Where astrophysics meets citizen science: Radio Galaxy Zoo EMU (RGZ-EMU)

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 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 deg^2 of the southern sky, reaching an rms of 25–30 μJy/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.

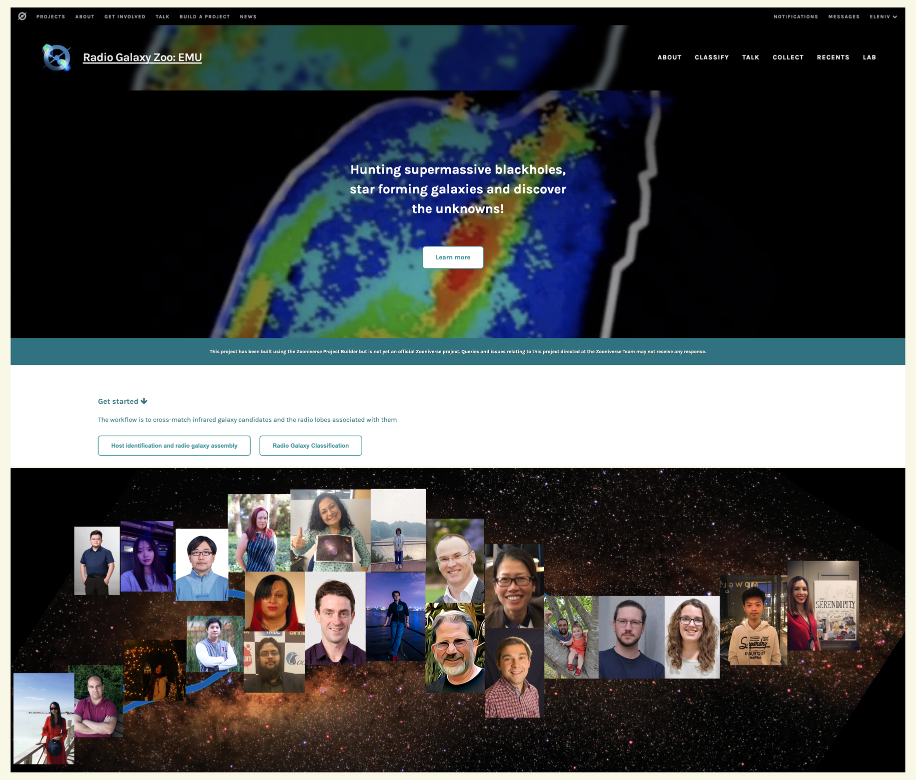

Fig. 1: (*Top*): **The RGZ-EMU Homepage** *It presents the two workflows: 1) source identification and assembly, and 2) source classification*, as well as details about the project, the team, and the education aspect of RGZ-EMU. (*Bottom*): Members of the RGZ-EMU team.

Novel methodologies

Anomaly detection to prioritise cutouts: Input data in the form of cutouts ($6' \times 6'$; Fig 2) are ranked using *complexity* [4, 5] to provide the most interesting objects to view and facilitate discoveries (Fig. 3).

A semantically meaningful taxonomy as a classification scheme: *Semantically meaningful tags* (Fig. 2–*Right*; [6] are derived (Fig. 4–*Left*), for the first time in astrophysics, using natural language processing (NLP). The aim is to provide a common language for classification between experts and to improve clarity in the communication between experts and citizen scientists.

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; [7]), and infrared (WISE; [8]) images.
- **Classify radio sources** using descriptive tags [9] that were derived using results from early RGZ-EMU classification experiments and natural language processing [6] .

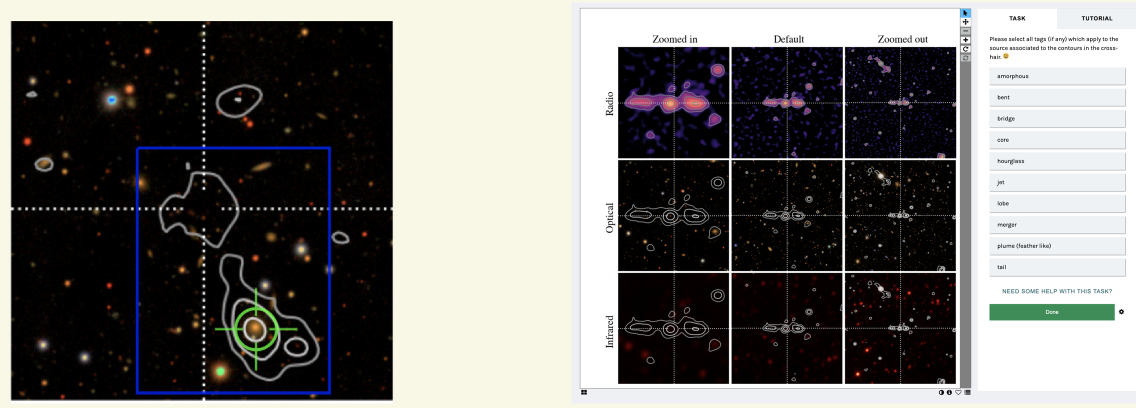

Fig. 2: Examples of the two workflows for the RGZ-EMU project release. *(Left:)* Zoom-in of the host identification (green circle) and source assembly (blue box) workflow. *(Right:)* Classification workflow based on tags. The setup of the cutouts in three different sizes ($3' \times 3'$, $6' \times 6'$, 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. 3). To facilitate discoveries, we applied the complexity methodology [4, 5] to provide the most interesting cutouts to view.

