

Crowdsourcing platform gets results

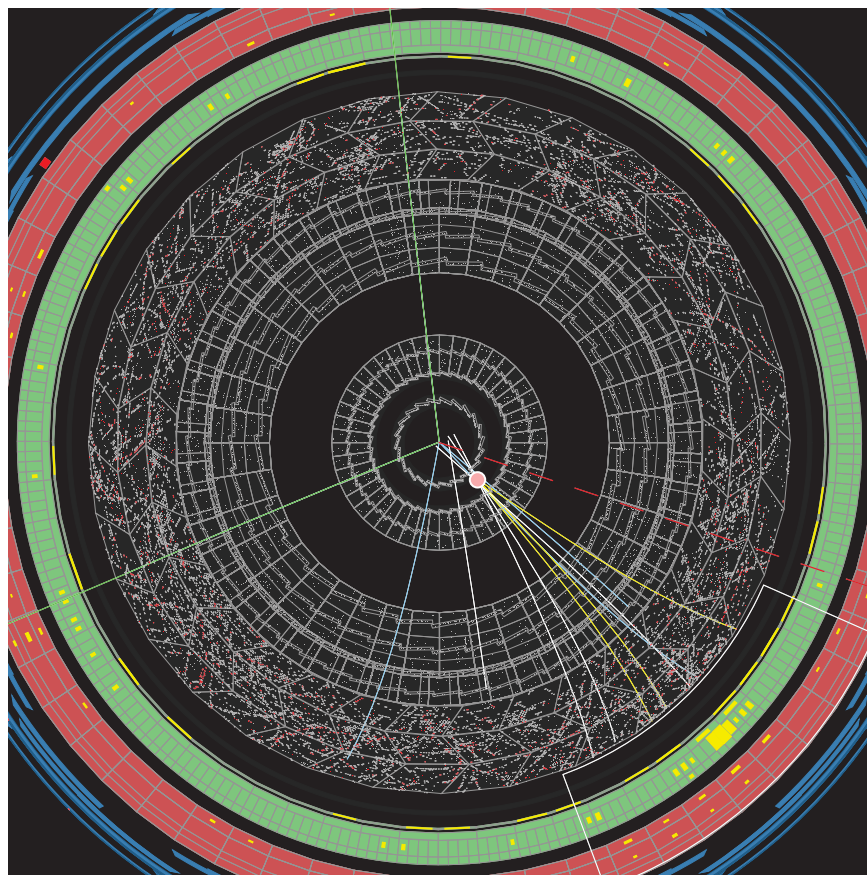
On the Zooniverse, data-analysis projects mesh public outreach and scientific discovery.

Within a day of being invited in 2007 to help classify galaxies by shape, volunteers were generating 70 000 classifications per hour—far above the 50 000 that a single graduate student had accomplished in one marathon week. “We were enormously surprised by this success,” says Oxford University’s Chris Lintott, who led the Galaxy Zoo project. “We realized there were lots of people who wanted to help.” Out of that experience grew the Zooniverse, an online platform where researchers post projects and invite “citizen scientists” to help analyze data.

The Zooniverse officially launched in late 2009 with grants from the UK and US totaling roughly \$4 million, and it now hosts more than three dozen interactive projects across many disciplines. Volunteers can hunt for fossils, identify individual humpback whales, or look for gravitational lenses, to name a few of the research projects on the site. Confidence in the results comes from redundancy: Typically 10 to 20 people are asked to look at a given image. As of early December, Zooniverse results had generated 87 peer-reviewed publications, plus more than a dozen studies on crowdsourcing in science.

Crowdsourcing analysis is best suited to projects that entail both huge amounts of data and tasks that humans outperform computers on. In particular, humans are better at pattern recognition. “All sorts of parameters go into the identity of a given galaxy,” says University of Minnesota astrophysicist Lucy Fortson, a cofounder of the Zooniverse and current board chair of its oversight organization, the Citizen Science Alliance. “The arms can be tightly bound or wide open, the image can be faint or not. So it’s difficult for a machine to say with certainty that a galaxy is spiral.” And, she adds, humans are good at spotting outliers. “They have the ability to ask the question, What the hell is this? Machines don’t have that ability.”

Until last summer, the Zooniverse team—about 30 people distributed



ATLAS COLLABORATION AND HIGGS HUNTERS ZOONIVERSE PROJECT

IN THE HIGGS HUNTERS PROJECT, citizen scientists look through images of proton collisions from the Large Hadron Collider’s ATLAS experiment at CERN. The volunteers are asked to mark tracks that begin away from the center, and to note the number of branches emanating from such vertices—eight in this example. The elusive jackpot would be a Higgs boson caught decaying into new particles.

among the University of Oxford, the Adler Planetarium in Chicago, and the University of Minnesota—put together project sites using information and data provided by researchers. But demand was high: Several proposals poured in each week and projects typically take months to build. So in July the Zooniverse team debuted a build-it-yourself tool.

