Proposal: Towards a Statistical Understanding of Galaxy Evolution
Potential Fellow: David O'Ryan
Potential Mentors & Collabs: Bruno Altieri, Guido de Marchi, Sandor Kruk & Bruno Merin

Research Rationale

Galaxy Morphology, Photometric Parameterisation and Galaxy Evolution

Galaxy morphology is a key indicator of galactic evolutionary history. We can make assumptions about this if it is a disk or elliptical, barred or unbarred, or if it has spiral arms, etc. For example, spiral galaxies are commonly gas rich and star forming compared to elliptical galaxies (Smethurst et al., 2022), evidence is growing that strongly barred galaxies are more likely to have active nuclear regions compared to their unbarred counterparts (Garland et al., 2023), and disk-dominated galaxies are likely to have merger-free histories (Simmons et al., 2017). To fully understand the link between a galaxy's morphology and evolutionary history we need large, statistically significant catalogues. The primary aim of my fellowship project will be to create these large morphological catalogues and combine it with the spectroscopic and imaging data of Euclid. With this ancillary catalogue of high-level science products (HLSPs), I will then explore the interplay of galactic nuclear activation with interaction.

Large morphological catalogues are available, primarily from the Galaxy Zoo collaboration (Lintott et al., 2008). These have been the largest and most accurate catalogues of detailed morphologies in the field, but they are still too small for the fine-grained and carefully controlled experiments needed to make major progress in our modern understanding of galaxy evolution. Larger datasets will shortly be available, but even citizen science approaches struggle to keep up with the scale of the latest generation of surveys.

`Zoobot` (Walmsley et al., 2023a), a recent development in representation learning from the Galaxy Zoo collaboration provides a solution to both of these problems. It is a convolutional neural network (CNN) trained to answer all thirty four morphology classification tasks in the Galaxy Zoo workflow (e.g. Fig. 4 of Walmsley et al., 2022). It has had significant success in a number of works, from successfully approximating the volunteer responses to 314,000 galaxies in the Galaxy Zoo DECaLS project (Walmsley et al., 2022), to finding interacting galaxies in the *Hubble Space Telescope* (HST) archives (O'Ryan et al., 2023), to successfully classifying the morphology of 8.7 million galaxies in the Galaxy Zoo: DESI project (Walmsley et al., 2023b). Thus, providing large sub-samples of different galaxy morphologies and pioneering insights into the interplay between morphology and the underlying physical processes driving galaxy evolution.

Another set of classifications important to galaxy evolution is photometric parameterisation. These are parameters measured from a galaxy's broad band photometry using spectral energy distribution (SED) fitting. The resulting catalogues provide the stellar masses, star formation rates, estimates of emission line luminosities, nuclear activity and a host of other parameters. These SED matches are conducted with well established astrophysical software such as the Easy and Accurate zPhot from Yale (EAzY; Brammer et al., 2008) and the Fitting and Assessment of Synthetic Templates (FAST; Kriek et al., 2009). Many examples of such ancillary catalogues have been produced, a prime example being the Cosmic Evolution Survey (COSMOS; Capak et al., 2007) catalogue. These contain the parameterisations of over 1.8 million galaxies focused on a $3^\circ \times 3^\circ$ area of the sky. This catalogue has seen significant use by the community, with the initial data release having over 650 citations.

However, the size of catalogues which overlap morphological and ancillary data remains limited.

Therefore, a key opportunity exists to enhance our understanding of the interplay between galactic morphology, their parameterisations and specific evolutionary processes. From my prior experience with state-of-the-art tools and the ESA Datalabs framework I am uniquely positioned to create this HLSP catalogue and to analyse it.

