

Exo-MerCat

a merged
exoplanet catalog

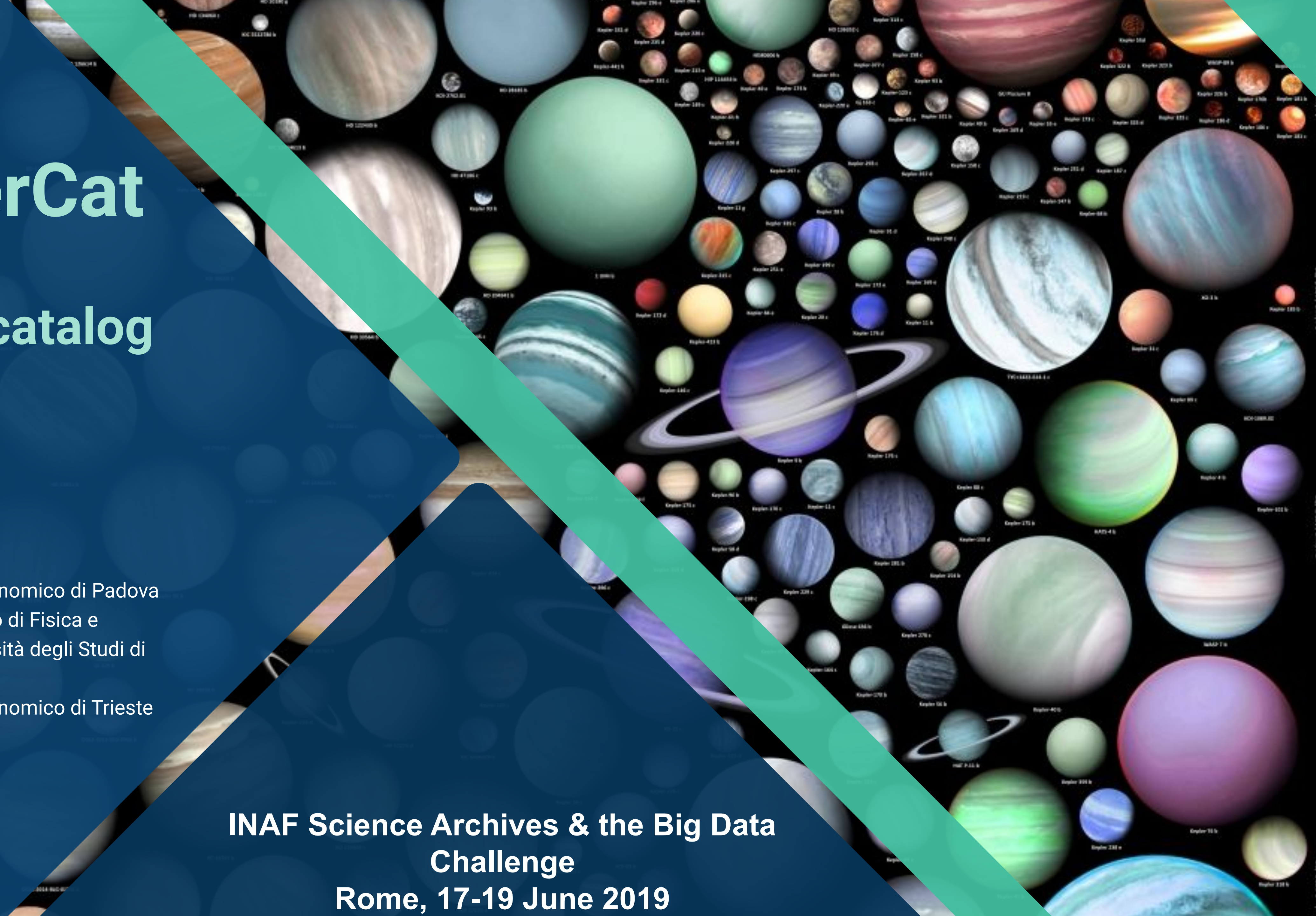
Eleonora Alei^{1,2},
Riccardo Claudi¹,
Andrea Bignamini³,
Marco Molinaro³

¹ INAF – Osservatorio Astronomico di Padova

² DFA-UNIPD – Dipartimento di Fisica e
Astronomia dell'Università degli Studi di
Padova

³ INAF – Osservatorio Astronomico di Trieste

INAF Science Archives & the Big Data
Challenge
Rome, 17-19 June 2019





Overview

- Online exoplanet catalogs: state of the art
- Raw statistics with the current datasets
- Known Issues: updates, errors, selection criteria
- Exo-MerCat: aims, description, efficiency
- Update workflow and VO resource

