End-to-end scientific processing in the LIneA Science Portal.

Luiz da Costa[1,2], Angelo Fausti Neto[1], Ricardo Ogando[1,2], Patricia Egeland[1], Marcio Maia[1,2], Eduardo del Peloso[1], Carlos Adean de Souza[1], Valter Machado[1], Cristiano Singulani[1]

[1]Laboratório Interinstitucional de e-Astronomia
Rua General José Cristino, 77 – CEP: 20921-400 – Rio de Janeiro – RJ – Brazil

[2]Observatório Nacional – Ministério de Ciência, Tecnologia e Inovação
Rua General José Cristino, 77 – CEP: 20921-400 – Rio de Janeiro – RJ – Brazil

{ldacosta,angelofausti,ogando,pbegeland,maia,peloso,carlosadean,valter,
singulani}@linea.gov.br

Abstract. *Dark Energy is one of the greatest mysteries of our time. To better understand that, the Dark Energy Survey (DES) has begun to observe in the end of 2012 a large volume of the universe that will generate a catalog of 400 million galaxies up to redshift ~ 1.4, being a unique sample for many years. Its flow from the telescope to a scientific article must follow a well established path that includes quality assurance, pruning and data sweeping for efficient analysis. This is specially true for DES key projects, when products and results must be vetted by the collaboration before going public. Here we describe the effort to create such an end-to-end system in the LIneA Science Portal from catalogs produced by National Center for Supercomputing Applications (NCSA) for DES Collaboration.*

1. Introduction

The acceleration of the universe is one of the greatest mysteries of our time. To better understand that, the Dark Energy Survey (DES, Flaugher et al. 2012) has begun to observe in the end of 2012 a large volume of the universe that will generate a catalog of 400 million galaxies up to redshift ~ 1.4, being a unique sample. Regardless of the science produced with this data set, its flow from the telescope to a scientific article must follow a well established path that includes quality assurance and data sweeping for a efficient run of a particular process that will add value to the catalog. This is specially true for DES key projects, when products and results must be vetted by the collaboration before going public. On BreSci 2012 (Ogando et al. 2012) we showed how the LIneA Science Portal was used to fine tune and calibrate scientific processes by running different experiment lines allowing to get to optimal final results. Here we describe the effort to create such an end-to-end system in the Science Portal from coadd catalogs produced by National Center for Supercomputing Applications (NCSA) for DES to these final scientific products.

2. Material and Methods

Observed tiles on a same region and filter (from the five available: grizY) are combined to increase the depth of the images and consequently the magnitude limit of the sources in the so-called coadd catalog. This is a multi-band catalog, hence each object parameter is multiplied by five. In fact, for DES Science Verification data, the raw catalog has more than 500 columns and 30 millions rows which are transferred to LIneA data center in a few minutes, thanks to an interplay between NCSA and RNP, and ingested using the German

Astrophysical Virtual Observatory (GAVO) Data Center Helper Suite (DaCHS) infrastructure. DACHS allows one to setup a Virtual Observatory compliant data center using XML resource descriptors that hold information to create tables, as well as perform other database operations, following common semantics that allows easy comparison between data sets (e.g. 2MASS (Skrutskie et al. 2006)).

Prior to scientific processing the catalogs go through a Quality Assessment (QA) pipeline that peruses for each filter the star/galaxy classification, magnitude limits, photometric calibration, image quality, and astrometry, comparing it to previous surveys that have similar or superior depth and quality (but obviously not the incomparable DES area*depth) (Figure 1).