Impact & Past Achievements

Galaxy Evolution Over the Last 8 Billion Years

My research has been focused on galaxy evolution, primarily the effects of interaction and merging on the growth of galaxies. The specific influence of galaxy interaction, when two galaxies come close enough to have a gravitational effect upon one another, are poorly understood. In general, interaction leads to a sudden increase in star formation (Moreno et al., 2021) often causing quenching (Das et al., 2023), severe morphological disturbance (Toomre & Toomre, 1972) and is sometimes linked to the ignition of nuclear activity (Comerford et al., 2015). The parameter space of interaction is very complex, to explore the variations in these effects and quantify their impact on evolution we require large catalogues of interacting and merging galaxies.

Creating these catalogues by visual inspection is a notoriously difficult task. Numerous works have shown that citizen science (Darg et al., 2010) or machine learning (Pearson et al., 2022) techniques often lead to significant levels of contamination by close pairs. These are galaxies which appear close together in the two dimensional plane of the sky but are at very different redshifts. I have contributed to numerous projects working to overcome this difficulty (e.g Martin et al., 2022; Margalef-Bentabol et al., submitted), and I led the first study combining `Zoobot` and ESA Datalabs to build a pure catalogue of interacting galaxies. In O'Ryan et al. (2023), I used `Zoobot` with ESA Datalabs to explore the *HST* archives for interacting galaxies (O'Ryan et al. 2023). We classified a total of 92 million sources, creating a catalogue of 21,926 interacting galaxies. This is the largest to date. This was the first published work where `Zoobot` was used to such a scale on a specific classification problem. This initial use of `Zoobot` was highlighted in the official release paper of the model (Walmsley et al., 2023a).

A secondary result of using `Zoobot` with ESA Datalabs was that a host of other objects of astrophysical interest were discovered in the archives. Because the morphological irregularities seen in interacting galaxies share some properties with other unusual objects, `Zoobot` also identified a small but interesting set of non-interacting objects of morphological or astrophysical interest. As a byproduct of my approach, I therefore created some of the largest catalogues of various astrophysical objects to date. Examples of these extra objects include ring galaxies, overlapping galaxies, quasars and edge-on proto-planetary disks. This approach, and internship, gave me a unique insight into how to use ESA Datalabs to efficiently provide me with a large repository of existing pipelines to stream line the ramp up of this project. It also demonstrated to me the applicability of ESA Datalabs in the field. Of the 21,926 interacting galaxies classified, two-thirds had no record on astrophysical data repositories like SIMBAD or the NASA Extragalactic Database. The 3-month internship that enabled this result also places me in a unique position to deliver further high-value, high-level ESA data products to the astrophysical community and to capitalise on the scientific opportunities they present.

This large, highly pure catalogue of intercting galaxies facilitates exploration of the physical processes underlying mergers and their consequences for galaxy evolution. One way I am investigating this is by using simulations to constrain the observations directly. I have developed software that uses an Markov-Chain Monte Carlo approach to infer the underlying parameters of interaction directly from observations. This code automates the methodology of Holincheck et al. (2016) within a rigorous statistical framework, allowing us to apply it to a much larger sample. It also incorporates the evolution of the stellar populations of each galaxy, and the enhancement in star formation. The release paper of this software (in prep.) shows it's application to the Holincheck et al. (2016) sample as a fiducial dataset. Once finalised, we will apply this the significantly larger O'Ryan et al. (2023) galaxy catalogue and directly quantify the effect of the underlying parameters of interaction with the tidal features that form and the physical processes that we observe.

A second approach I am adopting is to use ancillary data from different catalogues to find the underlying parameters of interaction. 3,503 galaxies within the O'Ryan et al. (2023) sample are in the COSMOS2020 catalogue (Weaver et al., 2022), with a wealth of measured multi-wavelength and ancillary data. As this work nears completion, I have made some insights into the interplay between galaxy interaction, the environment, nuclear activity, star formation enhancement and projected separation. This work, which will be published before I begin this fellowship, has involved the construction of pipelines and analysis techniques that will form the basis of those in this project.

