

The making of RGZ-EMU

The **Radio Galaxy Zoo for the Evolutionary Map of the Universe Survey (RGZ-EMU; Tang & Vardoulaki et al. in prep.)** is a direct spin off from the very successful RGZ citizen science project. With novel methodologies, i.e. **complexity** and **taxonomy**, and the combined efforts from specialists and citizen scientists we aim to tackle scientific questions about radio sources. **Data** are from the Evolutionary Map of the Universe (EMU) ASKAP [1] Pilot Survey at 944 MHz [2]. It observed 270 deg² of the southern sky, reaching an *rms* of 25–30 μ Jy/beam and a spatial resolution of 11–18 arcsec. The Selavy algorithm [3] was used to identify \sim 220K sources. The data are fed into the zooniverse.org platform (Fig. 1) into workflows we designed (Fig. 2), for the citizen scientists/zooters to access.

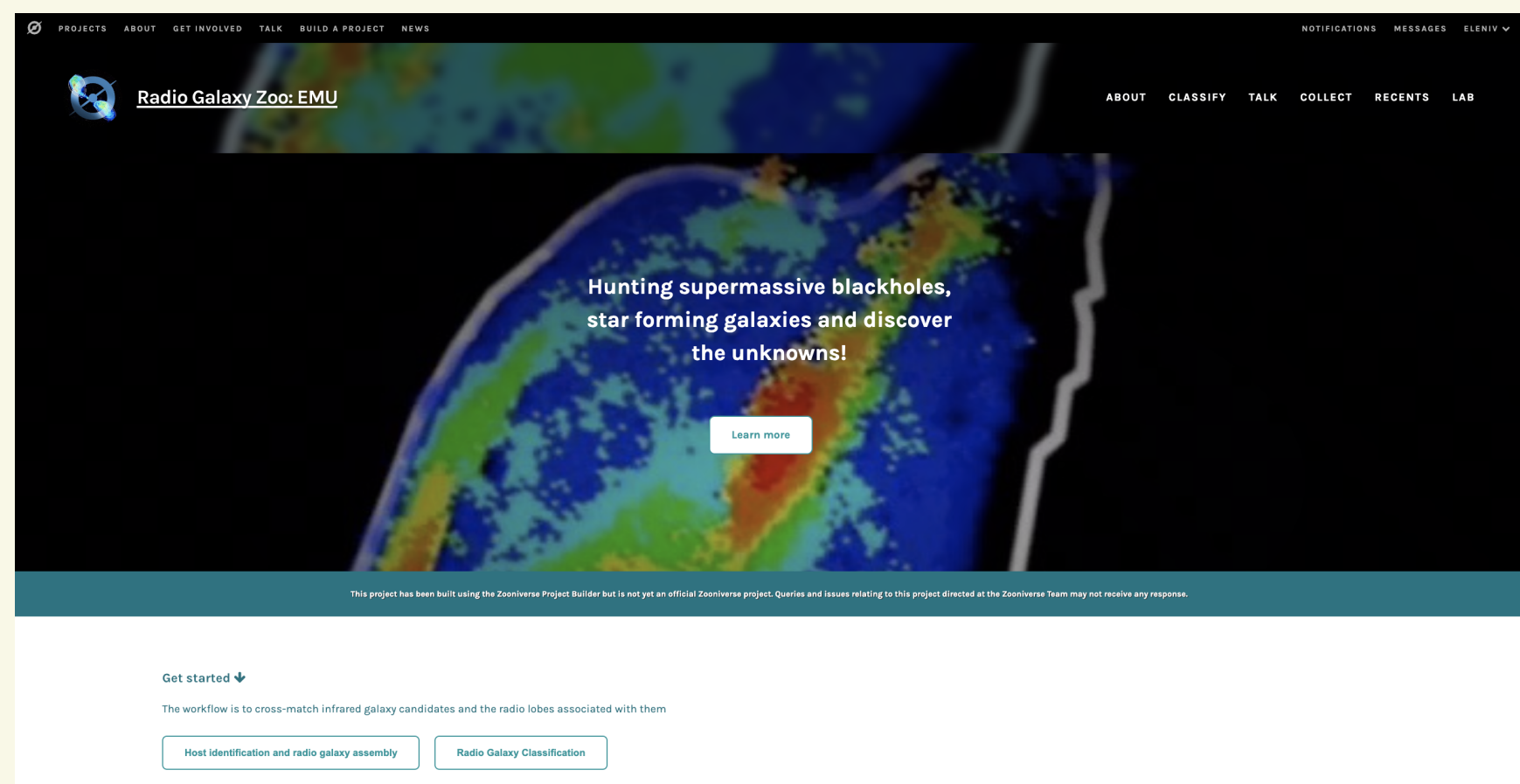


Fig. 1: The RGZ-EMU Homepage: It presents the two workflows: 1) *source identification and assembly*, and 2) *source classification*, as well as details about the project, the team, and the education aspect of RGZ-EMU.