NASA Exoplanet Archive (NASA)

NASA EXOPLANET ARCHIVE

A SERVICE OF NASA EXOPLANET SCIENCE INSTITUTE

[Home](#)[About Us](#)[Data](#)[Tools](#)[Support](#)[Login](#)

3,949
Confirmed Planets
05/02/2019

11
TESS Confirmed Planets
04/25/2019

564
TESS Project Candidates
04/23/2019

Exoplanets Orbit Database (ORG)

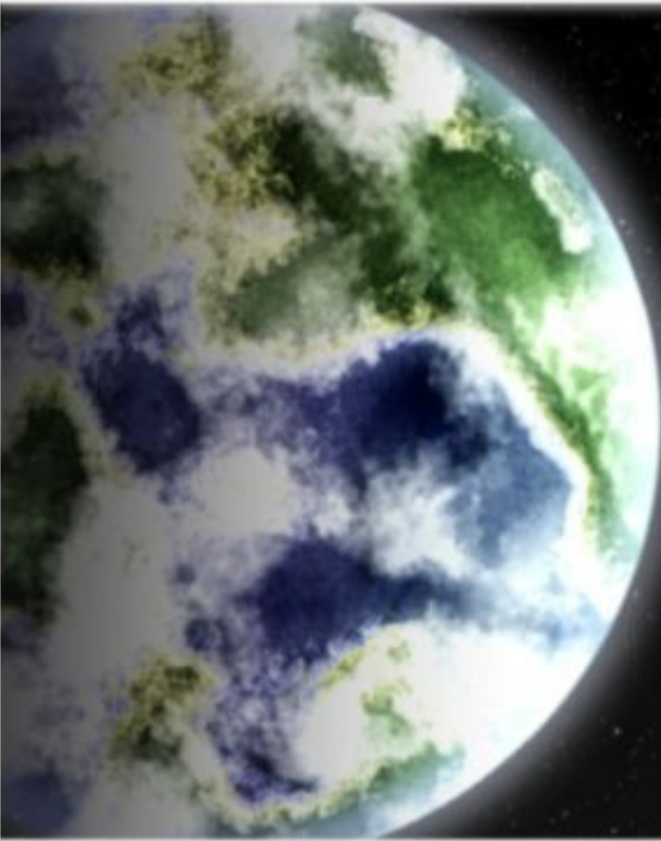
exoplanets.org

Exoplanets Data Explorer

Methodology and FAQ

Exoplanets Links

California Planet Survey



Table

Plots

Search

3237

26

3263

2485

5748

EOD Planets

Other Planets

Total Confirmed Planets

Unconfirmed Kepler Candidates

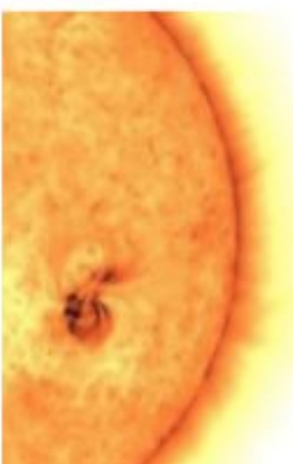
Total Planets

Planets with good orbits listed in the Exoplanet Orbit Database

Including microlensing and imaged planets

Confirmed planets + Kepler Candidates

Open Exoplanet Catalogue (OEC)



Open Exoplanet Catalogue

an open source database of all discovered extrasolar planets

The Open Exoplanet Catalogue is a catalogue of all discovered extra-solar planets. It is a new kind of astronomical database. It is decentralized and completely open. We welcome contributions and corrections from both professional astronomers and the general public.

Catalogue

Plots

All extrasolar planets

Habitable zone planets

Planets in binary systems

Correlations plots

Statistics

Number of confirmed exoplanets	3504
Total number of planets (including Solar System objects and unconfirmed exoplanets)	3791
Number of planetary systems	2657

Number of confirmed exoplanets

3504


Extrasolar Planets Encyclopaedia (EU)

Exoplanet.eu

HomeAll CatalogsDiagramsBibliography

The Extrasolar Planets Encyclopaedia

Established in February 1995
Developed and maintained by the [exoplanet TEAM](#)
update : May 6, 2019 (4065 planets)
Please report any problems to [vo.exoplanet](#)

