

# From stars to cells – harnessing the power of the crowd for research

**Helen Spiers**

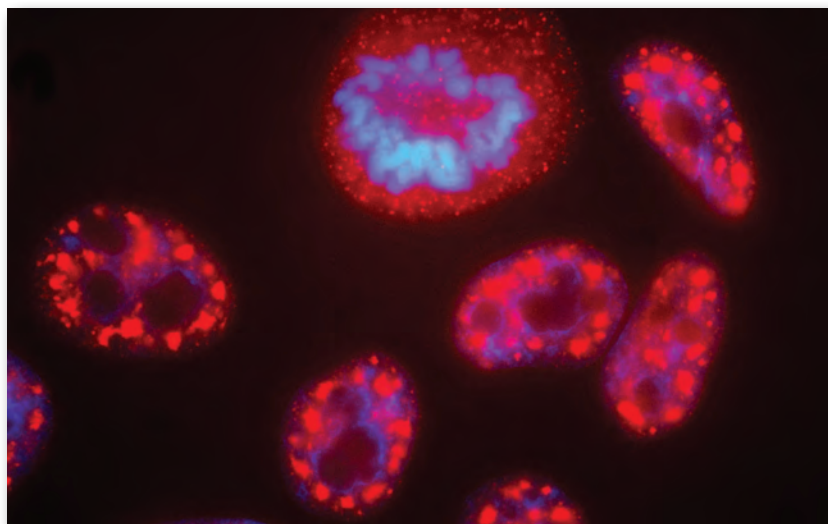
(University of Oxford, UK)

Modern research techniques allow more data to be generated than can be easily analyzed by the scientists who produce it. An original solution to this problem is to recruit volunteers to help with data analysis through online citizen science projects such as the Zooniverse.

The many technological innovations of the last century have culminated in a rising flood of data across diverse knowledge domains and academic disciplines. The volume of digital data generated is currently doubling every two years, with estimates predicting that there will be 44 trillion gigabytes of data globally by 2020.

Although such vast quantities of data provide a rich opportunity to garner novel insight and understanding, frequently, it is challenging to fully realize this potential. Often data analysis is hampered by inadequate or unavailable meta-data; too little is known about the data gathered and basic tagging and characterization is needed. Many computer algorithms have been generated to automatically perform data characterization tasks, however, such algorithms often remain outperformed by human-pattern recognition abilities.

Human time and effort is still necessary to adequately characterize many data types; but data is now being produced in such large volumes that analysts and researchers can't make sense of it alone. One innovative solution to this problem, that is gaining traction across multiple academic disciplines and domains, is to perform distributed data analysis through online citizen science.



Peering into the Cell Nucleus. Credit: Steve Mabon, Tom Misteli, NCI Center for Cancer Research, National Cancer Institute, National Institutes of Health. *Please note: this image is not part of the Zooniverse project*

## The origins of the Zooniverse

It is now over ten years since the first Zooniverse project, Galaxy Zoo, asked for public help with the morphological classification of galaxy images. The unexpected success of this project, which received more than 70,000 classifications per hour for its first few days following launch, led to the application of this approach to other astronomical research projects, including Solar Stormwatch and Moon Zoo, and eventually to the establishment of the Zooniverse platform.

## The expansion of online citizen science

Since these early astronomical origins, the Zooniverse has become the world's largest and most popular online citizen science platform with a community of more than 1.6 million registered volunteers contributing to over 110 research projects across multiple domains. The diverse assortment of projects currently on the Zooniverse platform range from transcribing documents written by Shakespeare's contemporaries in Shakespeare's World to aiding the study of climate change in Weather Rescue.

Although broad in research question and goal, these projects share the common principle of taking large data sets and asking volunteers to perform simple data-characterization tasks. Because we can ask more than one volunteer to look at each image, we can attain a level of accuracy greater than that achievable through the efforts of a single 'expert'. Not only does the application of online citizen science methodology result in large and accurate datasets that would be challenging to produce by other means, the process of having multiple volunteers examining each data point can result in unexpected, serendipitous discoveries; for example, a new class of galaxy was identified by volunteers in Galaxy Zoo, and the first planet in a four-star system was found through the efforts of volunteers contributing to the Planet Hunters project.

## Exploring new frontiers

Online citizen science methodology is being adopted by an increasing number of academic communities. One arena that is currently seeing a large amount of growth

in the number and variety of projects is the biomedical research community. Over the last year, the Zooniverse has launched multiple novel projects in this area, including Etch A Cell and Bash the Bug, with many more projects in the development pipeline.

Although developed by and hosted on the Zooniverse platform, these projects address very different research goals. For example, Etch A Cell, which is a collaboration with the Francis Crick Institute, seeks to improve methods for analysing images produced through electron microscopy. In this project, volunteers are asked to perform the manual segmentation of electron micrographs with a drawing tool, in a task very similar to that which would be performed by the researchers themselves. The data produced by the efforts of volunteers on the Zooniverse platform will be used to advance the automation of image segmentation, which is currently a significant bottle-neck in this area.

