Platform for Integration of Data Analisis

Juan B Cabral

UTN-FRC-LIS

FOF & ABRAS Meeting - 2014/APR/24 - IATE - Córdoba, Argentina

About Me

- Software engineer.

- Data scientist.

- Phd. Proposal:

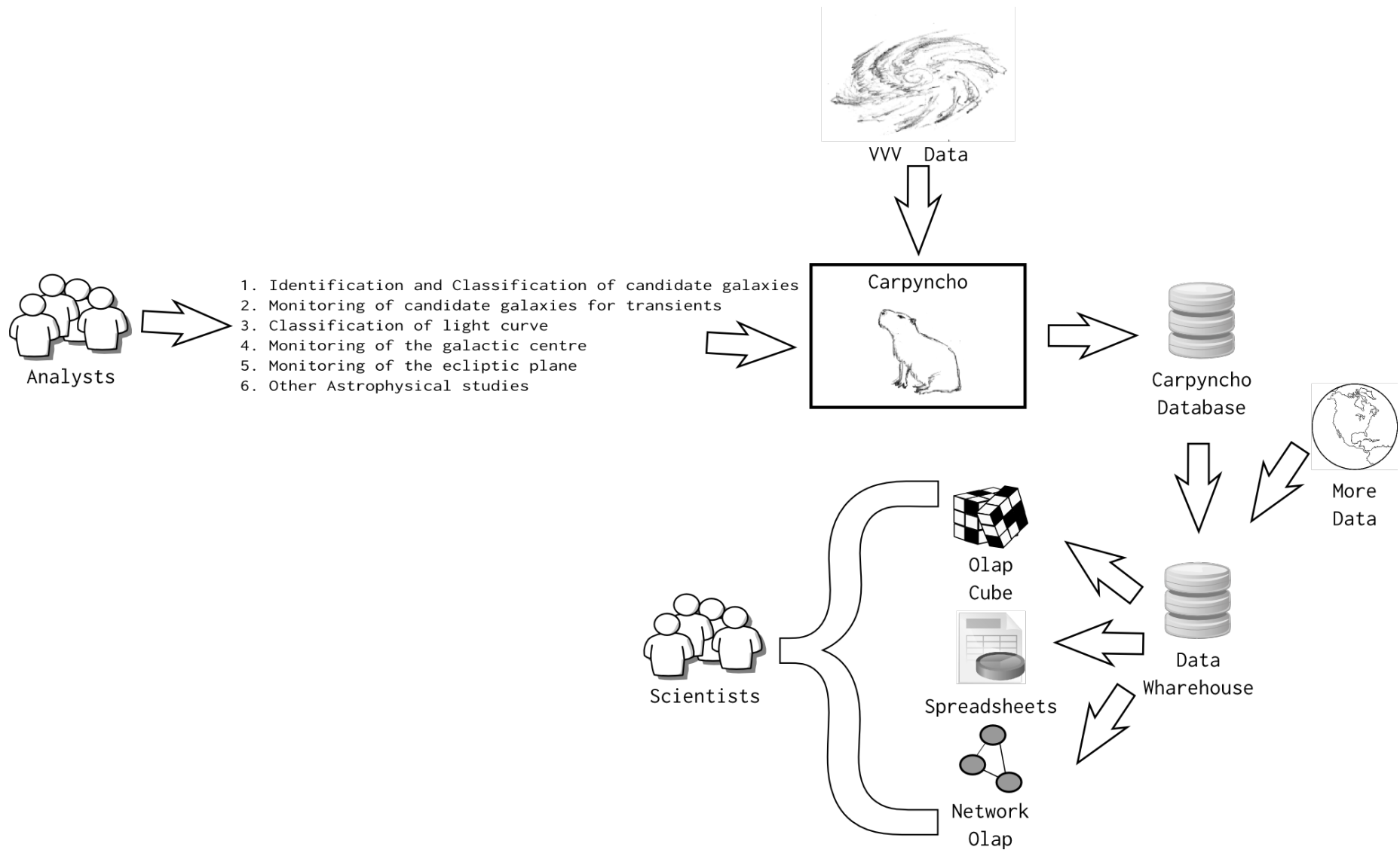Development of Datamining Techniques for Photometric Multiepoch Astronomical Surveys

