



Observatorio Astrofísico de Javalambre

Ingestion of Data in the J-PAS Catalogue

Name: UPAD-PRO-OAJ-9400-11

Version: 1

Date: 5 September,2013

Prepared by	Alessandro Ederoclite	
Reviewed by:	XXX	
Approved by:		
Released by:		

CENTRO DE ESTUDIOS DE FÍSICA DEL COSMOS DE ARAGÓN



Doc:	UPAD-PRO-OAJ-9400-11
Version:	1
Page:	2 of 7
Date:	5 September,2013

Change Records

Issue	Date	Section affected	Remarks
1	21 August 2013	All	First draft (AE)



Doc:	UPAD-PRO-OAJ-9400-11
Version:	1
Page:	3 of 7
Date:	5 September,2013

Table of Contents

Change Records	2
Table of Contents	3
1 Introduction	4
1.1 Scope and Purpose	4
2 Images, Catalogues and Hardware	5
2.1 Images	5
2.2 Catalogues	5
2.3 Hardware	5
3 The Process	6
3.1 Getting a new image	6
3.2 Completing a tile	6
3.2.1 The Reference Filter	6
3.2.2 Any Other Filter	6
3.3 Making a Data Release	7



Doc:	UPAD-PRO-OAJ-9400-11
Version:	1
Page:	4 of 7
Date:	5 September,2013

1. Introduction

1.1 Scope and Purpose

I have not understood part of the process that brings us to have the data in the "main" database and, apparently, there is no document explaining it. I, therefore, try to summarize here what I understand of this process.

Here I am focussing on the data-flow for J-PAS. This can, in principle, be generalised to J-PLUS or even to the data which are produced by Excalibur.

This is not intended as an official or complete document for the UPAD but rather trying to clarify ideas. It may, eventually, derive in a "real" document, though.



Doc:	UPAD-PRO-OAJ-9400-11
Version:	1
Page:	5 of 7
Date:	5 September, 2013

2. Images, Catalogues and Hardware

2.1 Images

Images produced by JPCam are fits files (one per CCD) with 16 extensions, corresponding to the 14 amplifiers of each CCD.

The "raw" files are the files "as" they come out of the camera. The "reduced" files are the ones which are processed by the pipeline in the UPAD. A "tile" is a region of the sky (about $30\text{arcmins} \times 30\text{arcmins}$) which is used as "sky unit" in J-PAS. A "tile image" is the combination of all the reduced images which belong to the same tile.

All images are "fits" images and comply with the fits "standard".

2.2 Catalogues

In order to measure the brightness of the sources which are found in J-PAS, we need to perform photometry on our images. The program which has been chosen for this task is SExtractor. SExtractor can be used in two ways: individual mode or double mode. "Individual mode" means that the image is searched for sources given a detection threshold. "Double mode" means that sources are measured using a reference image to locate the sources.

SExtractor will be run on both the individual images and the tile images. Individual images will be analysed in "individual mode" while tile images will be analysed both in "individual mode" and "double mode".

2.3 Hardware

The hardware that we are interested in is the "Unidad de Procesamiento y Archivo de Datos" (UPAD) and the "External Data Access Machine" (EDAM).

The UPAD is basically divided in two parts: a computation part and a storage part. Computation will have the 300 cores with the capability to reduce and analyse the data coming from the observatory. The storage part (2.5 PB) will be used to store the raw images as well as the reduced images and the resulting catalogues.

The EDAM is mostly a web-server, a "database machine" and some storage to distribute images "on demand".

3. The Process

3.1 Getting a new image

When a new image arrives to the UPAD, the "queue manager" sends it to the "computation machine" where the image is processed (cosmetics, photometry, WCS,...). The image and its catalogue (SExtractor "single mode") are then sent back to the storage. The catalogue is a "fits" produced by SExtractor. It is not included in any database but it is acknowledged as the main resource for variability studies.

3.2 Completing a tile

3.2.1 The Reference Filter

For detection purposes, J-PAS will take advantage of a deep image in r band. When a tile is completed, the files which belong to the tile are loaded into the computation machine and are combined. SExtractor is then run in "individual mode". The produced catalogue provides the "common fields" (e.g. coordinates) to the "double mode" database for each record.

A tricky point (for this and the following section) is how to know when a tile is completed and which images are parts of it. One could either note, for each image, to which tiles the image belongs or, on the other hand, note, for each tile, which images fall in that sky region. This second approach is being considered. Even more interestingly, only knowing which files belong to a sky region is not enough, one also needs to make sure that the tile is covered 100%. This may be done dividing the tile in n ($=100$, TBD) regions which can be used to check the coverage of the tile.

3.2.2 Any Other Filter

When a tile has been completed, all the images belonging to the tile are loaded again into the computation machine and are combined into a "tile image". The "reference image" is also loaded, in order to be able to use SExtractor in double-mode. It is chosen to load the reference image instead of keeping it into the computation machine since it would require a non-negligible amount of space in that machine and since each reference image is only used 58 times anyways.

The tile image is "SExtractor-ed" in both individual and double mode (the "double mode" here is actually "ColorPro") and then moved to the storage together with the catalogues which are produced in this step. Again, the catalogues are, preferably, fits files.

The "double-mode database" will be completed progressively. Some fields (e.g. celestial coordinates of the target) will be filled immediately when the r -band image is taken. The



Doc:	UPAD-PRO-OAJ-9400-11
Version:	1
Page:	7 of 7
Date:	5 September,2013

rest of the fields (e.g. fluxes in each band) will be filled as a new measurement arrives, thus "filling in" the database. One should be particularly careful with duplicate sources (due to the overlap between tiles).

The "single mode database" will be filled by matching sources by position "on the fly" (TBD).

3.3 Making a Data Release

The UPAD is not accessible from outside of ceFca. The solution, to distribute the scientific results of our surveys, is to provide a machine which hosts data which can be accessed.

SMC is going to tell us what "is" a "Data Release" (i.e. is a data release done when a region of sky is completed in all filters? Is a data release done every six months, whatever the coverage?..)

A data release is **not** going to be "copy-and-paste" of the UPAD database to the EDAM. There is going to be some "critical dumping" (i.e. removal and flagging of duplicates).
