

Twelve Years of Galaxy Zoo

Karen L. Masters¹ and Galaxy Zoo Team

¹Haverford College, Department of Physics and Astronomy, 370 Lancaster Avenue, Haverford, PA 19041, USA
 email: klmasters@haverford.edu

Abstract. The *Galaxy Zoo* project has provided quantitative visual morphologies for over a million galaxies, and has been part of a reinvigoration of interest in the morphologies of galaxies and what they reveal about galaxy evolution. Morphological information collected by GZ has shown itself to be a powerful tool for studying galaxy evolution, and GZ continues to collect classifications - currently serving imaging from DECaLS in its main site, and running a variety of related projects hosted by the *Zooniverse*; the citizen science platform which came out of the early success of GZ. I highlight some of the results from the last twelve years, with a particular emphasis on linking morphology and dynamics, look forward to future projects in the GZ family, and provide a quick start guide for how you can easily make use of citizen science techniques to analysis your own large and complex data sets.

Keywords. galaxies:fundamental parameters, galaxies:kinematics and dynamics, galaxies:evolution, galaxies:statistics, catalogs

1. Introduction

It has been more than twelve years since the launch of *Galaxy Zoo* (on 11th July 2007)†, and the project has been continuously active during that period, inviting volunteers to help provide galaxy classifications, using images from a variety of different sources (see Table 1); and with the science team making use of the aggregated and reduced classifications to investigate a wide range of topics in extragalactic astrophysics.

This invited talk for the IAU Symposium, “Galactic Dynamics in the Era of Large Surveys” was to cover “Key Results from *Galaxy Zoo*”. As well as providing a brief overview of twelve years of the project, I will focus on results (given the topic of the conference) most linked to disc galaxy dynamics and the structures this creates, and also look to the future plans for *Galaxy Zoo* and the *Galaxy Zoo* method.

2. The Galaxy Zoo Method

Galaxy Zoo was inspired by the 1 million galaxy sample of the original Sloan Digital Sky Survey (SDSS; namely the Main Galaxy Sample of Strauss et al.(2002), which was fully released in Data Release 7, Abazajian et al.(2009)). This was the original data set, and the original *Galaxy Zoo* (Lintott et al.(2008)) asked for simple classification into spiral or not; further asking spirals to be identified as “S” or “Z” shaped (to test claims of a preferred sense of rotation; Land et al.(2008)), or edge-on/can’t tell, and had categories for merging systems (and artefacts). This sample was classified with a median of 40 people per image in about eighteen months of operations, and the aggregated classifications were released in Lintott et al.(2011).

Following the success of the first phase, *Galaxy Zoo 2*, (GZ2) was launched, asking for more detailed classifications for the brightest ($r < 16$) 250,000 of the SDSS sample.

† www.galaxyzoo.org

Figure 1 shows an example set of responses for galaxy NGC 2771 from GZ2‡. The GZ2 classifications were reduced and released in Willett et al.(2013); an improved treatment of redshift bias on the spiral arm questions is presented in Hart et al.(2016), and these improved classifications are now also available.

Following the completion of classifications on the original SDSS imaging, *Galaxy Zoo* relaunched with all public images from optical Hubble Space Telescope surveys (COSMOS, GOODS, etc; these classifications were published in Willett et al.(2017)), as well as the images from the CANDELS survey (classifications available in Simmons et al.(2017)). Other survey images for which *Galaxy Zoo* have obtained classifications include UKIDSS and KiDS; these are not yet published. We have also shown images generated by the Illustris simulation (see Dickinson et al.(2018)), with the goal of testing how well such large scale simulations do at reproducing the population demographics of galaxy morphology. Currently live on the site are images from the Dark Energy Camera Legacy Survey (DECaLS; also see Section 7).

3. Summary of Published Results

The *Galaxy Zoo* science team have published over 60 papers in the last 12 years which make use of the classifications from different phases of GZ† At the time of writing there are twelve papers from the team with more than 100 citations each.

Attempting to categorize publications into types, GZ team papers can be broken down into:

- *Data papers* which describe the techniques to go from clicks to classifications (including user weighting, and accounting for image quality, and changes in how galaxies appear in images as a function of redshift). These are: Lintott et al.(2011), Willett et

‡ See www.visualise.galaxyzoo.org for more examples

† A complete listing of team papers can be found at www.zooniverse.org/about/publications.