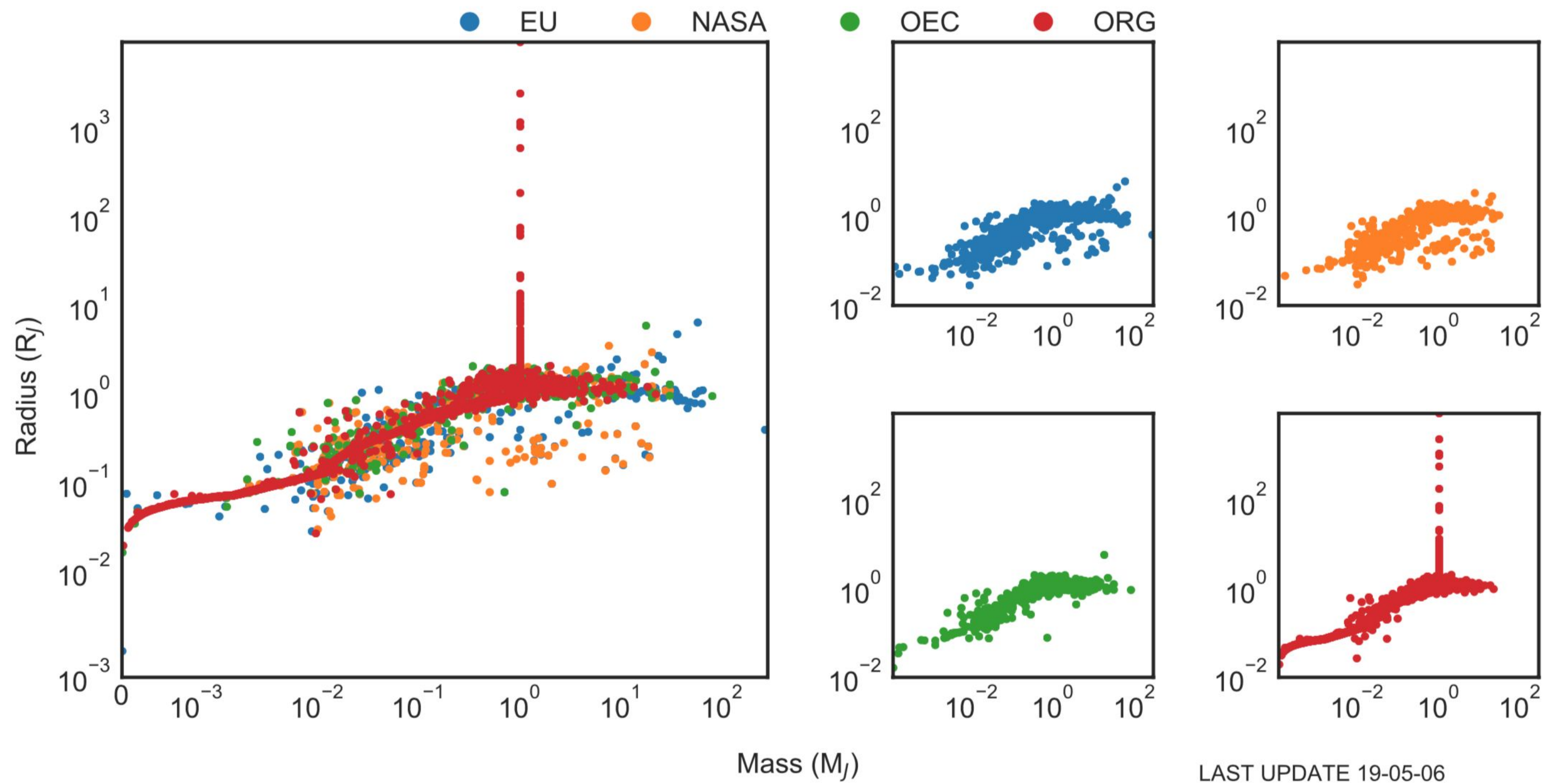


All Catalogs

Filter, sort, export — arbitrary data manipulations with the

update : May 6, 2019 (4065 planets)

Raw statistics



Query	NASA	ORG	OEC	EU
All planets	3949	5748	3793	6698
Confirmed	3949	3237	3674	4022
Candidates	0	2511	108	2670
False Positives	0	0	11	6
With radius	3083	4999	2917	5379
With mass	827	457	1120	1989
With minimum mass	738	29	273	988
With period	3845	5734	3678	6474
With mass or minimum mass	1533	481	1393	1989
With mass and minimum mass	32	5	0	988
With mass and minimum mass and radius	8	2	0	46
With mass or minimum mass and radius	679	420	537	774
With mass or minimum mass, and radius and period	670	420	525	754
All systems	2945	4716	2795	5424

Yikes!

