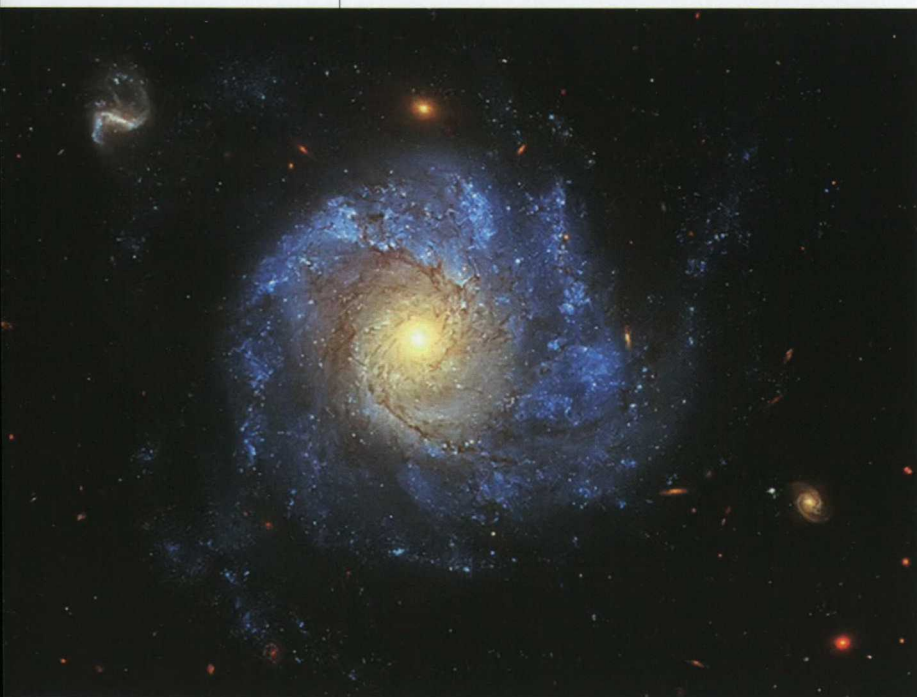# THE CITIZEN AS SCIENTIST

Not enough computing power for data processing? A ridiculous number of man-hours needed to visually inspect collections of raw data? Scientists facing such problems can seek help from volunteers equipped only with computers connected to the Internet. In other words, from practically everybody.

Photo 1:
Galaxy Zoo engaged volunteers to classify such objects as galaxies using images of the sky. This photograph shows a Hubble Space Telescope view of the galaxy NGC-1309; August-September 2005

Dr. Olek Michalski

S tatistical techniques are good at detecting regularities in data sets under analysis, but not so good at handling very rare events or situations which we suspect might follow certain patterns, but we have no good idea of what those patterns might look like until we find them. For example, we face such problems when combing through radio signals in the universe in search of signs of transmissions from extraterrestrial intelligences.

The very concept of establishing radio communication with extraterrestrial civilizations is as old as the history of radio. Even Nikola Tesla, Guglielmo Marconi, and Lord Kelvin suggested that radio technology could be used to contact inhabitants on the planet Mars. Although the way we imagine extraterrestrial intelligence has changed radically since that time, the belief that it is possible to receive transmissions from potential extraterrestrials lingers on. Listening for extraterrestrial signals began in the United States and in the USSR in the 1960s, with developments in radio astronomy leading to a dramatic increase in the amount of data in projects referred to as SETI (Search for Extraterrestrial Intelligence).

However, analyzing such data requires an ever-growing amount of computing power. In the mid-1990s, researchers from the University of California, Berkeley, pointed out that personal computers connected to the Internet could offer an alternative. Launched in May 1999, the project SETI@home allows users to install software on their home computers, which harnesses their unused computing power to analyze signals from the Arecibo radio telescope.

SETI@home proved extremely popular: 400,000 users volunteered to join the project within the year before the first version of the software was released and 200,000 actually downloaded and ran the program one week after its release. Although no spectacular success was reported (i.e. no signal coming from extraterrestrial intelligence was detected), the project paved the way for similar initiatives. Scientists have enlisted Internet users to study a variety of problems, ranging from the simulation of high-energy particle collisions (LHC@home), which serve as a point of reference for actual measurements at the Large Hadron Collider (LHC), to the modelling of protein-protein interactions (Rosetta@home) to help design new drugs.