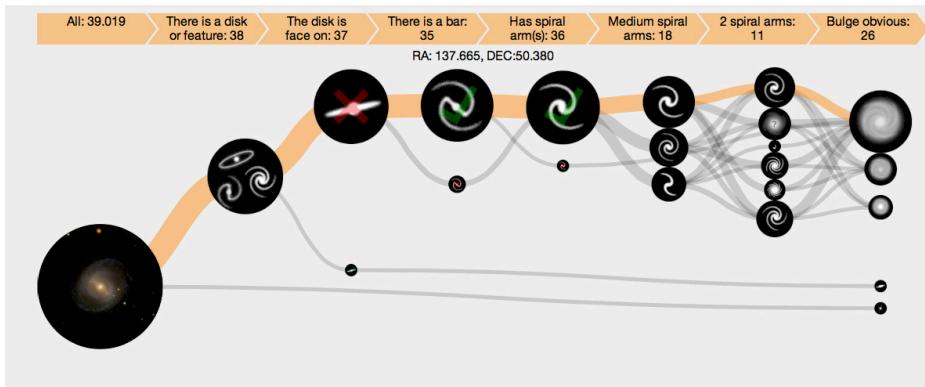


Figure 1. An example classification from *Galaxy Zoo* 2 in the format which is available at www.visualise.galaxyzoo.org. This galaxy is NGC 2771, which has a classification of (R')SB(r)ab in the RC3 (de Vaucouleurs et al.(1991)). The numbers in the orange bar at top represent the number of users (after a weighting scheme is applied which removes users who always provide highly inconsistent answers) who responded positively to the question indicated.

Table 1. Summary of Data Sets and Projects in the *Galaxy Zoo* Project

Project	Size (approx)	Imaging Source/ Classification Type	Data Access	Reference
Original <i>Galaxy Zoo</i>	1 million	SDSS Main Galaxy Sample (spiral/early-type/mergers only)	Public ^{1,2}	Lintott et al.(2011)
<i>Galaxy Zoo 2</i>	250,000	SDSS Main Galaxy Sample $r < 17$. Full classification tree.	Public ^{1,2}	Willett et al.(2013) Hart et al.(2016)
<i>Galaxy Zoo: Hubble</i>	120,000	Public Optical HST Surveys (COSMOS, GOODS etc)	Public ¹	Willett et al.(2017)
<i>Galaxy Zoo: CANDELS</i>	48,000	CANDELS Survey Galaxies	Public ¹	Simmons et al.(2017)
<i>Galaxy Zoo: MaNGA</i>	30,000	MaNGA Target Sample GZ2 style	Public ^{1,2}	Aguado et al.(2019) (SDSS DR15)
<i>Galaxy Zoo 4</i> <i>Galaxy Zoo: UKIDSS</i> <i>Galaxy Zoo: Illustris</i> <i>Galaxy Zoo: KiDS</i>		SDSS Extra Imaging (DR12) $r < 17$ UKIDSS Illustris Simulated Galaxies GAMA KiDS	Pending Pending Pending Pending	Dickinson et al.(2018) Holwerda et al.(2019)

Notes:

¹Public data is available at www.data.galaxyzoo.org

²Also available via SDSS Casjobs at www.skyserver.org

al.(2013), Willett et al.(2017), Simmons et al.(2017). In these papers you can also find comparisons of *Galaxy Zoo* morphologies with other published classifications and morphological proxies (e.g. structural parameters).

- Papers which look at *broad statistical properties of the morphologies of galaxies*. These have often focused on the comparison of colour and morphology (see Section 4). Examples include: Bamford et al.(2009), Skibba et al.(2009), Schawinski et al.(2009), Masters et al.(2010), Schawinski et al.(2014), Smethurst et al.(2015), Hart et al.(2016), Masters et al.(2019).

- Papers about *interesting/odd objects (or classes of objects) revealed by Galaxy Zoo*. Examples include the green peas (compact ionized galaxies, see Cardamone et al.(2009)) and Hanny’s Voorwerp (a light echo from a recently quiescent accreting black hole; see Lintott et al.(2009) and recent work by Keel et al.(2019)).