Problems

Selection Criteria

NASA:

- Unique reference;
- $M < 30 M_{Jup}$;
- Peer-reviewed data only.

ORG:

- Robust orbital measurements;
- $M_{pl}/M_* < 0.023$;
- Candidates and confirmed;
- M-R theoretical relations.

EU:

- $M < 90 M_{Jup}$;
- Candidates, announced and published planets.

OEC:

- Open-source, periodically checked by the maintainer.

Aliases

- Names appear in different formats;
- Whitespaces are present;
- Different aliases for the same planet;

Algieba, gamma Leonis:

in NASA: gam 1 Leo
in ORG: gamma Leo A
in EU: gamma 1 Leo
in OEC: Gamma Leonis

Coordinates

- Human errors (plus-minus signs);
- Not updated coordinates;
- Different epochs.

Proxima Centauri b (ra,dec):

in NASA: (217.428995,-62.679485)
in ORG: (217.448946,-62.681353)
in EU: (217.429167,-62.679444)
in OEC: (219.990850,-60.835619)

Updates

- False positives are present in the catalogs because of lags in the updates;
- New candidates have yet to be included in the database.



Aims

- Provide **greater uniformity** among the databases;
- More effective **associations** among the datasets;
- Identify and **correct errors**, to warn the catalog maintainers;
- Provide a direct **link** with most **stellar sources** archives;
- Provide the user with an intuitive **Graphical Interface** to download and filter data.



Description

Initialization

- Create a nested folder to contain all useful files;
- Use various **Virtual Observatory** tools to download raw datasets:
 - `wget` command to access NASA/ORG database;
 - `git` commands and an `*.xml` reader to access the OEC database;
 - VO TAP service for the EU database.





Description

Homogenization

- Selection of specific, useful columns;
- Grouping of stored aliases;
- **Removal of whitespaces** and standardization of name strings, following known notations and conventions;
- The planet name was stripped in Host star name + Letter, and those values stored separately;
- In the end... all four datasets looked very similar!



Description

Status check

- Download the **Kepler-K2 Objects of Interest** list with updated statuses from NASA Archive and Mikulski Archive for Space Telescopes (MAST);
- Compare the various entries and update if necessary the status of each planet (whether CONFIRMED, CANDIDATE, FALSE POSITIVE);
- If confirmed, update names with default ones.



Description

Alias Check

Globally, we expect up to **four occurrences** for the same planet (one per catalog). But a planet could be labeled with an **alternative name** and thus any software which matches strings won't recognize it as the same planet after all.

Therefore:

- All known aliases for the host stars were queried by performing a **VO TAP query** to SIMBAD.
- If one of the aliases for each star is found as a main identifier elsewhere in the databases, the code uniforms all occurrences.



Description

Coordinate Check

- For each host star, retrieval of the **mode** of right ascension and declination in degrees. If one or more values are different from the mode, these are replaced by the mode itself.
- If no mode is found (i.e. there is no most common value), no replacement is made.
- **Warnings** are printed to be sent to the catalog maintainers in order to encourage a check on particular values.



Description

Main ID retrieval

#IDs	Various archives and catalogs are queries by means of VO TAP connections and <code>pyvo</code> Python library.
0%	
95%	1. SIMBAD TAP query for exact match for the host star;
96%	2. SIMBAD TAP query for exact match for every available alias ;
99%	3. SIMBAD TAP query for coordinate match for the host star (tolerance 0.0005 degrees);
	4. VizieR TAP query for coordinate match in Kepler-K2 input catalogs;
100%	5. VizieR TAP query for coordinate match in GAIA DR2 catalog.



Description

Catalog retrieval

- The **global catalog** (concatenation of the four archives) is grouped by `MAIN_ID` and `Letter`.
- For every parameter, this function calculates the **relative error**, in order to choose the most precise dataset for each parameter (and its reference paper).
- A **default name** for the planet is chosen, but all aliases are stored.
- At this point, each group is collapsed in a single line, which may have measurements belonging to different papers and/or different catalogues.

