Expanding Big Data mining for Astronomy

R. Mor and X. Luri on behalf of the Gaia UB team

University of Barcelona ICCUB-IEEC

Data Mining tools have been very important in the last years for the scientific exploitation of astronomical datasets (e.g. Mor et al. 2018; Castro-Ginard et al. 2018; Mor et al. 2019; Romero-Gomez et al. 2019). Data Mining and the Big data infrastructures have allowed us to reach very relevant scientific results. Furthermore, the amount of data that is being generated by space and on ground missions is becoming extremely large and astronomers need very large computational clusters to analyse it. This fact reduces the number of teams able to treat the data as a whole. We are working to widen the Big Data mining available tools to enable the capability of analysis of a larger number of groups. **We expect to make the difference by providing the infrastructure, tools and methods to analyse these large amounts of data data, influencing the future approaches to treat data in space science.**

rmor@fqa.ub.edu

@RogerMor_C

Motivation: we are motivated by the open science concept

- **Open Science pursues a universal science were the results, data and knowledge are not capitalized only by a small group of people.**
- In this direction most of the ESA space mission data are public to ensure a universal accessibility
- However, new times come with large amounts of data (e.g. RVS spectra or CU8 data in Gaia)
  - Even the access to the data is universal the capability of analysing this data is not.
- **One of the challenges for the coming years is to provide Big data mining tools to a wider community boosting the use of large data sets to produce scientific results.**

Ultimate Goal

"The access to the data itself is not enough anymore"

- **We want to offer a Big Data and Data Mining platform through the European Open Science Cloud portal (EOSC) and other local and commercial systems**

    - To **enable** the **analysis** of the data to ensure the universal accessibility

Roadmap to the open science platform

- To develop a Prototype: Our Big data environment, Gaia Data Analytics Framework (GDAF)

- Ambition: To provide a self-deployable template to deploy GDAF in main the commercial cloud services.

- Challenge:To provide a self-deployable environment to deploy GDAF in both the commercial cloud services and local physical environments

SPACIOUS: H2020 project (under evaluation process)

We are in collaboration with: Barcelona Supercomputing Centre, Universidade da Coruña (UDC), University of Edinburgh,CNRS,Port d'Informació Científica (PIC/CIEMAT) and University of Lisbon (UL).

The SPACIOUS project aims to boost the scientific exploitation of ESA mission data by offering a new computational framework for astrophysical research that requires Big Data and Data Mining technologies to be addressed. SPACIOUS will raise the competitiveness of the EU scientific community by increasing its Big Data and Data Mining expertise and capabilities, providing an easier access to the use of these techniques.

The main idea behind SPACIOUS is to facilitate and open the analysis of ESA archives to the scientific community in new ways. SPACIOUS will make the difference by providing unified access to the data and at the same time by providing the infrastructure, tools and methods to analyse these data. **We expect SPACIOUS to become the tool enabling Big Data analysis of ESA data products.** Furthermore, we expect SPACIOUS to **make a change by influencing the future approaches to treat data in space science.**

To take away

- We are working in Big data and Data Mining platforms for astronomy
- In our first steps towards an Open Science platform:
  - We are working to expand our Big Data environment to the Commercial Cloud Services
  - We are in the preliminary stages for an efficient virtualization of our Big Data environment
- **Our ultimate Goal for the future is to offer and Open Science platform through European Open Science Cloud portal (EOSC) and other local and commercial systems**