- Papers which investigate the *morphology of specific types of galaxy, or the properties of galaxies with specific morphological features*. Examples include work on the host galaxies of Active Galactic Nuclei (AGN): Schawinski et al.(2010), Simmons et al.(2013), Galloway et al.(2015), Smethurst et al.(2016); mergers: Darg et al.(2010), Darg et al.(2010) or many papers on the properties of disc galaxies with bars: Masters et al.(2011), Masters et al.(2012), Cheung et al.(2013), Melvin et al.(2014), Simmons et al.(2014), Kruk et al.(2018). Section 5 will discuss our work on bars in more detail.

In addition to team papers, the public release of classifications has led to many more results making use of GZ classifications. Overall, at the time of writing† the classifications have been used or mentioned in over 1000 papers in the last 12 years. Figure 2 shows how this has grown since 2009 (note that 2019 data is not yet complete).

4. Colour and Morphology

One of the earliest contributions to come out of GZ classifications, was putting on a firm statistical footing the correlation between galaxy colour and morphology. While it

† Based on a literature search of the phrase ”Galaxy Zoo”

had been known for a long time (e.g. as commented by Zwicky et al.(1955)) that spiral galaxies were bluer than elliptical galaxies, it was the advent of large galaxy surveys (most notably SDSS) that led to the common perception that colour and morphology were entirely equivalent. Although this result was based on fairly small samples of galaxies with morphological classifications, it remains fairly common in the astronomical literature today.

In a pair of early papers, GZ classifications were used to look at the correlation between colour in morphology in a larger sample than ever before (Skibba et al.(2009), Bamford et al.(2009)). These papers revealed the existence of significant numbers of red spirals and blue ellipticals which were investigated in Masters et al.(2010) and Schawinski et al.(2010) respectively. Red spirals in particular were found to be quite common, with more than half of the most massive spirals being optically red (even when removing highly inclined spirals, and very early-type spirals with large bulges). Example images of red and blue spirals are shown in Figure 3. This demonstrated that it is important to recognise that selections based on colour cannot be interpreted as cleanly dividing the population into the two broad morphological types.

The imperfect correlation between colour and morphology makes sense physically. The morphology of a galaxy (to “first order”, ignoring things like dust) provides information on the orbits of stars within it. As such, it should be obvious that important clues to the formation history of galaxies is revealed by their morphologies, and this information is complimentary, but not identical to, their star formation history and chemical composition as revealed by photometry (i.e. colour) and spectra.

5. Bars and Spiral Arms

The more detailed classifications of GZ2 enabled the study of features in disc galaxies tied intimately to their dynamics, such as bars, and spirals.

5.1. Bars

The GZ team have also published a wide range of papers investigating bars in galaxies. Early results on bars revealed something unexpected at the time, that they were more

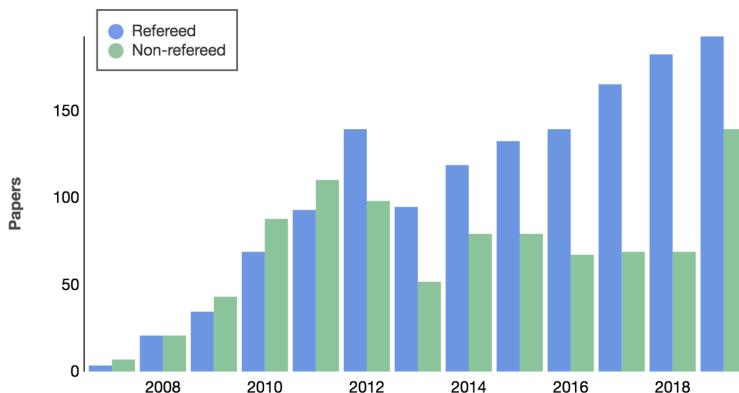


Figure 2. Count of papers (refereed in blue, non-refereed in green) mentioning the term ”*Galaxy Zoo*” over the last twelve years, since *Galaxy Zoo* launched in 2007. Credit: NASA/SAO Astrophysics Data System; ADS