The team continues to improve the build-it-yourself project capabilities. And with an eye to keeping up with the ever-growing onslaught of scientific data, the Zooniverse team is working to use its results from crowdsourcing to train computers to do more of the data analysis.

130 000 collaborators

Project discussion boards turn out to be where the crowd makes discoveries. On

the Milky Way project, for example, volunteers discovered nascent star clusters much younger than the ones the researchers had been looking for in false-color IR images from the *Spitzer* telescope. After a volunteer posted about compact bright yellow globs, the discussion prompted the researchers to add a new identification category. “Had the volunteers not started discussing the yellow balls, or had the color scheme been different, they may never have been identified,” says the Adler Planetarium’s Grace Wolf-Chase, a scientist on the project. Zooniverse volunteers have also discovered planets, other astronomical objects, and a new species of ocean worm.

Meg Schwamb, a postdoctoral fellow in Taiwan at the Academia Sinica’s Institute of Astronomy and Astrophysics, is

involved in several Zooniverse projects. As she sees it, she has 130 000 collaborators—and that's just on the Planet Four project, in which volunteers characterize seasonal patterns on the surface of Mars, such as the fan-shaped regions shown in the image at right. "By having so many people looking at the data and asking, 'Is that weird? Is that interesting?'" you pull out these rare gems you wouldn't otherwise have known existed."

Schwamb also works on Planet Hunters. That Zooniverse project has attracted around 300 000 volunteers to check light curves for dips in star brightness that may indicate a planetary transit. Many of the other projects have "gorgeous pictures of galaxies or animals," she notes, "but with Planet Hunters, we give them graphs. We showed that people are doing it to contribute to science."

Some volunteers stop by the Zooniverse website, analyze a few images, plots, or audio files, and then move on. Others get hooked on a particular project. Michael Purves, a retired meteorologist in Victoria, British Columbia, for example, spends several hours a day on a project called Old Weather. He goes through naval logs to recover weather conditions and to extract from handwritten annotations events of interest about battles, people, animal sightings, and the like. He has worked on the logs from a British ship during World War I and another that sailed up and down the Yangtze River in the 1930s; he is now focusing on a US Navy gunboat from the Spanish-American War. The historical information is collected by one set of researchers, and the weather data is for climate models.

BY CATEGORIZING SPECTROGRAMS from the Laser Interferometer Gravitational-Wave Observatory, citizen scientists will help train machines to better distinguish astrophysical signals from noise such as the beat note seen in this plot of frequency versus time.



IN PLANET FOUR, volunteers help study Martian climate and wind patterns by mapping dark seasonal fans near Mars's south pole in images from NASA's High Resolution Imaging Science Experiment (HiRISE) camera. Fan-shaped regions show up in spring and summer and are believed to be from dirt-carrying carbon dioxide geysers that rise a meter or so above the ice cap and fan out due to winds.

Volunteering on Zooniverse projects is "mentally stimulating without the pressures of work," says Joan Arthur, an administrator at the University of Oxford who volunteers as a moderator on Old Weather and also contributes to Penguin Watch, where volunteers mark penguins in images to help monitor changes in the Antarctic ecosystem. "For me, it's the fora that make the difference. Enjoying others' company, with the same purpose in mind is just a wonderful way of spending my time," she says. Arthur puts in upwards of 30 hours a week on the Zooniverse.

Of the 1.4 million registered users, about 100 000 are active volunteers,

says Grant Miller, Zooniverse liaison between researchers, project developers, and volunteers. Of the active users, about one-third are in the UK, one-third are in the US, and the others span the globe. The age range and level of education of volunteers are broad, encompassing children, students, working people, and retirees. According to Zooniverse surveys, the driving motivations for volunteers are a genuine interest in the research and the satisfaction of contributing to it.

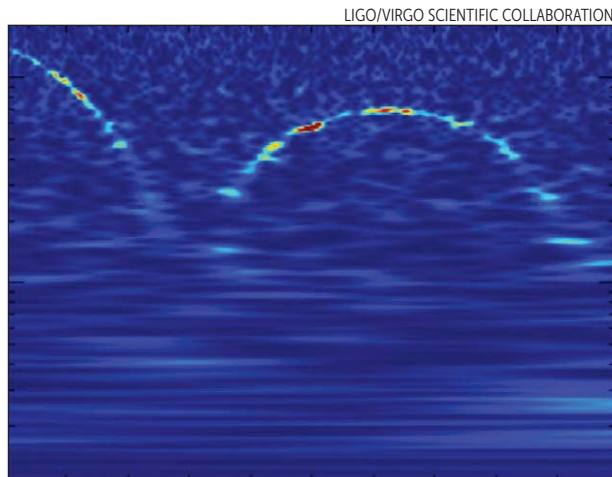
Build it yourself

With the build-it-yourself option, projects can go online quickly—they can be built in hours or days, depending on the project and whether scientists are prepared with their research descriptions, tutorials, and data. In the first five months that the option was available, more than 1300 project attempts were made, according to Miller. Many researchers just explore the Zooniverse's build-it-yourself process, he says, but to date 40 or 50 are "serious attempts at working projects."