Pre-launch Early Science results

The projects below were developed during internal testing and preparations for launching the project to the public.

- **Semantic Morphology Taxonomy** is derived via natural language processing (Fig. 3–*Left*) and provides a set of tags to the RGZ-EMU workflow (Fig. 2–*Right*; [7]). By using plain English terms, we can recover known scientific as well as rare sources with abnormal radio structures (Bowles et al. in prep.).
- **Self organising maps (SOM; Vardoulaki, Tang, et al. (in prep.))** are produced using the Parallelized rotation and flipping INvariant Kohonen (PINK; [10]) algorithm, with the purpose of radio-source classification and the discovery of peculiar radio sources (Fig. 4).

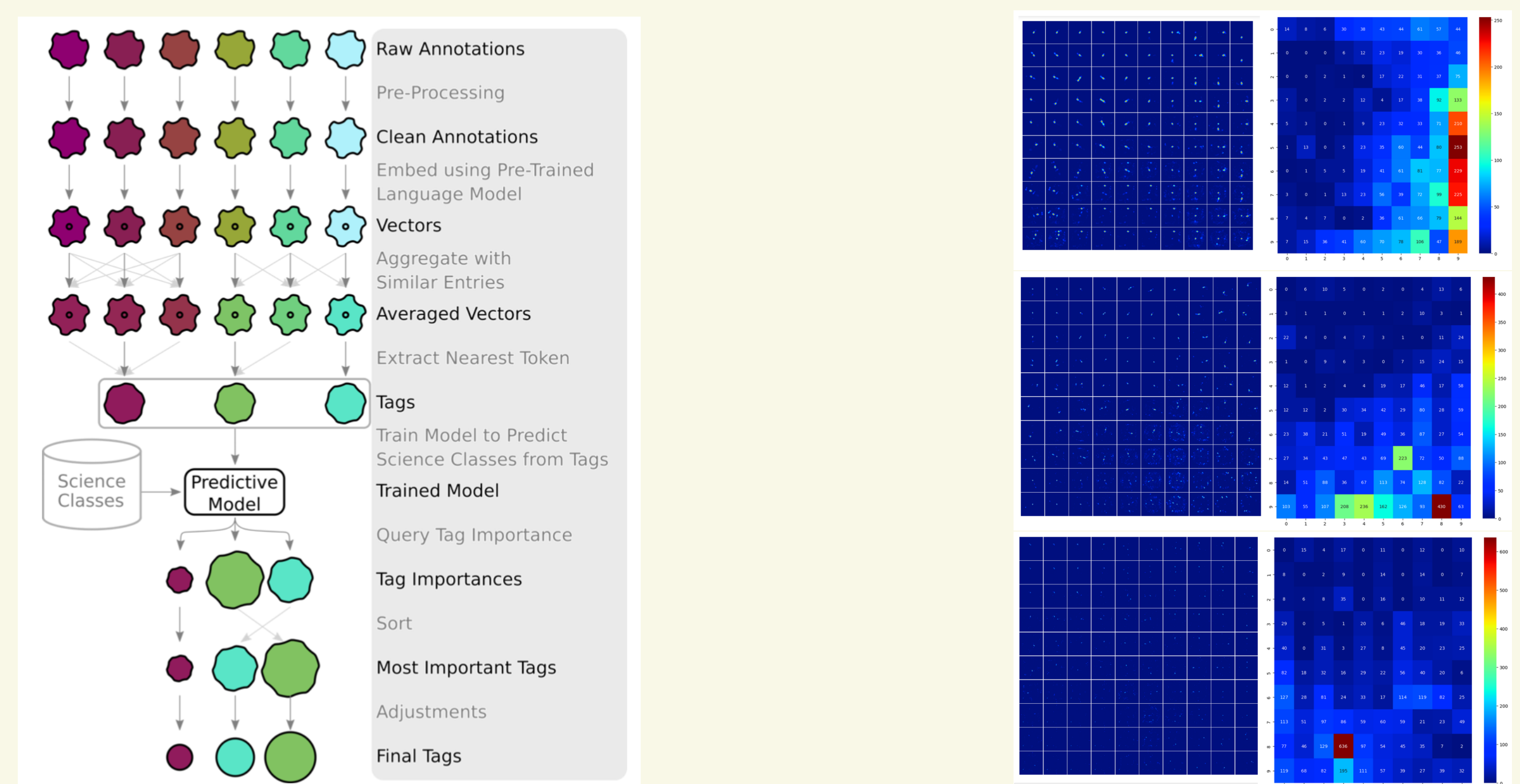


Fig. 3: (Left): Proposed workflow to computationally derive a semantically meaningful plain English taxonomy from a set of annotations (Bowles et al. in prep.). **(Right):** Testing the effect of different cutout sizes ($3' \times 3'$ – top; $6' \times 6'$ – middle; and $12' \times 12'$ – bottom) as input to self organising maps using PINK [?]. The SOM was run by students (Anees, Dinh, Medina, Meyounyo, Mirshanova, Pöppelmann) at the Darmstadt University of Applied Sciences.