Another Zooniverse biomedical research project launched this year is the award-winning Bash the Bug. This project is part of the large-scale international CRyPTIC consortium, which has the goal of improving both the diagnosis and treatment of Tuberculosis (TB). TB is currently responsible for more deaths each year than HIV/AIDs, and like all bacterial disease, Mycobacterium tuberculosis is evolving resistance to the antibiotics used to treat it. The CRyPTIC project seeks to advance understanding of how genetic variance in TB influences susceptibility to treatment with different antibiotics. The knowledge of which genetic mutations confer antibiotic resistance will allow more tailored treatment of patients, through the genetic sequencing of the TB strain they are infected with. Zooniverse volunteers are helping with this effort through classifying which antibiotic dose is effective at killing each strain through examining images from 96-well plates. During the six-months following launch, this project has received nearly half a million classifications.

## Engaging communities with your own Zooniverse project

Not only does online citizen science provide a means of collectively accelerating research, it also provides a simple and accessible means for anyone with an internet connection to make a contribution to authentic scientific research. Such increased participation with science can benefit professional and non-professional researchers alike, through enabling research that wouldn't be possible otherwise and providing an opportunity for further education, and encouraging deeper engagement with science. ■

*If you would like to build your own Zooniverse project, please visit the Project Builder interface where it is possible to build a project in minutes. [www.zooniverse.org/lab](http://www.zooniverse.org/lab)*



*Dr Helen Spiers is a Postdoctoral Associate in biomedical research and citizen science at the University of Oxford. She is also the Biomedical Research Lead within the international research group responsible for the Zooniverse, the world's largest and most popular platform for online citizen science. To date, over 110 research projects across multiple academic disciplines have been launched on the Zooniverse, allowing anyone with an internet connection to make an authentic contribution to real research, and for researchers to do studies that would not be possible otherwise. Email: [helen.spiers@physics.ox.ac.uk](mailto:helen.spiers@physics.ox.ac.uk).*

## Further Reading

- Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S. P., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A. S., and Andreescu, D. (2008) Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, *Monthly Notices of the Royal Astronomical Society* **389**, 1179-1189.
- Woodcock, J., Greenhill, A., Holmes, K., Graham, G., Cox, J., Young Oh, E., and Masters, K. (2017) Crowdsourcing Citizen Science: Exploring the Tensions Between Paid Professionals and Users, *Peer Production and Work*.
- Cox, J. (2015) How is success defined and measured in online citizen science? A case study of Zooniverse projects.
- Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., and Parrish, J. K. (2014) Citizen science. Next steps for citizen science, *Science* **343**, 1436-1437.
- Kim, J. S., Greene, M. J., Zlateski, A., Lee, K., Richardson, M., Turaga, S. C., Purcaro, M., Balkam, M., Robinson, A., Behabadi, B. F., Campos, M., Denk, W., Seung, H. S., and EyeWriters. (2014) Space-time wiring specificity supports direction selectivity in the retina, *Nature* **509**, 331-336.
- Kwak, D., Kam, A., Becerra, D., Zhou, Q., Hops, A., Zarour, E., Kam, A., Sarmenta, L., Blanchette, M., and Waldispuhl, J. (2013) Open-Phylo: a customizable crowd-computing platform for multiple sequence alignment, *Genome Biol* **14**, R116.
- Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popovic, Z., and Players, F. (2010) Predicting protein structures with a multiplayer online game, *Nature* **466**, 756-760.
- Khatib, F., DiMaio, F., Foldit Contenders, G., Foldit Void Crushers, G., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popovic, Z., Jaskolski, M., and Baker, D. (2011) Crystal structure of a monomeric retroviral protease solved by protein folding game players, *Nat Struct Mol Biol* **18**, 1175-1177.
- Candido Dos Reis, F. J., Lynn, S., Ali, H. R., Eccles, D., Hanby, A., Provenzano, E., Caldas, C., Howat, W. J., McDuffus, L. A., Liu, B., Daley, F., Coulson, P., Vyas, R. J., Harris, L. M., Owens, J. M., Carton, A. F., McQuillan, J. P., Paterson, A. M., Hirji, Z., Christie, S. K., Holmes, A. R., Schmidt, M. K., Garcia-Closas, M., Easton, D. F., Bolla, M. K., Wang, Q., Benitez, J., Milne, R. L., Mannermaa, A., Couch, F., Devilee, P., Tollenaar, R. A., Seynaeve, C., Cox, A., Cross, S. S., Blows, F. M., Sanders, J., de Groot, R., Figueroa, J., Sherman, M., Hooning, M., Brenner, H., Holleczer, B., Stegmaier, C., Lintott, C., and Pharoah, P. D. (2015) Crowdsourcing the General Public for Large Scale Molecular Pathology Studies in Cancer, *EBioMedicine* **2**, 681-689.
- Kawrykow, A., Roumanis, G., Kam, A., Kwak, D., Leung, C., Wu, C., Zarour, E., Phylo, p., Sarmenta, L., Blanchette, M., and Waldispuhl, J. (2012) Phylo: a citizen science approach for improving multiple sequence alignment, *PLoS One* **7**, e31362.
- Tsueng, G., Nanis, S. M., Fouquier, J., Good, B. M., and Su, A. I. (2016) Citizen Science for Mining the Biomedical Literature, *Citizen Science: Theory and Practice* **1**.
- Roskams, J., and Popovic, Z. (2016) Power to the People: Addressing Big Data Challenges in Neuroscience by Creating a New Cadre of Citizen Neuroscientists, *Neuron* **92**, 658-664.