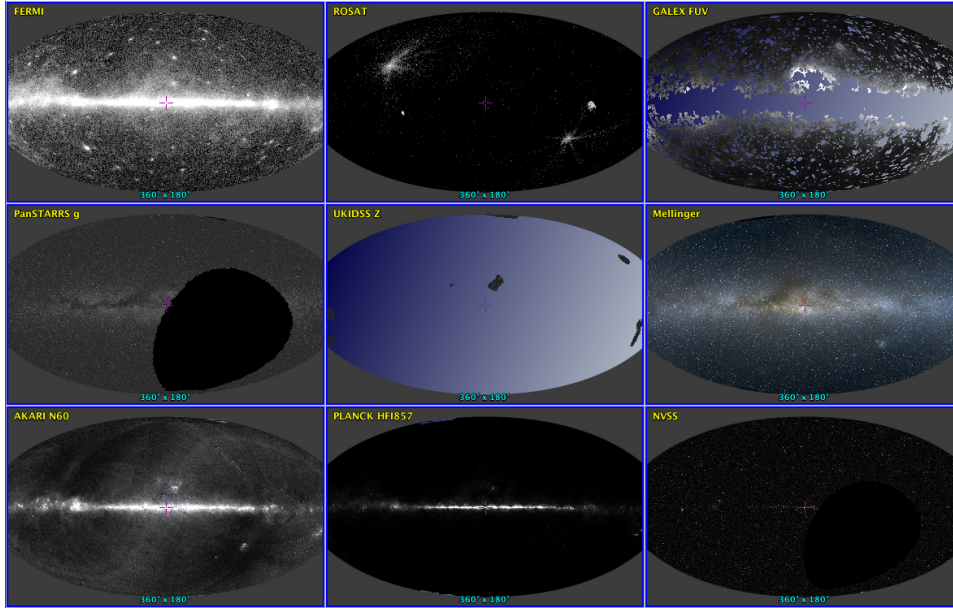
---

[3]`https://aladin.u-strasbg.fr/hips/list`

Figure 1.   Examples of different HiPS generated with HiPSgen and visualized with Aladin Desktop. These HiPS were selected as examples of the wide wavelength range and the diversity of resolutions or coverages that can be produced.

References

Boch, T., & Fernique, P. 2019, in Astronomical Data Analysis Software and Systems XXVII, edited by P. J. Teuben, M. W. Pound, B. A. Thomas, & E. M. Warner, vol. 523 of Astronomical Society of the Pacific Conference Series, 421

Boch, T., Fernique, P., Bonnarel, F., et al. 2020, in ADASS XXIX, edited by R. Pizzo, E. Deul, et al. (San Francisco: ASP), vol. 524 of ASP Conf. Ser., 999 TBD

Fernique, P., Allen, M., Boch, T., et al. 2017, HiPS - Hierarchical Progressive Survey Version 1.0, Tech. rep., International Virtual Observatory Alliance (IVOA)

Fernique, P., Allen, M. G., et al. 2015, Astron. and Astrophys., 578, A114