The Science behind

RGZ-EMU has three science goals, presented via two workflows in the zoonivers.org platform:

- **Identify the hosts of radio sources and assemble the different radio components** into one parent source using overlays with radio, optical (DES; [4]), and infrared (WISE; [5]) images.
- **Classify radio sources** using descriptive tags [6] that were derived using results from early RGZ-EMU classification experiments and natural language processing [7] .

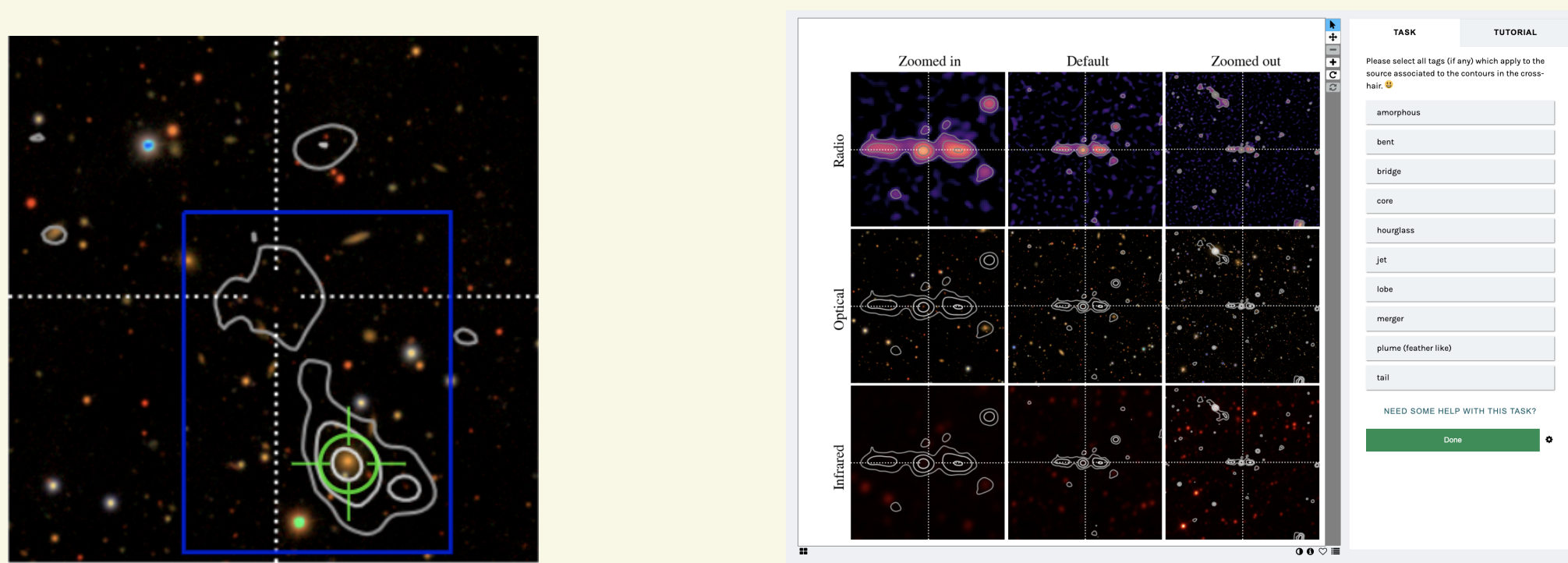


Fig. 2: Examples of the two workflows for the RGZ-EMU project release. **(Left):** Zoom-in of the host identification (green circle) and source assembly (blue box) workflow. **(Right):** Classification workflow based on tags. The setup of the cutouts in three different sizes ($3' \times 3'$, $6' \times 6'$, and $12' \times 12'$) enhances workflow functionality and helps identify 95% of giant radio galaxies.

Hunting the unexpected

Radio surveys are filled with strange radio structures (Fig. 4). To facilitate discoveries, we applied the complexity methodology [8, 9] to provide the most interesting cutouts to view.