common in redder spirals (Masters et al.(2011)). Work to compare what kinds of bars are identified by GZ classifications (with a threshold of $p_{\text{bar}} > 0.5$) demonstrated they were strong bars (SB in the traditional notation, but not SAB); while weaker bars can be identified by lower vote fractions this is at the risk of including larger numbers of erroneous classifications; see e.g. Willett et al.(2013)). Example images of a red barred spiral and a blue unbarred spiral are shown in Figure 3. Further work has confirmed a picture of strong bars being preferentially associated with massive, red, gas poor (Masters et al.(2012)) and low star forming (Cheung et al.(2013)) discs, and in addition showing evidence for secular growth of pseudo-bulges (Cheung et al.(2015)). Classifications on the public HST images, and CANDELS allowed investigation of bar fractions to be extended to higher redshift revealing how they become more common in the later Universe (Melvin et al.(2014), Simmons et al.(2014)). The clustering of GZ2 barred galaxies was investigated in Skibba et al.(2012), while Galloway et al.(2015) studied the impact galactic scale bars had on AGN properties, and Kruk et al.(2017), Kruk et al.(2018) revealed the presence of offset bars, and provided three component (bar, bulge, disc) photometric fits for a large sample from GZ2.

5.2. Spiral Arms

Spiral arms (the number and winding angle) are also described in GZ2 classifications. The properties of spiral arms in the population has been investigated in a series of recent papers. Hart et al.(2016), Hart et al.(2017a) provided an updated method to correct classifications for redshift bias, and showed that many arm spirals are associated with recent star-formation, and also that the distribution of SF differs between two-armed (“grand design”) and many armed (“flocculent”) spirals. The winding properties identified in GZ2 were then used to investigate spiral formation mechanisms; interestingly we find that the classic static density wave does not match GZ2 classifications on a population scale – they reveal a lack of strong correlation between bulge size and spiral arm pitch angle predicted by those models (Hart et al.(2017bs), Hart et al.(2018), Masters et al.(2019)) . Further work on this topic may reveal many interesting further constraints on spiral arm formation mechanisms.



Figure 3. Example images of red (left) and blue (right) spiral galaxies identified in *Galaxy Zoo*. These also demonstrate the barred/not barred selection from GZ2.

6. Spin Off Projects

The GZ technique of crowdsourcing the analysis of galaxy images can also be applied to more than just answering simple questions. New capabilities in *The Zooniverse Project Builder* (see Section 8) make this very straightforward now, but even before this existed there were efforts to make spin-off projects linked to GZ by science team members. Spin-off *Galaxy Zoo* projects run by the science team in part, or whole include:

- *Galaxy Zoo Mergers* - a custom built project in which volunteers were asked to run models of galaxy mergers and find the best match to images. (Holincheck et al.(2016))
- *Galaxy Zoo Bar Lengths* - this project ran in a Google interface to *Galaxy Zoo* and measured bar lengths on SDSS images. For details see Hoyle et al.(2011). A second *Galaxy Zoo Bar Lengths* was run more recently† to do expand on this work using Hubble images; results will be presented in Hutchinson-Smith et al. in prep.
- *Galaxy Zoo: 3D* - this spinoff uses *Project Builder* to display images of galaxies which are part of the SDSS-IV project, Mapping Nearby Galaxies at Apache Point Observatory (MaNGA; Bundy et al.(2015)), and is asking users to draw the locations of bars, spiral arms and foreground stars (as well as mark the centers of the galaxies)‡. The resulting masks can be used with MaNGA data to extract spectra from the respective features (e.g. as used by Peterken et al.(2019), Fraser-Mckelvie et al.(2019)). The project will be described in full in an upcoming Masters et al. in prep.
- *Galaxy Builder* - Another challenge for extragalactic astrophysicists is decomposing galaxy images into multiple components. While large databases exist of simple automated fits, significant human interaction is generally still required to ensure physical results. An attempt to involve crowdsourcing in this process is Galaxy Builder¶, which makes use of crowdsourced galaxy models to select the number of needed components (from disc, bulge, bar, and any number of spirals) and then initiate a final automated fit. This project will be described in full in an upcoming Lingard et al. in prep.. and is a custom project built inside the *Project Builder* framework.
- *Clump Scout* - very recently launched|| this project is asking volunteers to help identify star formation clumps in galaxies.

