



ariana.ferreira@sim.ul.pt

# TAKING THE CENSUS OF STAR CLUSTERS IN THE MILKY WAY

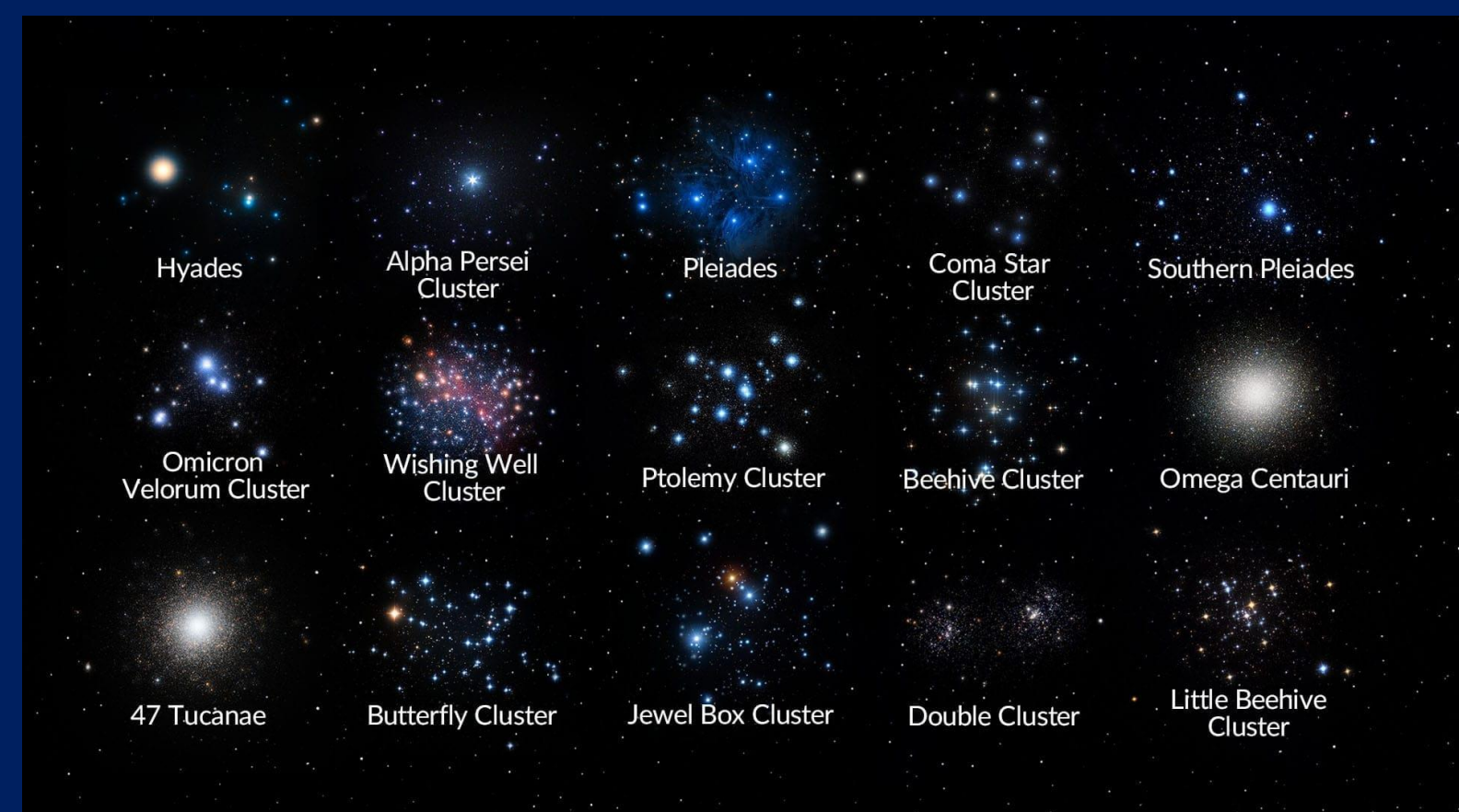
Ariana Dias<sup>1</sup>, André Moitinho<sup>1</sup>, Márcia Barros<sup>2</sup>

<sup>1</sup>CENTRA, Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, <sup>2</sup>LASIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa



## Motivation

- The Gaia mission has led to an explosion of reported OC discoveries driven by the wealth of data, recent developments in machine learning techniques, and the availability of open-source software;
- These reported discoveries are often published in catalogues, but the cross-identification with previously known OCs is not always accurate. This can lead to low-reliability lists with "discoveries" of clusters that were in fact known, a cluster in one list corresponding to more than one in another list, and even duplicated clusters in the same list, and other situations;
- Cross-identification of OCs is not a simple task, as they are often sparse and discrete stellar groups with irregular shapes, different sizes and with no clear boundaries;



- It is clear that the commonly employed method of relying on reported cluster centres and (highly uncertain) radii for cross-identification is not producing high-quality compilations of OCs;
- Nowadays, when hardly a month passes without the publication of a new catalogue of reported discoveries, it has become extremely hard to veto and integrate new findings into a carefully curated compilation of all known OCs.

## Goals

- To create a framework for merging published catalogues into a Master Catalogue of unique OCs;
- Noting that OCs are defined by their members and not by their position and radii, we are developing OC matching algorithms based on statistical comparisons of lists of OC members;
- Generate statistical reports on catalogue comparisons, and which can also be used for tuning the matching criteria;
- Offer this framework as a webservice as well as open-source software.

## Framework

- We start by identifying the publications with the OC catalogues we wish to join. These publications will be identified in the framework by their ADS bibcodes.
- Using the ADS API, we retrieve bibliographic information such as First Author, Title, Year of publication and DOI and put that into the table of catalogues in the Main Database;
- Using the bibcodes and the Vizier API, we retrieve, for each reference, the cluster tables (catalogues) and tables of individual stellar memberships when available. These tables are inserted into the data warehouse in their (heterogeneous) original formats;
- From the data in the warehouse, we create a database with a table of clusters and a table of memberships combining all references;
- All listed memberships are identified with a source id, currently Gaia DR3 source ids. For membership lists using other identifiers (or just coordinates), we crossmatch them into Gaia DR3;
- We define an order for merging the catalogues, for example starting with the oldest as the reference catalogue and merging catalogues in succession by year of publication. In each step, the reference catalogue is the result of the already merged catalogues;
- The merging is done by comparing the memberships sets assigned to clusters in the master catalogue with the catalogue being merged. Several comparison algorithms are being assessed;
- The final product is a master catalogue of OCs and their members, and statistical reports of the matching process.

## Current Status

- We are currently addressing 16 catalogues of OCs and their memberships, making in a total of 32 tables;
- Not all catalogues are available in Vizier. These were retrieved manually from the repositories indicated in the respective articles.
- In total, these tables list over 18 thousand clusters and more than 10 million stars.
- Currently, we are developing algorithms for cross-identification of members. A first naïve approach we have implemented is to simply identify common members in clusters from different catalogues and set a threshold for signaling a positive cross-identification;
- At this stage, we have some preliminary findings:
  - Some catalogues have incorrect or even non-existent source ids;
  - Some catalogues identify their discovered clusters with ordered numbers, no prefix. For those we adopted the prefixes assigned to them in the literature, when available;
  - As a by-product of the work, a list of OC aliases across different catalogues is produced;
  - As expected, we find reported OC discoveries with significant numbers of common members (>60%) with respect to previously known clusters.

Note: The references for the OC catalogues mention above are too many to fit in this poster. If you are interested in knowing them, please contact me.

## Framework Structure