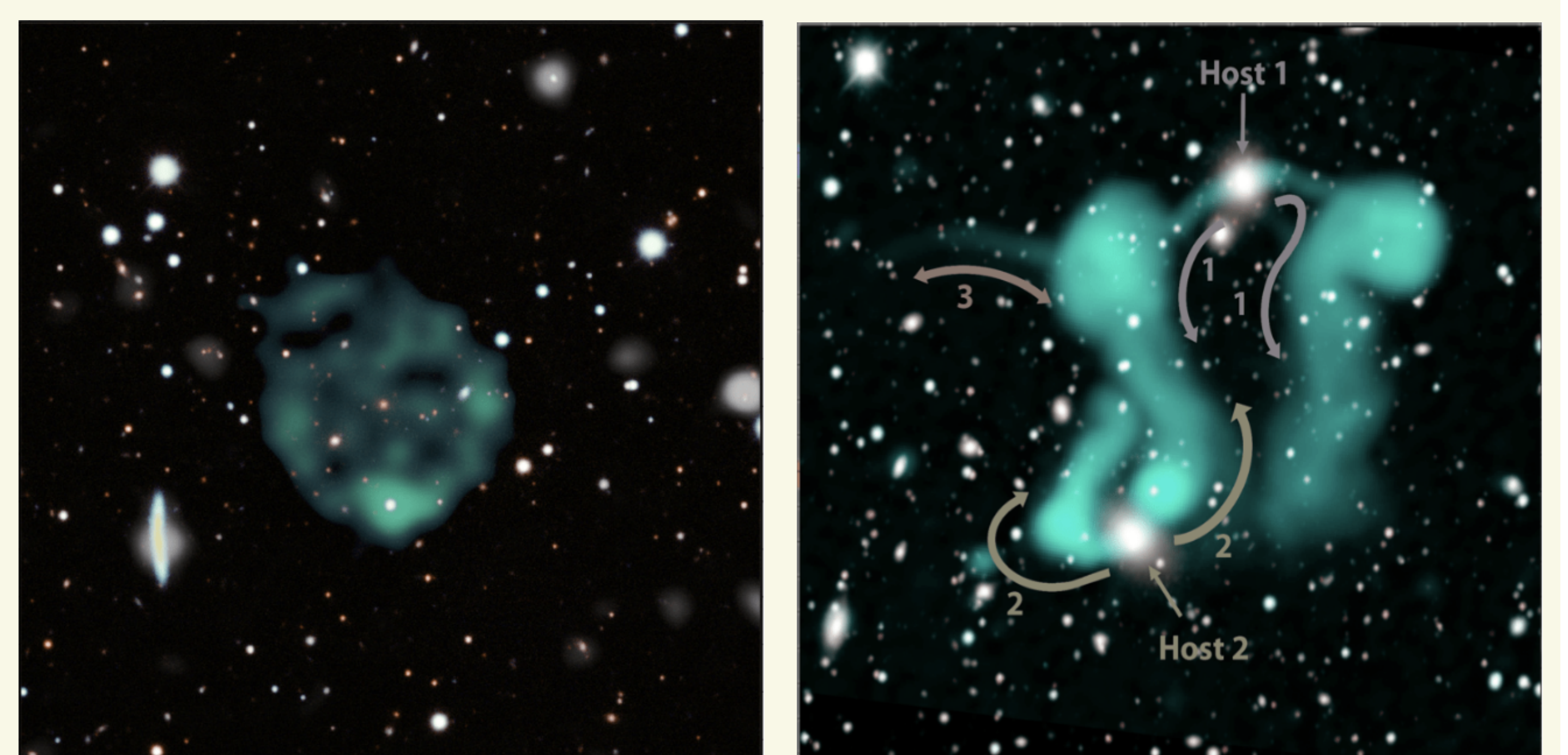


Fig. 4: Left: An example of Odd Radio Circle [2] . **(Right):** An example of Peculiar Radio Source [11] . The source consists of a group of distorted radio components, collectively known as PKS 2130–538, and nicknamed “the dancing ghosts”.

Novel methodologies

Anomaly detection to prioritise cutouts: Input data in the form of cutouts ($6' \times 6'$; Fig 2) are ranked using *complexity* [8, 9] to provide the most interesting objects to view and facilitate discoveries (Fig. 4).

Taxonomy as a classification scheme: *Semantically meaningful tags* [7] are derived (Fig. 3), for the first time in astrophysics, using natural language processing. The aim to provide a common language for classification between experts and to improve the communication between experts and citizen scientists.

References

- [1] R. Norris et al., PASA, **28**, 215, 2011[2] R. Norris et al., PASA, **38**, 3, 2021[3] M. Whiting & B. Humphreys, PASA, **29**, 371, 2012[4] T. M. C. Abbott et al., ApJS, **239**, 18, 2018[5] E. L. Wright et al., AJ, **140**, 1868, 2010[6] L. Rudnick et al., Galaxies, **9**, 85, 2021[7] M. Bowles, et al., arXiv:2210.14760, NeurIPS, 2022[8] G. Segal et al., PASP, **131**, 8007, 2019[9] G. Segal et al., arXiv:2206.14677, MNRAS, 2023[10] L. K. Polsterer et al., ASPC, **495**, 81, 2015[11] R. Norris et al., PASA, **38**, 46, 2021