7. The Future of Galaxy Zoo - Machines and Humans Working Together

For *Galaxy Zoo* itself, the next generation of sky surveys, moving to both deeper and higher resolution will challenge the volume of galaxies which are classifiable by the crowd. Luckily machine learning algorithms are also advancing. An initial study by Beck et al.(2018) demonstrated the use of a random forest algorithm to significantly speed up the first (binary) step on the question tree. Today, the main GZ site is running a new “Enhanced Workflow”, which uses a Bayesian model to prioritize the galaxies shown to volunteers. A Bayesian Convolutional Neural Network trains on the crowdsources classifications at the end of each week, and quickly becomes expert at classifying the simplest galaxies, while identifying more complex galaxy which need to be shown to the human volunteers for further classification. For more details on this see Walmsley et al.(2019).

† <https://www.zooniverse.org/projects/vrooje/galaxy-zoo-bar-lengths>

‡ <https://www.zooniverse.org/projects/klmasters/galaxy-zoo-3d>

¶ <https://www.zooniverse.org/projects/tingard/galaxy-builder>

|| <https://www.zooniverse.org/projects/hughdickinson/galaxy-zoo-clump-scout>

8. How to Build Your Own Zooniverse Project

The early success of *Galaxy Zoo* as a public engagement project (at its peak obtaining 70,000 classifications an hour; and noted by *The Guardian* in 2012^{††} as providing “no surer way to engage the public than to involve people in the research itself”) inspired the creation of *The Zooniverse*^{‡‡}, a platform which hosts similar citizen science projects across any research topic. *The Zooniverse* launched in 2010 to make a common community of citizen scientists, and now has over 1.7 million account holders. In the early days each project was custom built by the development team, which limited the rate at which projects could launch. A significant innovation came with the launch of *Project Builder*^{¶¶} in 2015. This facility allows researcher to build their own projects, which can be submitted to *The Zooniverse* for peer review and beta testing if they wish to be promoted to the millions of account holders, or simply run independently. A simple project based on a few images can be built in a matter of minutes, entirely in the browser, and facilities exist to provide several different types of common crowdsourcing activities (e.g. question trees, drawing, annotation). A total of just over 200 projects have been formally launched by *The Zooniverse*, out of more than 9000 which have been built on *Project Builder* (300 of which have independently collected more than 100 classifications).

Zooniverse Project Builder projects don’t have to be public facing. One example of use within astronomy was for the visual inspection of spectra obtained for the MaNGA Spectral Library (Yan et al.(2019)), where 28 members of the MaNGA science team helped to visually inspect the quality of 10,797 stellar spectra.

9. Summary and Conclusions

In this short paper, based on an invited review talk given at the “Galactic Dynamics in the Era of Large Surveys” conference in Shanghai, China in July 2019 I give a whistle-stop tour through the first twelve years of the *Galaxy Zoo* project, highlighting in particular results related to the imperfect correlation between galaxy colour and morphology, as well as results related to the disc galaxy dynamical structures: galactic bars and spirals. Almost a teenager now, *Galaxy Zoo* is a mature technique, which has had a significant impact on the field of extragalactic astrophysics. New innovations in both crowdsourcing, and the combination of humans and computers working together ensure that the project is still relevant and productive today, has a long future to come.

References

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
- Aguado, D. S., Ahumada, R., Almeida, A., et al. 2019, ApJS, 240, 23
- Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, MNRAS, 393, 1324
- Beck, M. R., Scarlata, C., Fortson, L. F., et al. 2018, MNRAS, 476, 5516
- Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7
- Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, MNRAS, 399, 1191
- Cheung, E., Athanassoula, E., Masters, K. L., et al. 2013, AJ, 779, 162
- Cheung, E., Trump, J. R., Athanassoula, E., et al. 2015, MNRAS, 447, 506
- Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010, MNRAS, 401, 1552
- Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010, MNRAS, 401, 1043
- Dickinson, H., Fortson, L., Lintott, C., et al. 2018, The Astrophysical Journal, 853, 194

^{††} <http://www.guardian.co.uk/science/2012/mar/18/galaxy-zoo-crowdsourcing-citizen-scientists>
^{‡‡} www.zooniverse.org

