

On the future of the TERAPIX data centre

Henry Joy Mc Cracken, P. Hudelot

June 15, 2017

1 Introduction and background

TERAPIX (“Traitement Élémentaire, Réduction et Analyse des PIXels”) data processing centre was founded at the IAP in 1997 and has been funded by INSU, CEA, PNCG and the IAP. The centre was *labélisé* as a “Service d’Observation” S05. The primary aim of the centre has been to provide assistance for the reduction of data from array camera detectors; at the end of 1990s, individual astronomers had access to neither the necessary software or computing resources to do this. This task has been divided into three activities: developing software, providing assistance in the usage of this software to users at their home institutes, and reducing data at the TERAPIX center on computer hardware installed at the centre. Funding for TERAPIX computer hardware has been provided by the CSA and from individual grants from PI astronomers at the level of 20-30k euros/year.

At the beginning, TERAPIX data was almost exclusively dedicated to processing optical mosaic camera data from the Canada-France-Hawaii telescope (CFHT) such as UH8K, CFH12K and MegaCam. However, in recent years near-infrared detectors such as CFHT’s WIRCAM and the VISTA telescope’s WIRCAM detector have played an increasingly larger role. Throughout the lifetime of the centre, data has come from two principal sources, principal-investigator (PI) driven programs and “Survey” programs. Around 200 PI programs have been processed in total. These programs are typically based on a few nights’ observation at CFHT or similar-sized telescopes, although several PI-driven programs, such as deep field surveys like “WUDS” or COSMOS *u*-band followup amounted to several tens of nights. PI data has not come only from France, but also from other countries. Generally, the processing steps carried out at TERAPIX involve astrometric and photometric calibration of individual images followed by combination of these images to produce final combined images. “Detrending”, the removal of the instrumental

signature from each image, has been traditionally carried out by teams closer to the instruments (Such as the Elixir team at CFHT)

Concerning surveys, TERAPIX processed seven releases of the Canada-France Legacy-Survey (CFHTLS), a medium-sized five-band MEGACAM survey comprising two main components, a $\sim 100 \text{ deg}^2$ “wide” survey and a “deep” survey comprising four fields observed repeatedly. The TERAPIX releases for the CFHTLS are archived at the Canada Astronomy and Data Center (CADC) and, for the last release, at the CDS in Strasbourg.

TERAPIX developed a pipeline for near-infrared data sets, and this pipeline was successfully converted to work with data from ESO’s VIRCAM camera on the VISTA telescope. From 2007, an increasing fraction of time TERAPIX was devoted to processing near-infrared datasets, including the “UltraVISTA” project (one of the public surveys carried out with VIRCAM on VISTA). Three data releases for the UltraVISTA survey have been publicly released by ESO through their “phase 3” system. The UltraVISTA survey finished in 2017, and TERAPIX will deliver one further release, DR4.

Since the start, developing software has been an important aspect of operations at. Many tools developed by Emmanuel Bertin at TERAPIX (he left the centre in 2006) are now widely used in astronomical community. These tools have continued to be developed as part of the “Astromatic” project.

Staffing levels at TERAPIX have been modest. At the peak of CFHTLS operations (2005-2010) there were three astronomers, two engineers and temporary engineering staff. Today (2017) onwards, the center has three engineers (P. Hudelot, Renato, O. Herent) and one astronomer (H. McCracken). Since 2005 the TERAPIX computer cluster has been managed by the computer service at the IAP.

With the selection of the Euclid mission in 2011, P. Hudelot assumed direction of the TERAPIX data center. In 2015, a review of the TERAPIX operation was called by the CSA in order to better define the future of the center, especially in the context of missions like Euclid and the arrival of other data centers. The CSA was particular concerned that a large fraction of TERAPIX resources was devoted to reducing data from surveys like UltraVISTA.

2 Summary of workshop

Following the recommendations of the CSA review, a one day workshop was held at the IAP the 5th of January 2017. Before the workshop, it was announced to the community that production would stop at TERAPIX in 2017 and that no further PI programs would be accepted. TERAPIX has

engaged to deliver the last UltraVISTA data release (DR4) by Autumn 2017. The aim of this workshop was to solicit community input concerning the future direction the TERAPIX service after the end of production. This document summarises the main conclusions from this workshop.

2.1 Lessons learned from TERAPIX (P. Hudelot)

P. Hudelot summarised TERAPIX activities. The delivery of the CFHTLS releases were initially intended to be once every six months but the centre rapidly fell behind this schedule, principally because no automatic pipeline was available and production of the pipeline was very time-consuming.