Research Plans for Fellowship

This project has six key goals which will be achieved with four scientific results (RN) and two deliverables (DN). I have defined four **work packages** (WPs) that follow the full process of this project across eleven different tasks. For a full breakdown of the expected timeline, specific tasks and their WPs please see Figure 1. The project goals are:

1. WP1:
   (a) To create a fully open, and accessible image repository of every extended source at rest frame in the *Hubble* and JWST public archives (T1.1; D1).
   (b) To apply `Zoobot` to the above created repository (T1.2, T1.3; R1). Thus, enabling others to more easily deploy `Zoobot`.
2. WP2:
   (a) To utilise Euclid data to create a catalogue of multi-wavelength and photometric parameterisation data (T2.1, T2.2, T2.3; R2).
3. WP3:
   (a) To match these catalogues and create the catalogue of high-level science product data, while fully diagnosing their accuracy and completeness (T3.1, T3.2; R3). If we only used the HSC, this would be at least 126 million objects.
4. WP4:
   (a) To create the largest catalogue of spectroscopically confirmed interacting galaxies from the *NISP* instrument from Euclid (T4.1; D2).
   (b) To use this catalogue in the context of the interplay between galaxy interaction and the ignition of nuclear activity. Specifically, I will answer when in an interaction nuclear activity begins by breaking down each interaction into chronological stages. I will also investigate the stellar masses and mass ratios of the interaction that are required for ignition (T4.2, T4.3; R4).

Synergies with ESA Missions

This project is directly related to multiple ESA missions, with the two primary missions being the development of the archives with ESA Datalabs and the new flagship mission Euclid. The archives, and the new infrastructure built around it, form a cornerstone of this project. Without

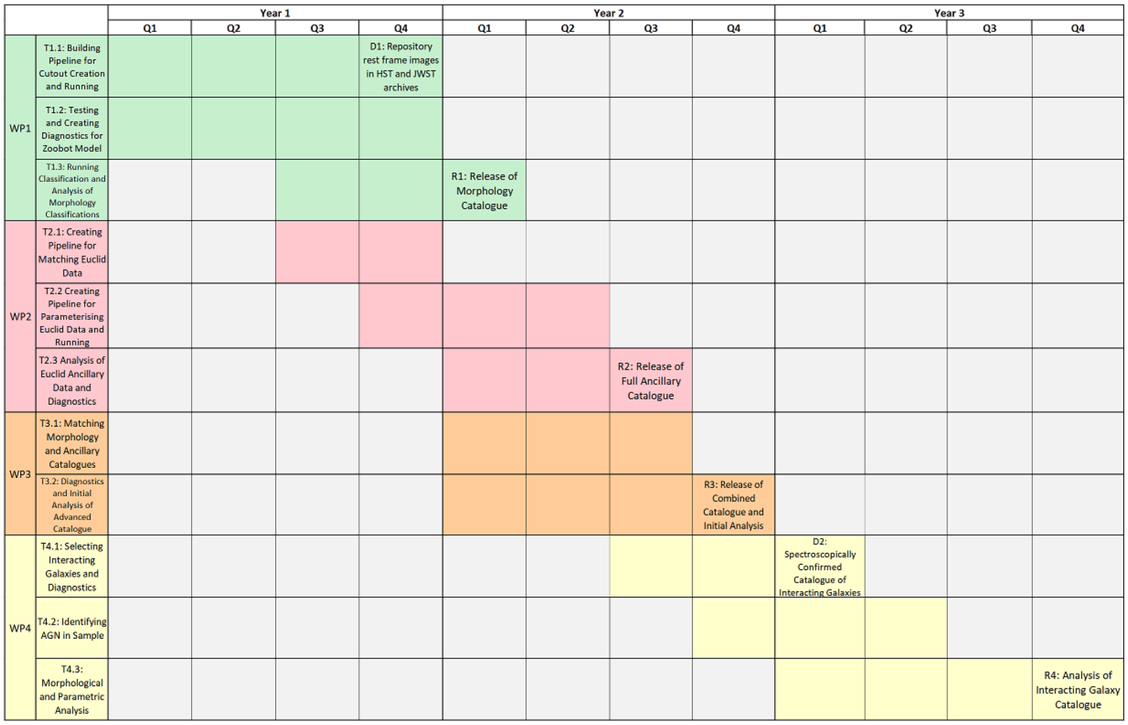

Figure 1: GANTT chart of proposed project timeline. Includes workpackes (WPN), scientific results (RN) and deliverables (DN).