Directors:

Dr. Pablo Granitto (CIFASIS-CONICET)

Dr. Sebastian Gurovich (IATE-OAC-UNC)

Global Vision of the Project

Citizen Science

The systematic collection and analysis of data; development of technology; testing of natural phenomena; and the dissemination of these activities by researchers on a primarily avocational basis [http://www.openscientist.org].

It's not new

- Amateur astronomy
- Ornithology

Projects:

- Galaxy Zoo (http://www.galaxyzoo.org/)
- Mechanical Turk (https://www.mturk.com)

Carpyncho

- The first phase of this project involves the eyeball detection/classification of galaxies and the monitoring of supernovae within them.
- Since we wish to make eye-ball detections ofgalaxies in the VVV to increase sensitivity, the production of colour composite RGB images is first required so we create colour composite RGB images with the Montage toolkit.
- Fractions

30 arcsec
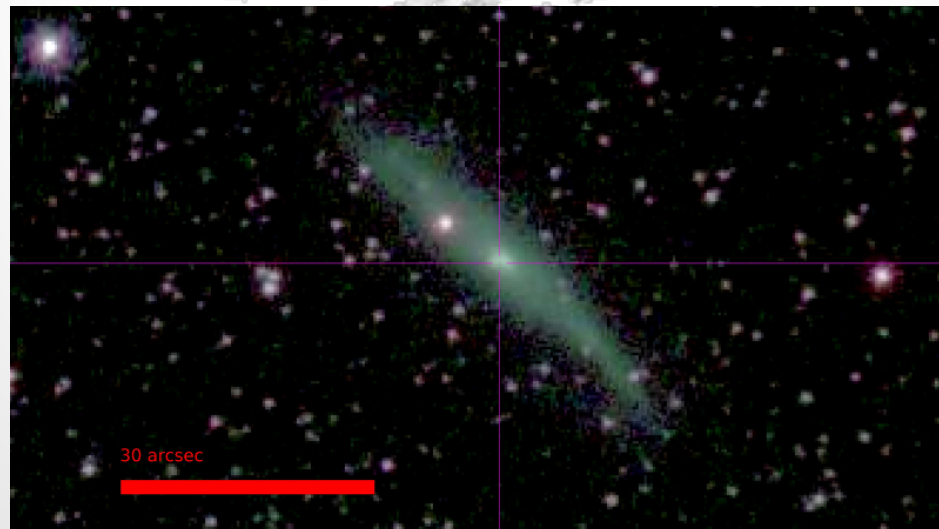
Carpyncho (Galaxy Search)
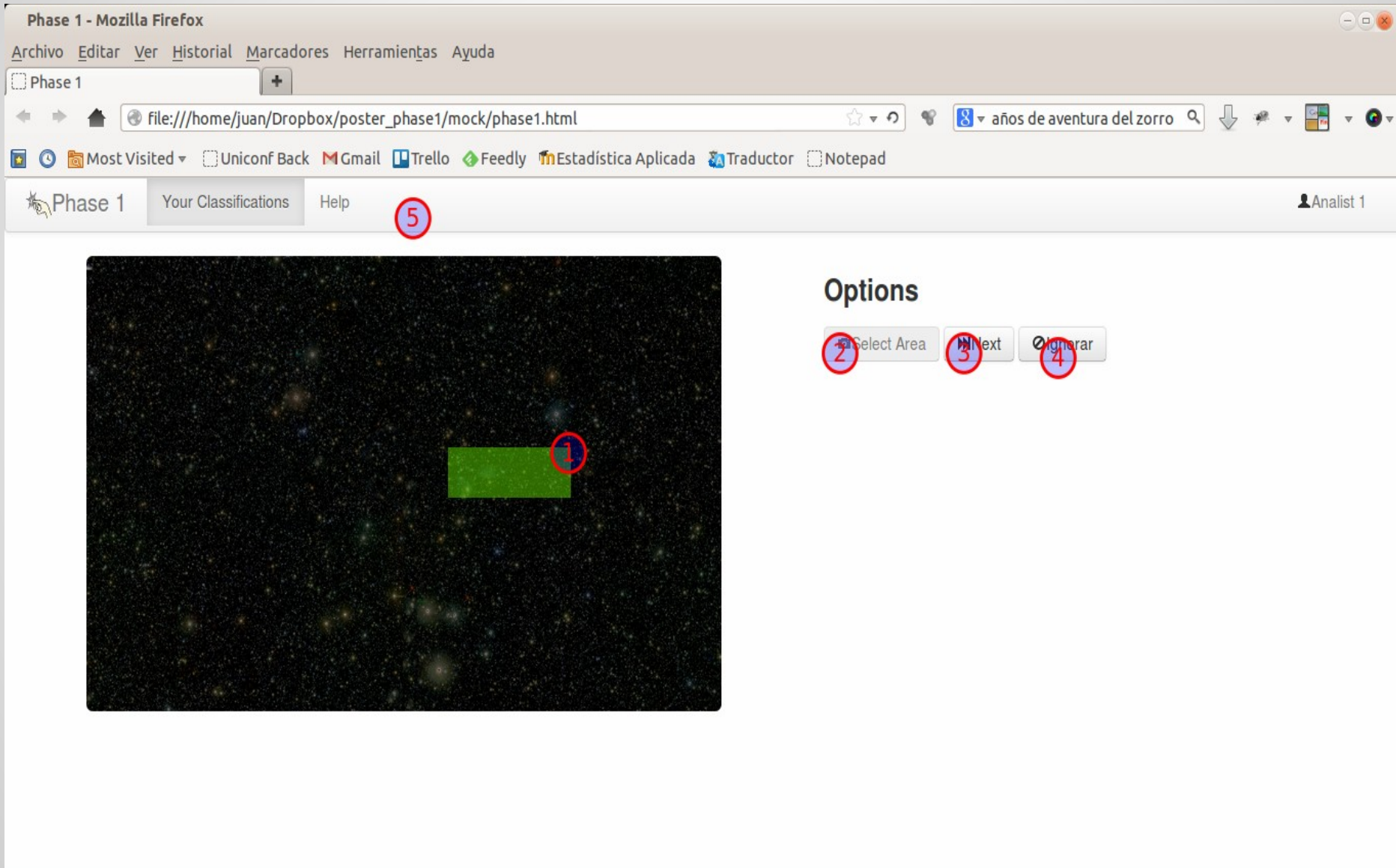
Carpyncho (Future)

- The interface will be adapted to allow the inclusion of movies made for all galaxy 'fractions' in different epochs. Analyzers will be able to click on the frame and any noticeable transient in similar fashion as described in galaxy search.
- When a supernova is detected, the CASU 1.3 pawprint and tile catalogues will be queried at the detected position/epoch and the detection manually verified. We will construct basic light curves from the data at all epochs and will perform time series analysis.
- Our analysis will be used to estimate the type of SNae, its distance, stellar metallicity from the galaxy mass-metallicity relationship, galaxy stellar mass, etc.