Gazing into the heavens

In such computing projects, the help provided by volunteers is extremely important, but their involvement is solely limited to offering access to their computers. More often than not, however, conscious human participation proves necessary, usually when the need arises to discover objects whose characteristics vary or prove difficult to describe. A good example is offered by the images of almost 1 million galaxies captured as part of the Sloan Digital Sky Survey, conducted at Apache Point Observatory in New Mexico. Kevin Schawinski, who classifies galaxies based on their images, described the task as "mind-numbing" after just one week. Consequently, he decided to enlist volunteers. Together with Chris Lintott from Oxford University, Schawinski created a website that enabled volunteers to help classify galaxies captured on the images. He hoped to engage 20,000-30,000 volunteers. Ultimately, however, the project attracted over 100,000 users who classified 40 million objects in half a year.

Volunteers completed the task faster than expected. In addition, they quickly started to spot various irregularities, which on closer inspection turned out to be new astronomical objects. Such contributions proved invaluable to scientists.

Not just galaxies

The project, named Galaxy Zoo, proved so successful that it spun off similar extended-version projects and prompted Lintott to create the platform Zooniverse, where volunteers can choose from an entire spectrum of projects, from inspecting particle trajectories registered at the LHC, through describing video recordings from camera traps in a jungle, all the way to helping digitize and analyze First World War unit diaries. All these tasks involve searching for and classifying visual patterns – the kind of thing humans are still much better at than computers.

But does this exhaust the possibilities for putting volunteers to work in the service of science?

One of the most serious problems facing biomedical sciences is our limited knowledge of the structures of proteins. Although it is easy to identify the sequence of amino acids, or the building blocks, that make up a given protein, we have difficulties determining the physical shape this chain takes on through a process known as protein folding, and that shape is what largely determines its function. The number of possible shapes is very high, because the chains are long and very flexible. Checking all possible combinations one by one to select the most likely shape takes a lot of time despite the employment of sophisticated computational methods. One of the attempts to resolve this problem involves the game Foldit, which relies on the human ability to find solutions to spatial problems.

In Foldit, players can fold proteins themselves: they see the model of a molecule and can manipulate its shape using various tools, scoring points for the quality of spatial conformations. Since May 2008, over 57,000 users have taken on the challenge and they have been very successful. For example, their efforts helped scientists unlock the structure of the Mason-Pfizer monkey virus retroviral protease (which causes the monkey equivalent of AIDS), which is crucial to the functioning of the virus, after more than a decade of unsuccessful attempts to determine its shape. Despite these encouraging results, the creators of Foldit still have the same overarching goal: namely to find out whether humans' pattern-recognition and puzzle-solving abilities make them more efficient than existing computer programs at pattern-folding tasks, and if so, to try to teach these techniques to computers.

A revolution in the world of science?

Despite the considerable assistance that comes from it, the participation of volunteers in "doing science" raises doubts, especially as to the accuracy of the findings. Andrew Westphal, who enlisted volunteers in the project Stardust@home (researching small particles in outer space), observed that not all participants were equally reliable. Some volunteers would simply open files without actually analyzing their content in order to move up in the ranking of scores. And so, the researchers introduced a mechanism to evaluate the work being done by volunteers, and to determine the minimum number of users that must flag a sample before it can be accepted. The creators of Galaxy Zoo, on the other hand, assert that their findings are reliable, despite noticing a certain systematic bias in the classification of spiral galaxies as spinning clockwise or counterclockwise, which resulted from the specific characteristics of human perception. In order to quantify this error, they presented certain data to volunteers in the form of mirror images.

Using the work of volunteers also poses a certain ethical problem. The authors of an article recently published in the *Journal of Medical Ethics* point out the risk of addiction to participation in computational projects and demand that such projects should be supervised by bioethics committees. Most opinions, however, are less skeptical, seeing such "citizen science" projects as ways not only to resolve the problem of insufficient computing power and imperfect algorithms but also to create a bridge between the world of science and the public at large, in addition to engaging young people. ■

Further reading:

www.setiathome.ssl.berkeley.edu – Seti@home

www.galaxyzoo.org – Galaxy Zoo

http://fold.it – Foldit

E. Hand (2010). Citizen science: People power. *Nature* 466 (7307), 685–7. doi:10.1038/466685a, PMID 20686547.

Khatib F., DiMaio F., Foldit Contenders Group et al. (2011). Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural and Molecular Biology* 18, 1175-1177.