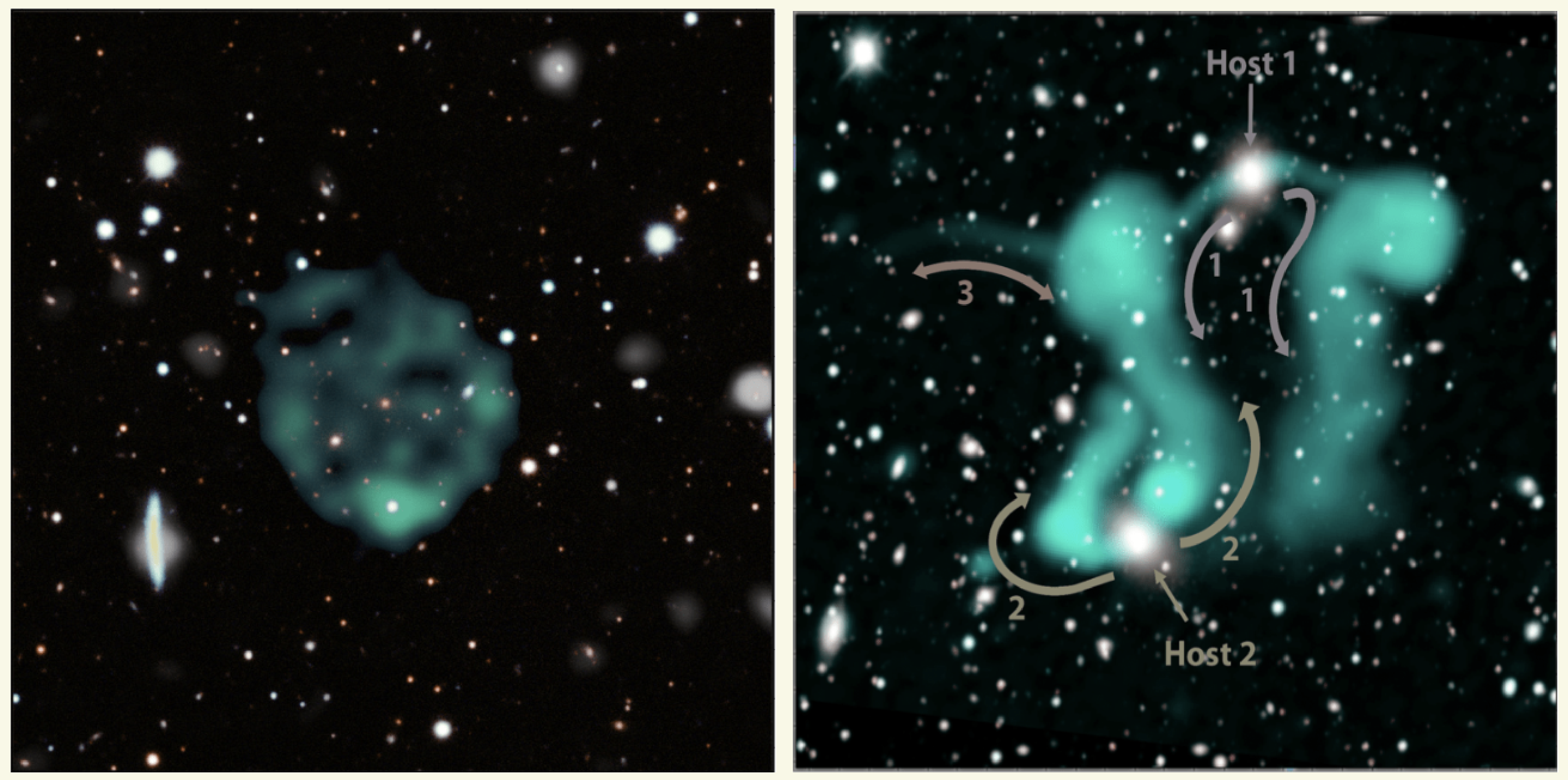

Fig. 3: *Left*: An example of Odd Radio Circle [2] . *(Right:)* An example of Peculiar Radio Source [10] . The source consists of a group of distorted radio components, collectively known as PKS 2130–538, and nicknamed "the dancing ghosts".

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. 4–*Left*) and provides a set of tags to the RGZ-EMU workflow (Fig. 2–*Right*; [6]). These are meant to **replace astro jargon** (e.g. FRI, FRII, WAT). 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; [11]) algorithm, with the purpose of radio-source classification and the discovery of peculiar radio sources (Fig. 3).