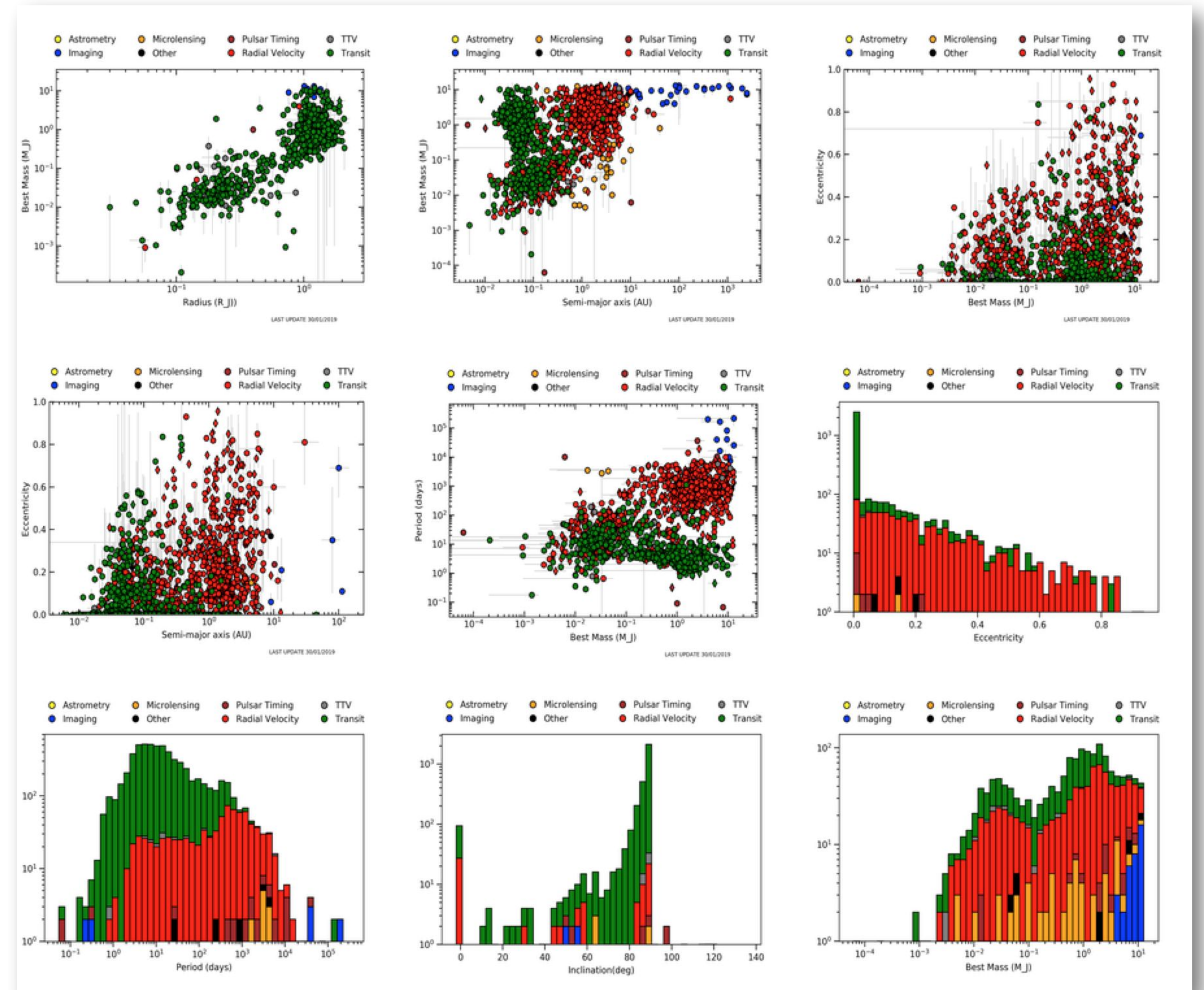


Description

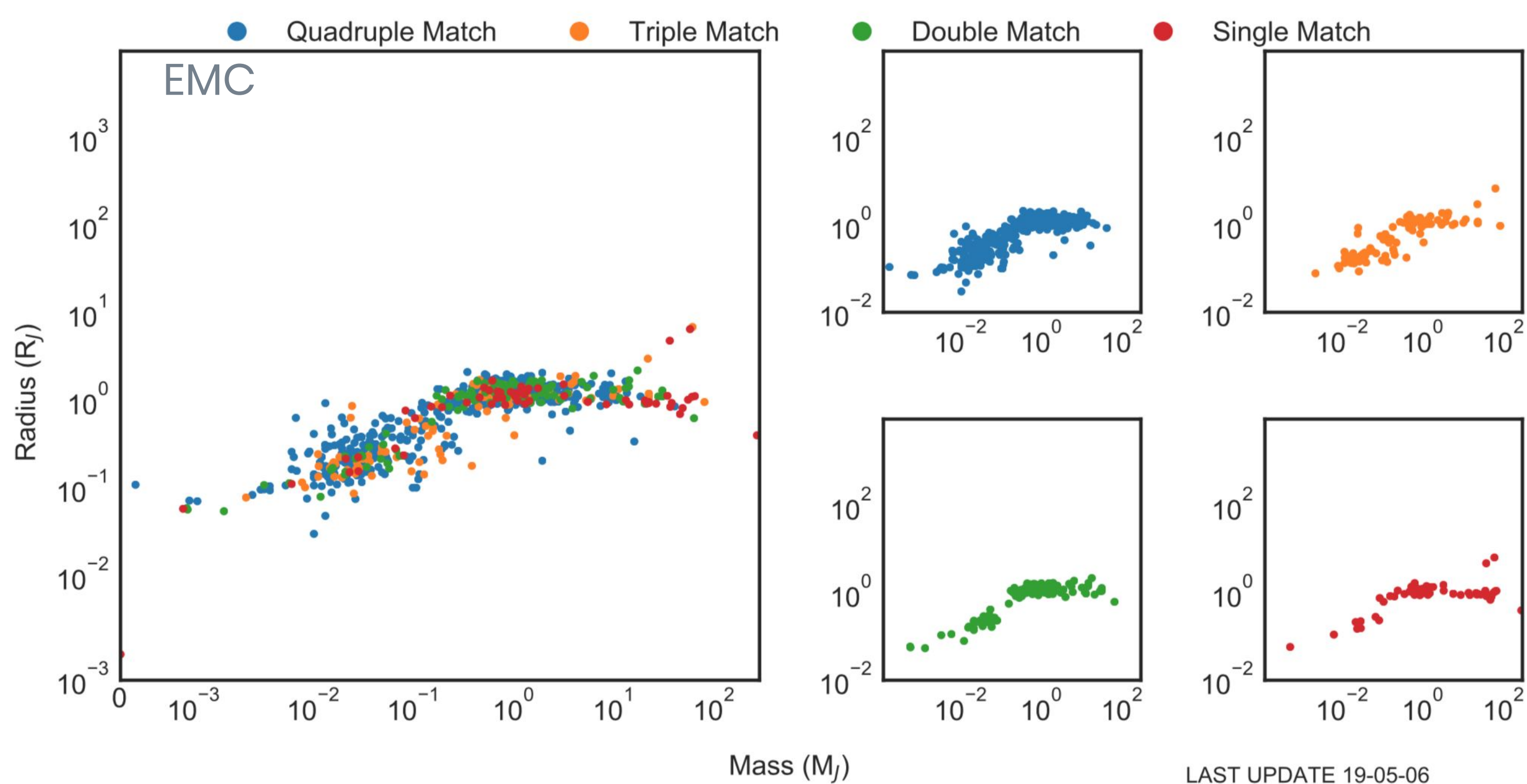
GUI

- An open-source Graphic User Interface is available to **directly download** the MEC and to **filter** data, as well as to make some **plots**.

