PROJECT NOAH/DAN DOUCETTE

Some contributors to citizen-science initiatives, such as Project Noah, sport a tattoo of the project's logo.

CITIZEN SCIENCE

Computer sharing loses momentum

Competition and education needed to keep people engaged.

BY NICOLA JONES

The family of '@home' volunteer computing projects is growing ever more diverse. Spare time on a personal computer can now be donated to anything from finding alien life to crunching climate models or processing photos of asteroids. But enthusiasm is waning. The 47 projects hosted on BOINC, the most popular software system for @home efforts, have 245,000 active users among their 2.7 million registrants, down from a peak of about 350,000 active users in 2008 (see 'Slumping @home').

David Anderson, the founder of BOINC (Berkeley Open Infrastructure for Network Computing) and a computer scientist at the University of California, Berkeley, has several explanations for the slip. He says media coverage has declined now that volunteer computing is more than 15 years old. A shift to mobile-computing devices has probably also hurt — BOINC can run on an Android phone while charging, but uses too much battery power when unplugged. And the site has been unable to attract a broad demographic of volunteers.

"Essentially, we have a bunch of middle-aged, male computer nerds," says Anderson. "We have thought long and hard about ways to break out of that category, using Facebook, for example, but none of that has been all that successful."

On 20–22 February, at the 3rd Citizen Cyberscience Summit in London, conference-goers will trade tips on how to entice volunteers into projects ranging from BOINC-style distributed computing to more-active 'citizen-science' projects, in which users are asked to donate not just their time but also their brains.

The desire to keep numbers up is not just academic. If distributed computing flourishes, serious money can be saved, says Francois Grey, coordinator of the Citizen Cyberscience Centre, based in Geneva, Switzerland. He notes that the Chinese Academy of Sciences in Beijing has been monitoring the economic benefits of CAS@home, which uses volunteers' computing time for projects such as predicting protein structures. The academy estimates that US$20 million has been saved since it launched CAS@home in September 2010, by using donated computing power rather than buying it from a company such as Amazon.

Grey predicts that funding bodies might at some point enforce the use of volunteer computing whenever possible, rather than allowing grant money to be used for supercomputer time or cloud-based services. "It's very delicate. There are big IT companies with vested interests in selling supercomputers to universities," he says. "But I think it's something that will happen at some point."

For volunteer computing to be used in a bigger way, participation rates need to keep up. Perhaps the most obvious motivator — money — is deemed a bad idea. "Small amounts of money are too trivial, and may be almost insulting," says Grey. "It goes against the idea of volunteering." Only one BOINC project — IBM's World Community Grid, an umbrella initiative that oversees a batch of biomedical projects aimed at goals such as drug discovery — has partnered with a scheme that allows volunteers to earn virtual cash (which can be exchanged for real money) for their time. This had a measurable but small overall impact, says Anderson, earning the grid as many as 15,000 new volunteers, bringing the total so far up to almost 650,000.

A more powerful motivator is pleasure. This can be achieved by turning participation into a game. FoldIt, for example, asks volunteers to optimize protein folding, which requires a mix of intellect and intuition that some describe as similar to chess. Competition can also provide pleasure. Many projects offer scoreboards and awards such as virtual titles or badges to mark progress; some people have become so devoted that they have had the badges tattooed on their bodies. In the BOINC world, groups of volunteers have formed teams that compete to donate the most time over a designated period. These competitions offer a short-term boost, but the effect wears off, says Anderson.

Engaging participants in the core science mission is by far the best motivator, says Oded Nov, who studies links between new technologies and human behaviour at New York University. That includes giving participants credit in scientific papers and showing them how their help is advancing research. The World Community Grid, for example, hosts regular Q&A sessions with its project scientists. "Education is a great motivator," says Nov.

That could be one reason why the Zooniverse — the largest host of citizen-science schemes — has not seen a decline in participation. Its family

SLUMPING @HOME

The past several years have seen a decline in the number of active users in the BOINC family of volunteer computing projects.

WILLY DE ZUTTER, BOINCSTATS

of 22 projects asks volunteers to do everything from identifying galaxy types in astronomical images to transcribing historical weather records. Robert Simpson, a developer and head of communications for the Zooniverse team, says that the five-year-old scheme has 930,000 registered participants and that there is fairly consistent interest in new projects.

Quantifying the effects of different motivational tools is difficult, says Grey, whose cyberscience centre has received funding to explore the possible benefits of common rules and credit schemes across different platforms such as BOINC and the Zooniverse. "Because of its grass-roots nature, everyone's doing their own thing; there's no common metric," he says.