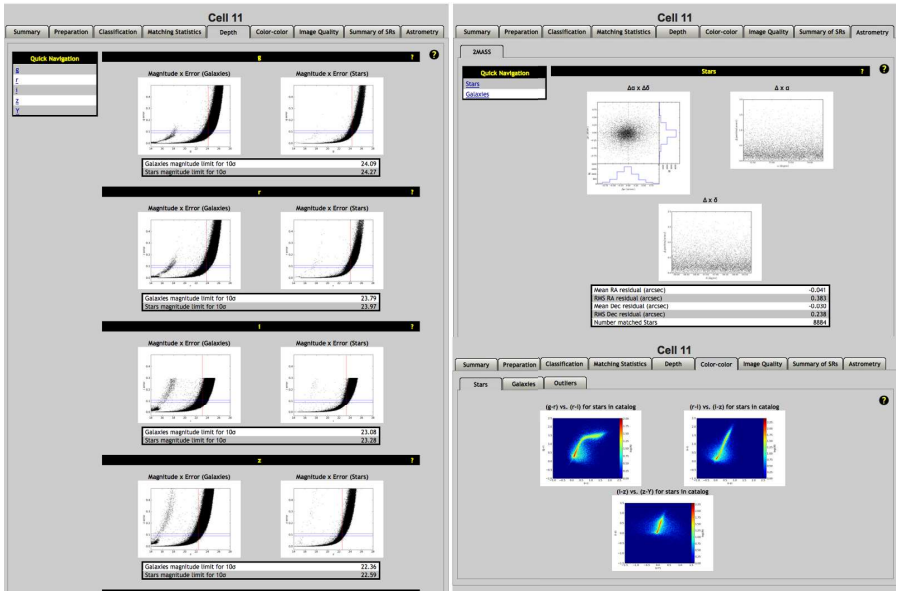

Figure 1: Quality Assessement pipeline screens for photometry depth (left), astrometry (top right) and colors (bottom right).

A key measurement in astronomy is distance, which in extragalactic astrophysics often translates into a redshift. Several methods to measure photometric redshifts are available in the portal allowing the production of the so-called Value Added Catalogs (VAC). In particular, one method that can be readily applied to the data is the template-fitting - in opposition to machine learning technique, which depends on training sets from spectroscopic redshift measurements, which are much more difficult to obtain, specially at higher redshifts. LePhare (Arnouts et al. 2006) is one of the main codes used, providing not only photometric redshifts but also quantities relevant to Galaxy Evolution studies. The analysis itself is also performed in the Science Portal, getting to the point of measuring cosmological constants, closing the loop of the DES end-to-end science processing.

3. Results

With the QA in place, one can make decisions on whether a sample has science quality, and if so, perform assisted selections (e.g. magnitude limit and color cuts) to get to the science end. At this stage, the collaboration is still looking at different ways of vetting the data quality, and for that purpose a number of new Acceptance Tests are being added to the QA pipeline. Anyway, to test the data flow the coadd catalogs have been ran through LePhare pipeline, producing Value-added catalogs, that allows scientific measurements by pipelines already installed in the portal, such as Angular Correlation

Function, Cluster Finding, as well as Galaxy Evolution analysis which is allowed by the LePhare template-fitting.

4. Conclusions and Perspectives

DES is a large collaboration that will analyze a massive amount of data during its 5 years duration. In order to produce reliable scientific deliverables where everyone in the collaboration can quote their provenance, quality and timely analyze them, one needs a framework to assist in such task. The LIneA Science Portal is implementing an end-to-end system that can assess the quality of data produced by NCSA, add value to it allowing scientific analysis, all in a single integrated environment.

References

Arnouts, S et al. (2006) "Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey" Astronomy & Astrophysics, 457, 841

Demleitner, M. "GAVO DaCHS Software Distribution" http://vo.ari.uni-heidelberg.de/soft/dachs

Flaugher, Brenna L.; (2012) "Status of the Dark Energy Survey Camera (DECam) project", Ground-based and Airborne Instrumentation for Astronomy IV. Proceedings of the SPIE, Volume 8446, article id. 844611, 15 pp. (2012).

Ogando, R. L. C. et al. (2012) "A web-based e-Science Analysis Center for large-collaboration astrophysics analysis. Application to the Dark Energy Survey" BreSci 2012, CSBC

Skrutskie, M. F. et al. (2006) "The Two Micron All Sky Survey (2MASS)" The Astronomical Journal 131 1163