To be hosted on the Zooniverse website, projects must have clear-cut tasks and science goals, and must require analysis for which humans do a better job than computers. Once accepted, projects get the benefits of the platform's visibility, a newsletter that alerts registered users, and occasional high-profile launch events by, for example, the BBC. Typically, says Miller, tens of thousands of visitors will check out a new project in the first few days.

In converting to a build-it-yourself model, not only can the Zooniverse platform accommodate more projects—it was previously limited by money for paying project developers—but do-it-yourselfers can build projects for use by their own internal groups without posting to the public. That feature may be attractive to protect privacy if medical images, national security projects, or other proprietary data are involved.

Looking ahead to larger volumes of data being collected at faster rates, the Zooniverse team is focusing on machine learning. Even with crowdsourcing, humans won't be able to handle the data from, for example, the Large Synoptic



Survey Telescope, which every night will generate terabytes of data and send several billion alerts to other telescopes for follow-up observations.

Between a rock and a penguin

An early, ongoing foray into machine learning on the Penguin Watch project involves feeding the crowd-sourced findings back to computers to train them to distinguish penguins from rocks. Another example in the works involves data from the Laser Interferometer Gravitational-Wave Observatory. Even when the upgraded LIGO experiment reaches full sensitivity (see PHYSICS TODAY, September 2015, page 20), a true signal is predicted to occur only about once a month, whereas glitches—signals from anomalous sources—are expected to be picked up hundreds of times a day. A lot can be vetoed by computers, says LIGO scientist Shane Larson of the Adler Planetarium and Northwestern University, “but there is a gray area, where the computer can’t determine if something is signal or noise.”

That’s where the Zooniverse comes in. “We are building a circular pathway from data to citizens to see if they can help classify the glitches,” says Larson. Starting later this year, the results from volunteers will be fed back to computers to try to teach them to recognize patterns as effectively as people. To protect the LIGO team’s dubs on discovery, data for the Zooniverse project will not be time-stamped, and volunteers will see a given signal from only one of LIGO’s two interferometers; data from both are necessary to identify an actual gravitational wave.

For more general application, the University of Minnesota’s Fortson and colleagues aim to train computers by having them work on the same data at the same time as humans do. A key example, she says, is identifying new classes of objects, “which humans are really good at, but where machines fail.” By having humans and computers pass data back and forth, the machine training will be faster, she says, and computers will be able to handle more types of data.

The goal is for computers to learn which tasks should be done by humans and which by computers, says Laura Trouille, Adler Planetarium’s head of citizen science and Zooniverse codirector. “Only in combination do we have a chance of optimizing data analysis.”

Toni Feder

Projected Field Electromagnets

Open access to the surface of thin film samples for laser, synchrotron radiation, electrical, or microscope probes.

5201 - In-Plane Projected Field

Peak Field, B_x:

>0.3T

5203 - Vertical Projected Field

Peak Field, B_z:

>0.5T over Ø2mm area;

>0.8T at a single point

5204 - Vector Projected Field

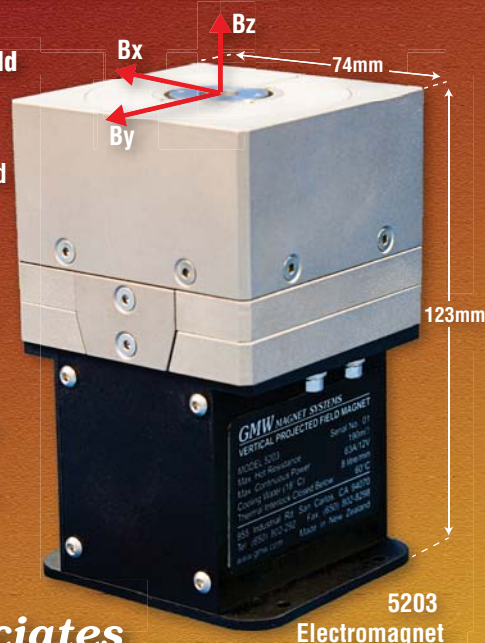
Peak Field, any vector:

>0.3T

5205 - Vector Projected Field

Peak Field, B_z:

>25mT over Ø40mm area



GMW Associates
www.gmw.com • sales@gmw.com

Impact a broader audience



Computing in Science & Engineering (CISE) appears in the IEEE Xplore and AIP library packages, representing more than 50 scientific and engineering societies. Why publish in a journal that serves only one society or scientific field?

SUBMIT AN ARTICLE

computer.org/web/peer-review/magazines
George K. Thiruvathukal, gkt@cs.luc.edu