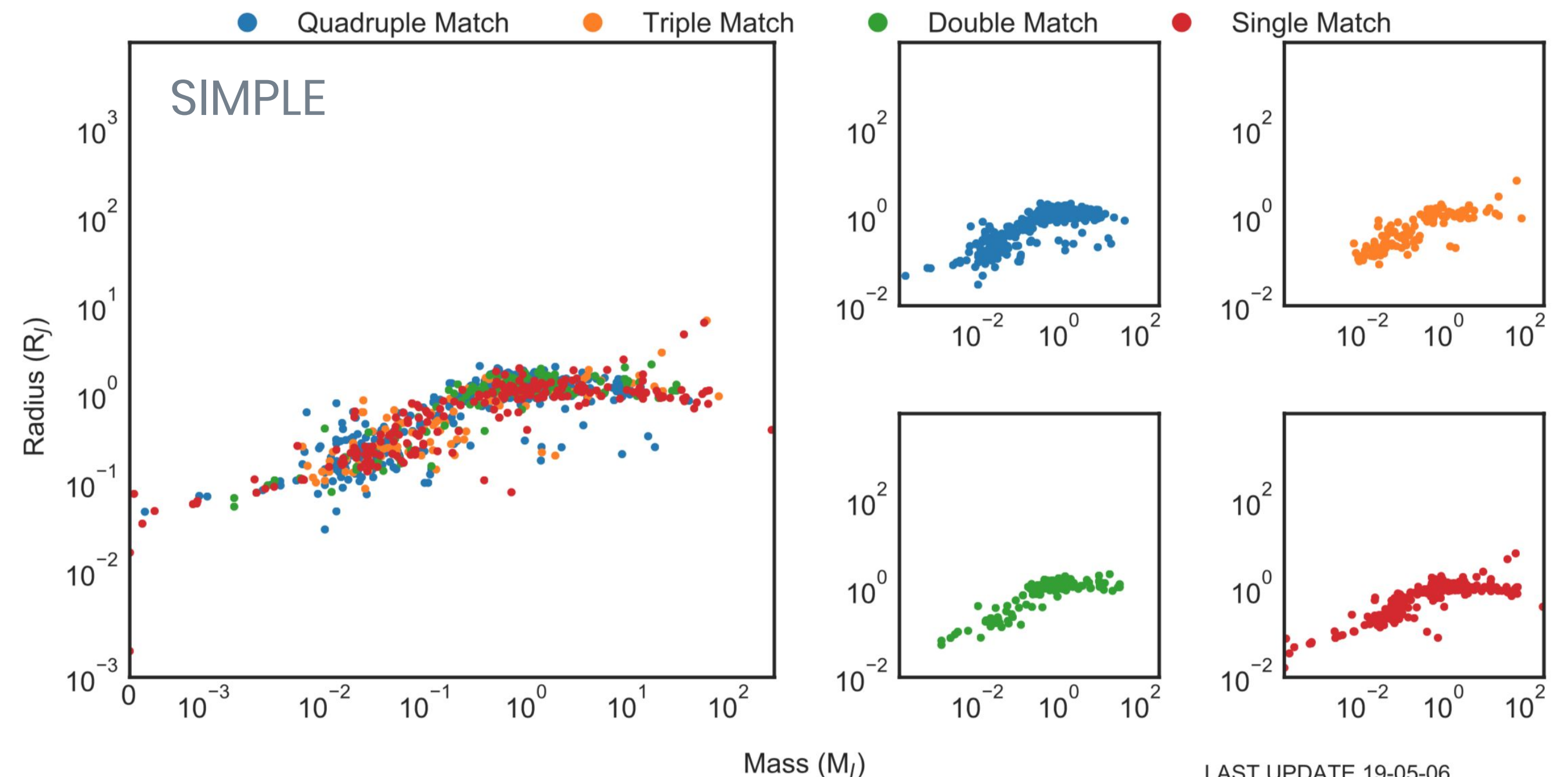
Parameter	MINIMUM	MAXIMUM	Unit	<input type="checkbox"/> only confirmed
Mass	<input type="text" value="Any"/>	<input type="text" value="Any"/>	M_J	<input checked="" type="checkbox"/> Msini <input checked="" type="checkbox"/> Mass
Radius	<input type="text" value="Any"/>	<input type="text" value="Any"/>	R_J	Discovery Method <input checked="" type="checkbox"/> All <div> Radial Velocity Transit Astrometry Imaging Microlensing TTV Pulsar Timing Other </div>
Period	<input type="text" value="Any"/>	<input type="text" value="Any"/>	days	
Semi-major axis	<input type="text" value="Any"/>	<input type="text" value="Any"/>	AU	
Eccentricity	<input type="text" value="Any"/>	<input type="text" value="Any"/>		
Inclination	<input type="text" value="Any"/>	<input type="text" value="Any"/>	degrees	
Folder Name	<input type="text" value="20190509/"/>			
<input type="button" value="Advanced Plot"/>			<input type="button" value="Plot"/>	