One thing is certain: there is still plenty of spare brainpower to access. "US citizens alone spend 200 billion hours watching television a year," says Simpson. "We only need to tap a tiny fraction of that." ■

PUBLISHING

Elsevier opens its papers to text-mining

Researchers welcome easier access for harvesting content, but some spurn tight controls.

BY RICHARD VAN NOORDEN

Academics: prepare your computers for text-mining. Publishing giant Elsevier says that it has now made it easy for scientists to extract facts and data computationally from its more than 11 million online research papers. Other publishers are likely to follow suit this year, lowering barriers to the computer-based research technique. But some scientists object that even as publishers roll out improved technical infrastructure and allow greater access, they are exerting tight legal controls over the way text-mining is done.

A few years ago, scientists complained that publishers were stymieing ambitious plans to use computer software to pull out information from published papers. Some researchers who ran software to harvest data from online articles found their programs blocked, and those who asked for permission found themselves trapped in tortuous case-by-case negotiations — even though they had already paid subscription fees for access. Max Haeussler, a computational biologist at the University of California, Santa Cruz, for instance, spent more than three years arguing with publishers for permission to extract DNA data from 3 million articles to annotate an online map of the human genome (see *Nature* **483,** 134–135; 2012).

"It was a legitimate criticism, that people sent text-mining requests in to publishers and they bounced around for a time without any response," admits Chris Shillum, vice-president of product management for platform and content at Elsevier. The publisher previously considered requests "case by case", he says — but it now wants to make text-mining permissions quicker and easier to obtain. "What we've tried to do is take the practical barriers away."

Under the arrangements, announced on 26 January at the American Library Association conference in Philadelphia, Pennsylvania, researchers at academic institutions can use Elsevier's online interface (API) to batch-download documents in computer-readable XML format. Elsevier has chosen to provisionally limit researchers to 10,000 articles per week. These can be freely mined — so long as the researchers, or their institutions, sign a legal agreement. The deal includes conditions: for instance, that researchers may publish the products of their text-mining work only under a licence that restricts use to non-commercial purposes, can include only snippets (of up to 200 characters) of the original text, and must include links to original content.

"Finally, someone is showing that there is no need to be afraid of text-mining analysis."

"Finally, someone is showing that there is no need to be afraid of text-mining analysis any more," says Haeussler.

Researchers working on the Human Brain Project — a European consortium that plans to use a supercomputer to recreate everything known about the human brain — have already used Elsevier's interface to do text-mining, says the project's spokesman Richard Walker, who is based at the Swiss Federal Institute of Technology in Lausanne. "We are very pleased with it. It resolves genuine technical issues," he says.

And neuroscientist Shreejoy Tripathy at the University of British Columbia in Vancouver, Canada, worked with Elsevier last year to pull out information on neuron physiology from thousands of articles (see neuroelectro.org). Text-mining is not yet well known, he says, but he hopes that the easier access will kick off its greater adoption among scientists. "As more papers get published that use text-mining, other researchers like myself — who are neuroscientists and not programmers — will see the need for the technique," he says.

Shillum says that Elsevier is ahead of the curve — but that other publishers are likely to follow soon. CrossRef, a non-profit collaboration of thousands of scholarly publishers, will in the next few months launch a service that lets researchers agree to standard text-mining terms and conditions by clicking a button on a publisher's website, a 'one-click' solution similar to Elsevier's set-up.

And, in the past year, large institutions and pharmaceutical companies have started to ask for text- and data-mining rights when renegotiating site licences, says Jessica Rutt, rights and licensing manager at Nature Publishing Group (NPG), the publisher of this journal. Anyone with those rights may mine NPG content. Many publishers are also experimenting with delivering text-minable content to pharmaceutical companies for an extra fee, she adds.

But some researchers feel that a dangerous precedent is being set. They argue that publishers wrongly characterize text-mining as an activity that requires extra rights to be granted by licence from a copyright holder, and they feel that computational reading should require no more permission than human reading. "The right to read is the right to mine," says Ross Mounce of the University of Bath, UK, who is using content-mining to construct maps of species' evolutionary relationships.

National governments are also weighing in on the issue. The UK government aims this April to make text-mining for non-commercial purposes exempt from copyright, allowing academics to mine any content they have paid for. And the European Commission, worried that barriers to computational research could hinder scientific innovation, is also examining the issue. It has convened a group chaired by Ian Hargreaves, an intellectual-property specialist at Cardiff University, UK, who recommended the changes to UK law, to examine the economic impact of text- and data-mining for scientific research and barriers to its use. The panel will reach conclusions by the end of February.

"Our plan is just to wait for the copyright exemption to come into law in the United Kingdom so we can do our own content-mining our own way, on our own platform, with our own tools," says Mounce. "Our project plans to mine Elsevier's content, but we neither want nor need the restricted service they are announcing here." ■