KDD, Machine Learning and Data Mining

- **KDD:** The nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [Fayyad, 1996]

- **Data Mining:** Data mining is the application of specific algorithms for extracting patterns from data [Fayyad, 1996]

- **Machine Learning:** Field of study that gives computers the ability to learn without being explicitly programmed [Samuel, 1959]


Figure: An Overview of the Steps That Compose
the KDD Process. [Fayyad, 1996]
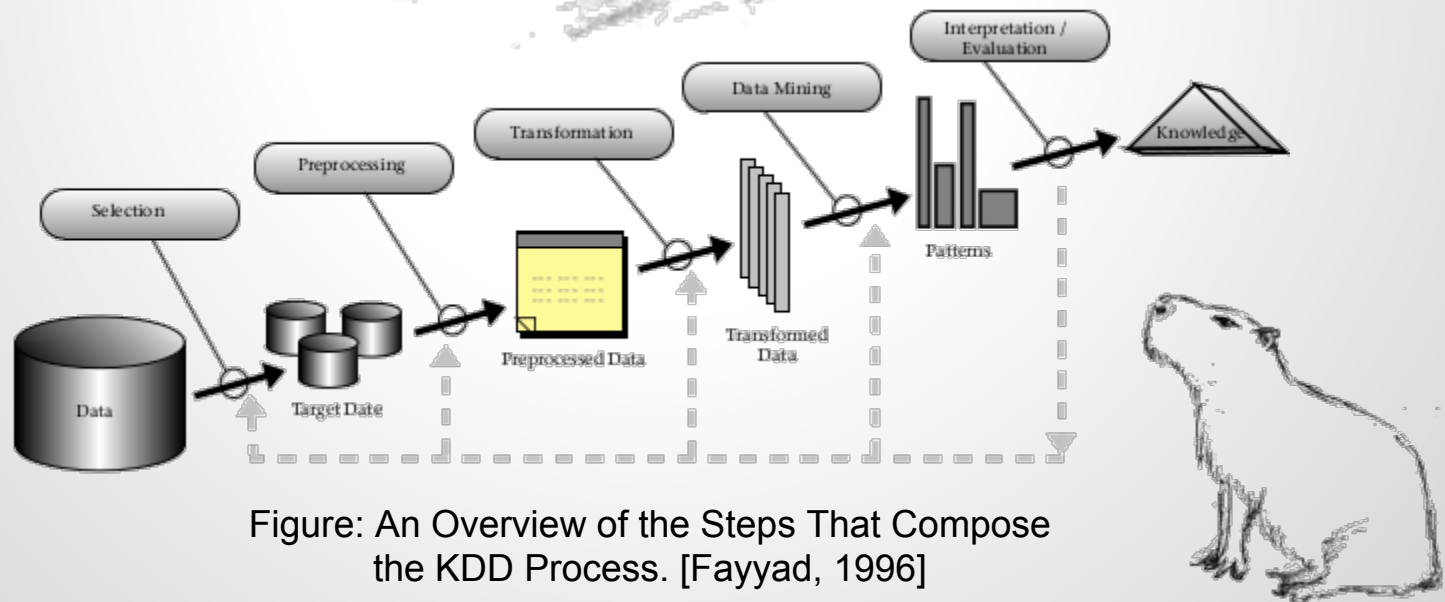
Global Vision of the Project (with concepts)

VVV Data

Carpyncho

Carpyncho Database

More Data

1. Identification and Classification of candidate galaxies
2. Monitoring of candidate galaxies for transients
3. Classification of light curve
4. Monitoring of the galactic centre
5. Monitoring of the ecliptic plane
6. Other Astrophysical studies
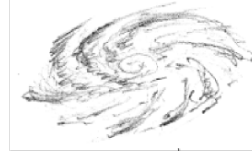
Analysts

Citizen Science

Scientists

Olap Cube

Spreadsheets

Network Olap

Data Wharehouse
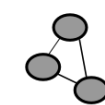
KDD + DM + ML + BI

Business (???) Intelligence

- BI, in simple words, makes interpreting voluminous data friendly. Making use of new opportunities and implementing an effective strategy can provide a competitive market advantage and long-term stability [Rud, 2009]
- Multidimensional data analisis of astrophysical parameter space.

"Friendly"?!?!?

DEMO

Questions?

- **Contact:** jbc.develop@gmail.com
- **Homepage:** http://carpyncho.vvv-tools.com.ar
- **This Slide:**

THANKS!