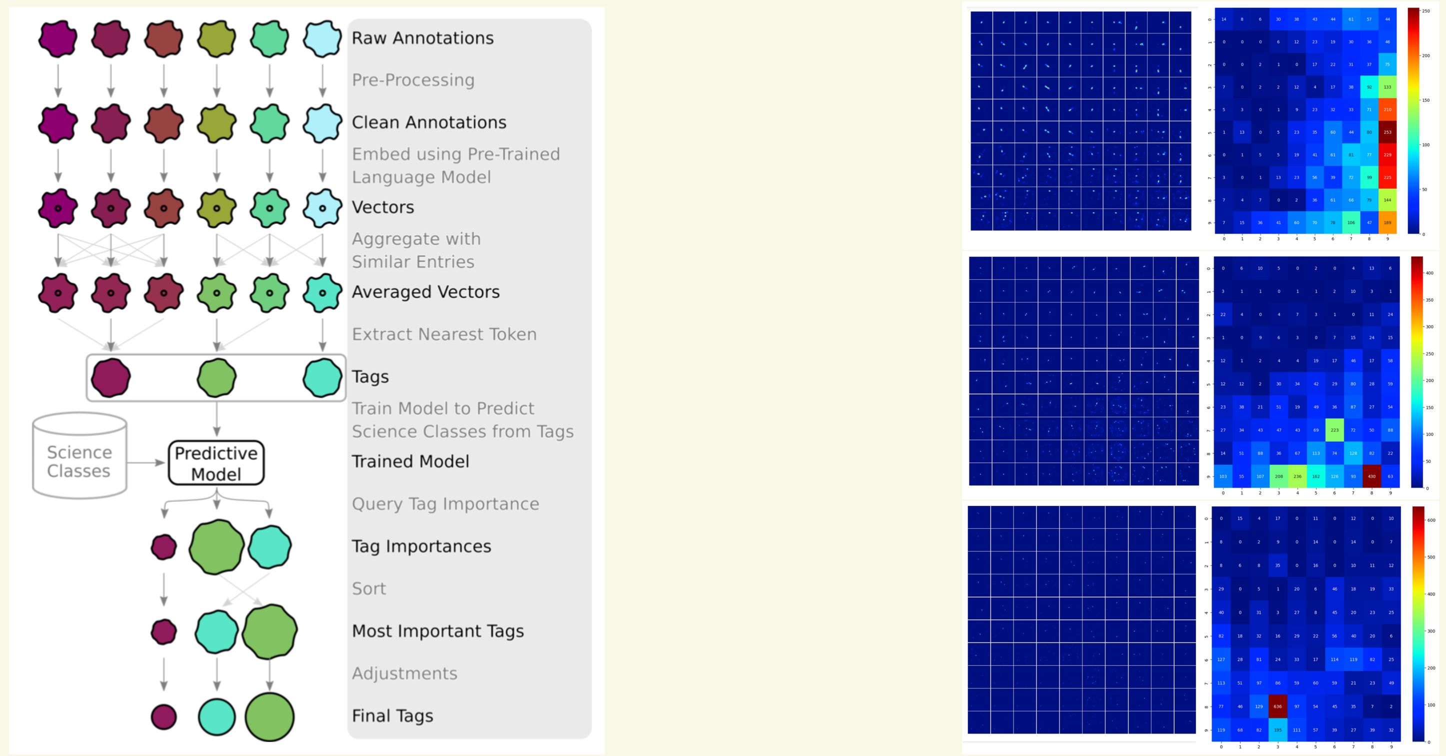

Fig. 4: *(Left:)* Proposed workflow to computationally derive a semantically meaningful plain English taxonomy from a set of annotations (Bowles et al. in prep.). *(Right:)* Testing the effect of different cutout sizes ($3' \times 3'$ – top); $6' \times 6'$ – middle; and $12' \times 12'$ – bottom) as input to self organising maps using PINK [11] . The SOM was run by students (Anees, Dinh, Medina, Meyounyo, Mirshanova, Pöppelmann) at the Darmstadt University of Applied Sciences.

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] G. Segal et al., PASP, **131**, 8007, 2019
[5] G. Segal et al., arXiv:2206.14677, MNRAS, 2023
[6] M. Bowles, et al., arXiv:2210.14760, NeurIPS, 2022; https://github.com/mb010/Text2Tag
[7] T. M. C. Abbott et al., ApJS, **239**, 18, 2018
[8] E. L. Wright et al., AJ, **140**, 1868, 2010
[9] L. Rudnick et al., Galaxies, **9**, 85, 2021
[10] R. Norris et al., PASA, **38**, 46, 2021
[11] L. K. Polsterer et al., ASPC, **495**, 81, 2015

Eleni Vardoulaki **Hongming Tang et al.**
elenivard@gmail.com hongmingt@mail.tsinghua.edu.cn
Thüringer Landessternwarte Tautenburg; Tsinghua University