Since the end of the CFHTLS, the scientific direction of TERAPIX was not clearly defined. In addition, it was not clear what the priorities of the centre should be: for instance, how important should it be to reduce PI data from astronomers outside of France in the context of limited resources?

2.2 Results of the TERAPIX survey (H. Aussel)

H. Aussel presented the results of a community survey. Around 35 people responded to the survey. In general, the results confirmed the conclusions of the CSA report: many people indicated that without TERAPIX they could **continue** to reduce data. It also confirmed that a large number of astronomers continued to work directly on pixel data. It was noted no pipeline is available to reduce WIRCAM data. WIRCAM data reductions will be no longer supported at the CADC.

2.3 Astromatic (Emmanuel Bertin)

The “Astromatic” software suite, developed by Emmanuel Bertin, is heavily used and well cited (this software contains several well-known packages such as **SExtractor**). The project is currently funded at the level of 20 keur / year by other scientific projects which use this software in their pipelines (e.g., the Dark Energy Survey). Almost all code development is done by E. Bertin, although the hope is to move to a more open, community-drive distributed model, like the scheme adopted by the Large Synoptic and Survey Telescope. This could encourage others to contribute more the Astromatic code base. Community involvement is the future to competing successfully with other large open-source software packages. Development, testing and execution of Astromatic software is carried out on small computer cluster maintained separately by E. Bertin.

3 Discussion and conclusions

Some of the principal conclusions from the meeting:

- TERAPIX is no longer competitive compared to other processing centres. The lack of a fully automatic pipeline means that processing data is heavily labour-intensive, and the number of surveys which can be processed is small. Consequently the impact of the centre in the community is minimal. It was noted that today most of the large MEGACAM CFHT proposals (even those led by French PIs) are processed at the CADC.
- It was noted by that no publicly-available and maintained pipeline exists to reduce WIRCAM or VIRCAM data (the CADC data center is ending support for their pipeline in the coming year). This means with the closure of TERAPIX there is now no way to reduce medium-to-large sized WIRCAM programs.
- Given this, it is important to preserve expertise gained throughout the project in terms of data processing and software. At the very minimum the UltraVISTA / VIRCAM pipeline should be documented and made available to everyone.
- It is not at all clear what could be the role of a TERAPIX-like data center which does not reduce data, or if such a service, without the access to data, could be relevant.
- Within the current structure, engineering manpower available to continue TERAPIX at IAP is extremely limited. As the launch date of Euclid approaches, this is expected to decline further. This poses a real problem to maintain the existing cluster of machines and also to carry out the functions of the service.

4 A future path?

It is clear that TERAPIX cannot continue in its present form. Given the context of limited personnel resources we have searched for a solution which would allow at least the expertise gained in ground-based data reduction and catalogue extraction to be maintained. These skills are still needed: successful exploitation several new datasets in the Euclid deep fields and calibration fields (such as the UltraVISTA extension survey) are heavily dependent on skills gained at TERAPIX.

Therefore, we have identified two engineers and one astronomer at IAP who would be able to maintain a subset of machines and participate in the reduction of these forthcoming data. These machines will be augmented with a grant of 40 keur from the DIM-ACAV and we will ask for other partners in the UltraVISTA project for hardware contributions. Our current plan is to fulfil the engagements we have at TERAPIX (notably the UltraVISTA DR4 release) and once this is completed, transition to the new structure. This will happen at the end of 2017, after the end of TERAPIX.

This plan will assure French participation in the extension to the UltraVISTA survey and the reduction of Spitzer data over the Euclid deep fields. Both these data sets have great scientific value and strengthen French participation in several key international projects (notably, JWST and Euclid). Furthermore, we expect both these data sets to lead eventually to publicly-released catalogues which will be highly cited. This new project would take place in the context of preparation of external data sets for Euclid (in particular on the deep survey fields), an approach which has been agreed with the director of the IAP and the PI of Euclid.

This new structure will be managed by H. J. McCracken with assistance from P. Hudelot. We plan to search for additional resources (students, postdocs) in order maintain production in the context of this new structure, as well as to assure scientific production. We would like to assess how well this new project functions for at least a year before deciding if will apply for *labellisation*.

A key question is what should happen with software packages developed in the context of the TERAPIX project. Our plan is that the staff currently in place will participate in the DR4 release and that during this time we will automate and as much as possible our pipelines. Our short-term objective will be documentation and code refactoring and attempting to deliver a pipeline which can be used with minimum amount of intervention during the production phase for the UltraVISTA extension and the Spitzer data (from 2017 onwards).

Given available resources , it is unlikely that the new structure will be able to offer all the services provided by TERAPIX. However, we will at least maintain the ability to reduce large quantities of pixel data, and maintain the expertise gained in the context of the TERAPIX project.