Results

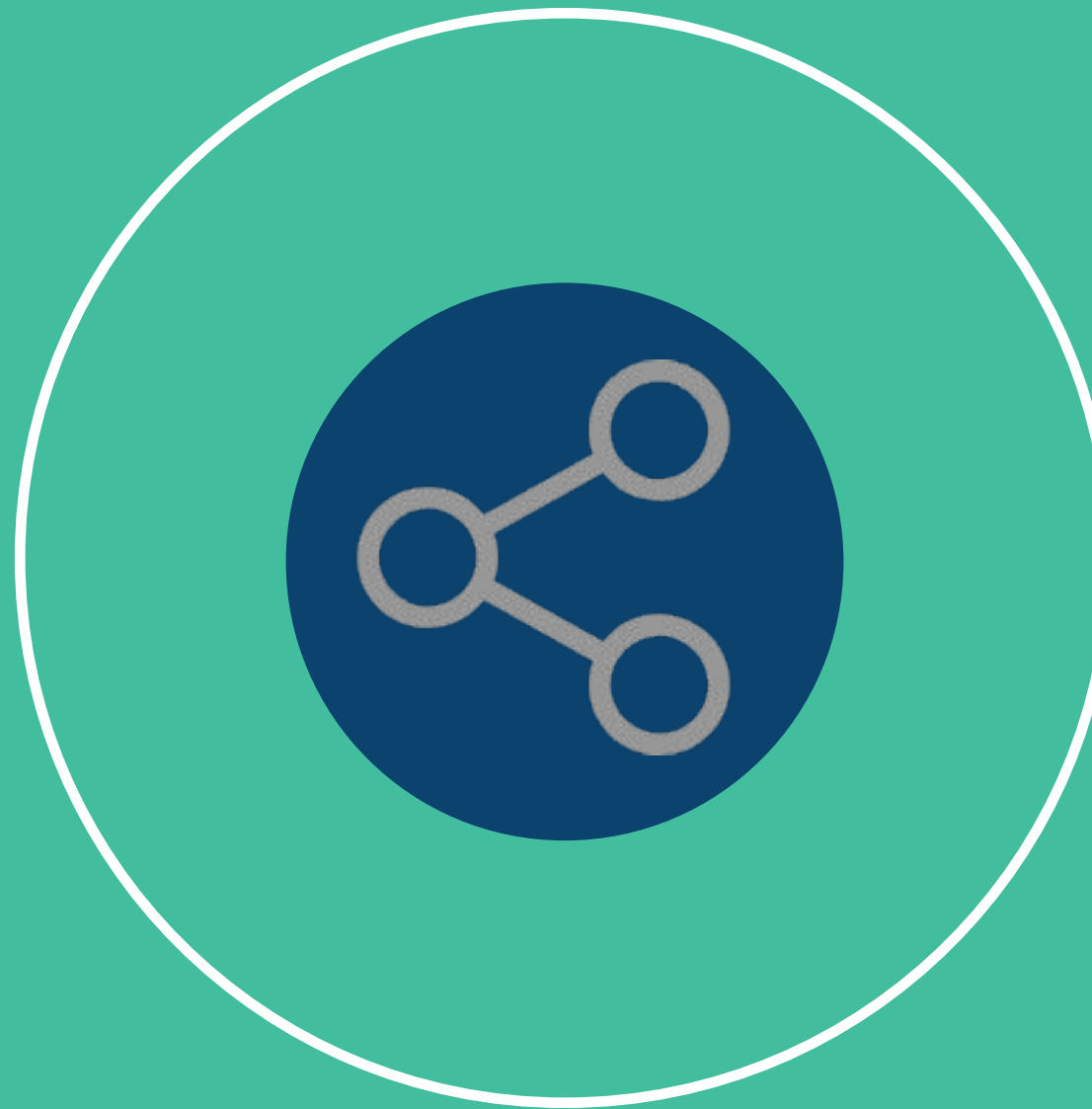


Samples	EMC RUN	SIMPLE RUN
All Planets	7396	10083
All Confirmed/Candidate Planets	6920	10067
Quadruple Matches	3190	2930
CONFIRMED	3129	2902
CANDIDATE	60	28
FALSE POSITIVE	1	0
Triple Matches	359	429
CONFIRMED	299	403
CANDIDATE	56	25
FALSE POSITIVE	4	1
Double Matches	2350	528
CONFIRMED	385	474
CANDIDATE	1960	52
FALSE POSITIVE	5	2
Single Match	1497	6196
CONFIRMED	351	1010
CANDIDATE	680	5173
FALSE POSITIVE	466	13



Update Workflow and VO resource

- Periodic updates (once a week).
- Workflow described via the Common Workflow Language, useful for the versioning of the input files.



- Registered as a VO resource (IVOID: ivo://ia2.inaf.it/catalogues/exomercat)
- The catalog is accessible by all VO-aware TAP-enabled applications (<http://archives.ia2.inaf.it/vo/tap/projects>)



Conclusions

- **Exo-MerCat** aims to standardize, correct and collect the most precise data from all available archives.
- It allows an easy **statistical analysis** of the current sample of exoplanets by reporting the updated status, the source catalogs, and the reference papers for each parameter. A **GUI** is provided to filter data, make easy plots and histograms.
- It is a **VO resource** accessible through VO-aware applications and a direct link to most famous stellar catalogs is provided.
- **To-do list**: possibility to query for one or more versions of the catalog; stellar datasets retrieval.
- But a standardization for exoplanet-related data is due! A new **Data Model** for such data needs to be developed.

Thank you!