^{¶¶} <https://www.zooniverse.org/lab>

- Fraser-McKelvie, A., Merrifield, M., Aragón-Salamanca, A., et al. 2019, MNRAS, 488, L6
 Galloway, M. A., Willett, K. W., Fortson, L. F., et al. 2015, MNRAS, 448, 3442
 Hart, R. E., Bamford, S. P., Willett, K. W., et al. 2016, MNRAS, 461, 3663
 Hart, R. E., Bamford, S. P., Casteels, K. R. V., et al. 2017, MNRAS, 468, 1850
 Hart, R. E., Bamford, S. P., Hayes, W. B., et al. 2017, MNRAS, 472, 2263
 Hart, R. E., Bamford, S. P., Keel, W. C., et al. 2018, MNRAS, 478, 932
 Holincheck, A. J., Wallin, J. F., Borne, K., et al. 2016, MNRAS, 459, 720
 Holwerda, B. W., Kelvin, L., Baldry, I., et al. 2019, AJ, 158, 103
 Hoyle, B., Masters, K. L., Nichol, R. C., et al. 2011, MNRAS, 415, 3627
 Keel, W. C., Bennert, V. N., Pancoast, A., et al. 2019, MNRAS, 483, 4847
 Kruk, S. J., Lintott, C. J., Bamford, S. P., et al. 2018, MNRAS, 473, 4731
 Kruk, S. J., Lintott, C. J., Simmons, B. D., et al. 2017, MNRAS, 469, 3363
 Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179
 Lintott, C. J., Schawinski, K., Keel, W., et al. 2009, MNRAS, 399, 129
 Lintott, C., Schawinski, K., Bamford, S., et al. 2011, MNRAS, 410, 166
 Land, K., Slosar, A., Lintott, C., et al. 2008, MNRAS, 388, 1686
 Masters, K. L., Mosleh, M., Romer, A. K., et al. 2010, MNRAS, 405, 783
 Masters, K. L., Nichol, R. C., Hoyle, B., et al. 2011, MNRAS, 411, 2026
 Masters, K. L., Nichol, R. C., Haynes, M. P., et al. 2012, MNRAS, 424, 2180
 Masters, K. L., Lintott, C. J., Hart, R. E., et al. 2019, MNRAS, 487, 1808
 Melvin, T., Masters, K., Lintott, C., et al. 2014, MNRAS, 438, 2882
 Peterken, T., Fraser-McKelvie, A., Aragón-Salamanca, A., et al. 2019, MNRAS 2135
 Schawinski, K., Lintott, C., Thomas, D., et al. 2009, MNRAS, 396, 818
 Schawinski, K., Urry, C. M., Virani, S., et al. 2010, AJ, 711, 284
 Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, MNRAS, 440, 889
 Simmons, B. D., Lintott, C., Schawinski, K., et al. 2013, MNRAS, 429, 2199
 Simmons, B. D., Melvin, T., Lintott, C., et al. 2014, MNRAS, 445, 3466
 Simmons, B. D., Lintott, C., Willett, K. W., et al. 2017, MNRAS, 464, 4420
 Skibba, R. A., Bamford, S. P., Nichol, R. C., et al. 2009, MNRAS, 399, 966
 Skibba, R. A., Masters, K. L., Nichol, R. C., et al. 2012, MNRAS, 423, 1485
 Smethurst, R. J., Lintott, C. J., Simmons, B. D., et al. 2015, MNRAS, 450, 435
 Smethurst, R. J., Lintott, C. J., Simmons, B. D., et al. 2016, MNRAS, 463, 2986
 Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 124, 1810
 de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G., et al. 1991, Third Reference Catalogue
 of Bright Galaxies (RC3).
 Walmsley, M., Smith, L., Lintott, C., et al. 2019, MNRAS submitted; arXiv:1905.07424
 Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835
 Willett, K. W., Galloway, M. A., Bamford, S. P., et al. 2017, MNRAS, 464, 4176
 Yan, R., Chen, Y., Lazarz, D., et al. 2019, ApJ submitted, arXiv:1812.02745
 Zwicky, F., Gates, H. S., & Taylor, D. 1955, Report to Office of Naval Research, 1, 1