ESA Datalabs and the unique kind of access to archival data it provides, this project would not be possible. The second is Euclid. Using Euclid's all-sky survey Near-Infrared Spectrometer and Photometer, we will gain an estimated 30 million spectra in the initial data release. We will use these spectra to extrapolate each sources photometric parameterisations with well established software.

This project will also use the ESA-NASA mission JWST. JWST data is specifically available on ESA datalabs and can be accessed by simple adaption of our pipelines. The combined public archive across many projects is certainly powerful, but we will need to pay specific attention to combine it into a large high-redshift dataset, which will be of high interest to the community and complementary to the wider-area but shorter wavelength coverage of Euclid. The ancillary data JWST provides, as well as its unique spectroscopy and imaging of high redshift systems will allow us to create a more complete catalogue.

My previous experience with ESA Datalabs, `Zoobot` and utilising the ancillary data of the COSMOS catalogue put me in a unique position to create and analyse these new advanced catalogues. While building the pipelines to create cutouts, or conduct classifications, would certainly take time to build the fact that I have much of these pipelines ready for use will significantly streamline the ramp up of the project. From this project, I also hope to fully demonstrate the power of Euclid combined with that of ESA Datalabs and the archives. By the end of my ESA Fellowship, I will have created a large, HLSP catalogue of galaxy morphology and parameterisation. I will have demonstrated the full potential of ESA Datalabs, and will have combined data from the *HST*, JWST and Euclid - the three greatest observatories of our time. I will release these catalogues to the community, with the aim of enabling it to have easy access to HLSP for highly efficient research. I will also have answered my own scientific questions of: What is the specific effect of galaxy interaction on nuclear activity? When in an interaction does nuclear activity clearly begin? And what photometric parameters are required for the maximum likelihood that ignition occurs?

References

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503 • Capak P., et al., 2007, ApJS, 172, 99 • Comerford J. M., Pooley D., Barrows R. S., Greene J. E., Zakamska N. L., Madejski G. M., Cooper M. C., 2015, ApJ, 806, 219 • Darg D. W., et al., 2010, MNRAS, 401, 1552 • Das A., Pandey B., Sarkar S., 2023, Research in Astronomy and Astrophysics, 23, 025016 • Garland I., et al., 2023, MNRAS, submitted • Holincheck A. J., et al., 2016, MNRAS, 459, 720 • Kriek M., van Dokkum P. G., Labbe, Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, ApJ, 700, 221 • Lintott C. J., et al., 2008, MNRAS, 389, 1179 • Margalef-Bentabol B., et al., sub, A&A, submitted • Martin G., et al., 2022, MNRAS, 513, 1459 • Moreno J., et al., 2021, MNRAS, 503, 3113 • O'Ryan D., et al., 2023, ApJ, 948, 40 • Pearson W. J., et al., 2022, A&A, 661, A52 • Simmons B. D., Smethurst R. J., Lintott C., 2017, MNRAS, 470, 1559 • Smethurst R. J., et al., 2022, MNRAS, 510, 4126 • Toomre A., Toomre J., 1972, ApJ, 178, 623 • Walmsley M., et al., 2022, MNRAS, 509, 3966 • Walmsley M., et al., 2023a, The Journal of Open Source Software, 8, 5312 • Walmsley M., et al., 2023b, MNRAS, submitted • Weaver J. R., et al., 2022, ApJS, 258, 11