

## An Introduction to the Zooniverse

**Arfon M. Smith, Stuart Lynn, Chris J. Lintott**

Adler Planetarium, 1300 S Lakeshore Drive, Chicago, IL 60605  
team@zooniverse.org

### Abstract

The Zooniverse (zooniverse.org) began in 2007 with the launch of Galaxy Zoo, a project in which more than 175,000 people provided shape analyses of more than 1 million galaxy images sourced from the Sloan Digital Sky Survey. These galaxy 'classifications', some 60 million in total, have subsequently been used to produce more than 50 peer-reviewed publications based not only on the original research goals of the project but also because of serendipitous discoveries made by the volunteer community.

Based upon the success of Galaxy Zoo the team have gone on to develop more than 25 web-based citizen science projects, all with a strong research focus in a range of subjects from astronomy to zoology where human-based analysis still exceeds that of machine intelligence. Over the past 6 years Zooniverse projects have collected more than 300 million data analyses from over 1 million volunteers providing fantastically rich datasets for not only the individuals working to produce research from their projects but also the machine learning and computer vision research communities.

The Zooniverse platform has always been developed to be the 'simplest thing that works', implementing only the most rudimentary algorithms for functionality such as task allocation and user-performance metrics. These simplifications have been necessary to scale the Zooniverse so that the core team of developers and data scientists can remain small and the cost of running the computing infrastructure relatively modest. To date these simplifications have been acceptable for the data volumes and analysis tasks being addressed. This situation however is changing: next generation telescopes such as the Large Synoptic Sky Telescope (LSST) will produce data volumes dwarfing those previously analyzed. If citizen science is to

have a part to play in analyzing these next-generation datasets then the Zooniverse will need to evolve into a smarter system capable for example of modeling the abilities of users and the complexities of the data being classified in real time.

In this session we will outline the current architecture of the Zooniverse platform and introduce new functionality being developed that should be of interest to the HCOMP community. Our platform is evolving into a system capable of integrating human and machine intelligence in a live environment. Data APIs providing realtime access to 'event streams' from the Zooniverse infrastructure are currently being tested as well as API endpoints for making decisions about for example what piece of data to show next to a volunteer as well as when to retire a piece of data from the live system because a consensus has been reached.