

# EPIC, a C++ service oriented inter-process messaging framework and it's usage in PFS Reduction Pipeline

C. Vidal, P.Y Chabaud and LAM scientists  
Laboratoire d'Astrophysique de Marseille

## Abstract

We present EPIC: a C++ library for managing inter process communication in the Prime Focus Spectrograph Data Reduction Pipeline (PFS/DRP). The aim of EPIC is to provide a consistent framework for building distributed software that can be efficiently used in various kind of astronomical data processing pipeline. The Prime Focus Spectrograph multi fiber system will allow more than 2000 simultaneous spectral observations of astronomical targets at the same time, leading to the need of a high throughput system for processing data reduction of each observations. This framework provide a load balanced, messages and services oriented system, where nodes communicate with each other using a set of C++ API. Because of the underlying tools used by EPIC, each service can be run either on a distant computer or an other process on the same computer. This poster shows a basic usage of this framework and quickly describes the protocol used to support inter services communication.

## CeSAM

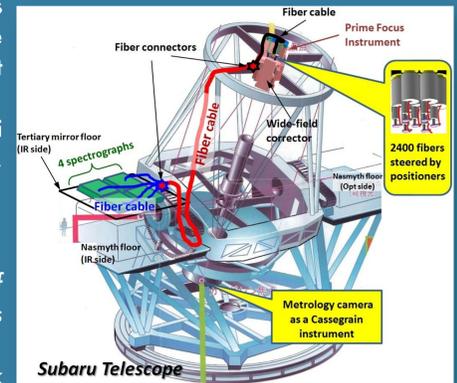


The Astrophysical laboratory of Marseille (LAM) has merged all his scientific computer activity into one entity: The Astrophysical Data Center Of Marseille (CeSAM)  
CeSAM offers software development solutions in several domains including WEB, Data Base Engineering, numerical simulations, image processing and process pipelining.

CeSAM activities spreads over a wide range of projects including COROT, FIREBALL, GALEX, and in development projects such as EUCLID, PLATO, SVOM and PFS

## Prime Focus Spectrograph - Data Reduction Pipeline

The Prime Focus Spectrograph (PFS) is part of the Subaru Measurement of Images And Redshifts (SuMIRe) project. It is one of the main future instruments of the Subaru 8.2 meters telescope at Manau Kea, Hawaii.  
The PFS instrument is an optical/near-infrared multi fiber spectrograph that targets galaxy surveys, and studies of galaxy/AGN evolution.



PFS will provide the unique capability of obtaining 2400 cosmological targets simultaneously.  
The scientific requirements ask for a quick responses and analysis of those targets, thus driving the need of an efficient processing pipeline.

The Data Reduction Pipeline (DRP) is split in several parts, managed by multiples working groups:

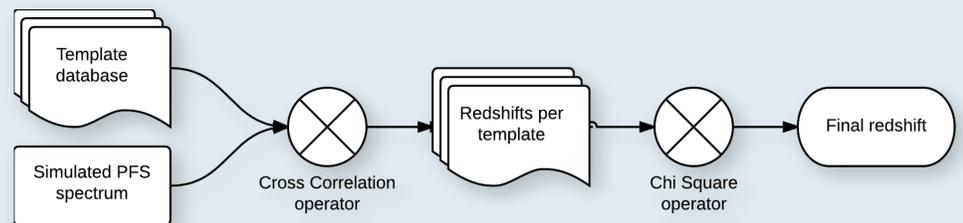
- Raw measurements of spectra's informations are done by the Ipmu institute, Japan.
- 2D reduction pipeline is done by a Princeton's university team.
- 1D reduction pipeline and redshift computation is done by LAM.

## Redshift measurement in PFS :

Because of the large amount of data to be proceeded for the PFS pipeline, fast processing is a crucial need. We provide a *Linux and Mac* tool to compute redshifts value of incoming spectrum. This tool rely on two C++ framework developed at LAM:

- A core framework for basic file I/O and threading systems.
- And a scientific framework for spectrum centric operations (such as: Spectrum loading, Continuum subtraction, Fast extraction of mean and standard deviation, Cross-correlation...)

The capacity of PFS to acquire 2400 targets (and so potentially 2400 spectra) leads the CeSAM to develop it's own pipelining framework: EPIC. Currently in development, EPIC will provide tools to develop high performances pipelines for todays spectroscopic survey.



## Inter Process Communication Framework : EPIC

Today's data processing pipelines, such as the PFS one, require complex assembling of interconnected processes, each responsible of it's own task: Fetching data from distant databases, pre-processing, analyzing and extracting informations. Making them communicate in an uniform way can be a challenging task.

Aware of this issue, and continuing his will to create high quality and reusable softwares, CeSAM has started the EPIC project.

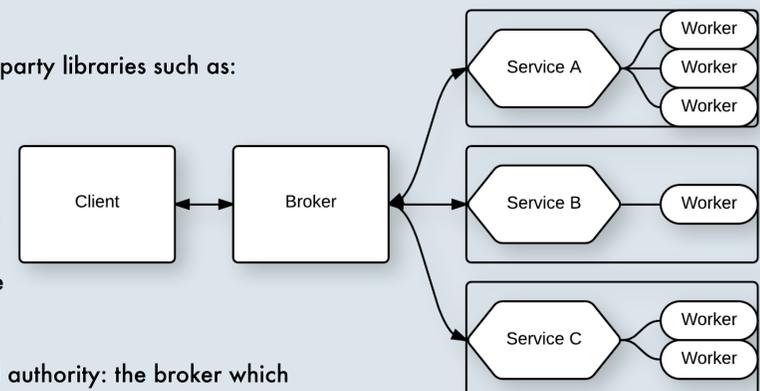
EPIC is a cross platform C++ framework aimed to provide simple, powerful and reliable inter-process messaging services for both networked and non-networked pipelines.

The framework heavily relies on well known and robust open source third party libraries such as:

- ZeroMQ for the inter process and network communication.
- Boost for Threading services, serialization system, I/O operations
- CMake for cross-platform project description.

EPIC is meant to create a service oriented system, where each service can be seen as a stage of a pipeline. Services are provided by one or more Workers allowing each stage of the pipeline to be parallelized.

Messages between client and services are routed via a unique centralized authority: the broker which act as a centralized server for the entire system, providing a unique entry point for controlling and observing state of the system.



## Conclusion and perspectives:

Several improvements needs to be done in order for the EPIC framework to be used by the scientific community:

- Add a python layer on top of it, allowing non C++ developer to use it
- Improve documentation and exemples
- Improve unit testing coverage and robustness

Contacts and informations:

[clement.vidal@lam.fr](mailto:clement.vidal@lam.fr)  
[pierre-yves.chabaud@lam.fr](mailto:pierre-yves.chabaud@lam.fr)  
[christian.surace@lam.fr](mailto:christian.surace@lam.fr)

