
Science wikinomics

Mass networking through the web creates new forms of scientific collaboration

A wiki, “from the Hawaiian meaning ‘quick’, is an online content management system to facilitate the collaborative editing of documents through an in-browser text interface, which are then immediately presented as publicly visible web pages” (<http://en.wikipedia.org/wiki/Wiki>). The definition comes straight from Wikipedia, the best known example of how wikis can be used to create collaborative websites and to power the communities around them. But beyond Wikipedia itself, an entire wiki world is flourishing that eases the sharing of knowledge, and is a world we have only just begun to explore.

An increasing share of the world’s most influential companies are already counting on ‘weapons of mass collaboration’ to foster innovative power, according to Don Tapscott and Anthony Williams, authors of the 2006 book, *Wikinomics. How Mass Collaboration Changes Everything*. “Individuals now share knowledge, computing power, bandwidth, and other resources to create a wide array of free and open source goods and services that anyone can use or modify,” they wrote. “All one needs is a computer, a network connection, and a bright spark of initiative and creativity to join in the economy” (Tapscott & Williams, 2006).

Innovative businesses are certainly embracing the new era of ‘wikinomics’ by

harnessing the knowledge and creativity of many thousands of external collaborators. As Rosabeth Moss Kanter at the Harvard Business School (Boston, MA, USA) noted, shared values, principles, and social and environmental responsibility are at the core of these new business structures, transforming them from “impersonal machines into human communities [who gain] the ability to transform the world around them in very positive ways” (Kanter, 2008). Yet, as Kanter remarked, the contributions of many dynamic parts must “add up to a unified purpose and accomplishment”—a collaborative concept that scientists have long understood.

When it comes to science, mass collaborations that take advantage of the ‘wiki’ world have been running for about a decade in the form of volunteer computing, in which members of the public donate their computing resources to the benefit of scientific progress. “The majority of the world’s computing power is no longer in supercomputer centers and institutional machine rooms. Instead, it is now distributed in the hundreds of millions of personal computers all over the world,” commented David Anderson—founder and Director of the Berkeley Open Infrastructure for Network Computing (BOINC; see Sidebar A) project—at a conference in 2003 (Anderson, 2003).

Sidebar A | A cradle for volunteer computing

Crucial to some of the most important projects based on volunteer computing is the World Community Grid (WCG), the largest public humanitarian grid in existence (www.worldcommunitygrid.org). Launched in 2004 by IBM (Armonk, NY, USA), WCG has run ten projects so far, five of which have been completed. “World Community Grid has more than 415,000 members and more than 1.1 million devices, but this is just a fraction of the estimated 1 billion PCs around the world that could be used,” commented Robin Willner. “The computing power needed for the research community is almost limitless.”

WCG acts as a useful tool to complete a certain stage of research such as complex calculations to hasten the progress of projects into further phases of development, explained Willner. “World Community Grid leverages infrastructure resources to help expedite calculations, normally requiring many months, and produces results in mere days. So the project must be one that could be grid-enabled.”

Prospective projects in medical, environmental and basic research areas are solicited through the WCG web site and selected by an Advisory Board consisting of senior IBM officials, as well as members from key foundations, academic institutions and public agencies. Willner explained that WCG support is only available to projects conducted by public and not-for-profit organizations, and that WCG gives priority to research that has the potential to assist economically disadvantaged communities and developing countries, or to provide initial data that could open new fields of inquiry. “All projects must serve to promote human welfare directly or indirectly by advancing knowledge in areas that contribute to the overall goal. Results must be made available to the world research community, and all research results and findings will be made available in the public domain,” Willner noted. “The goal is to assure that World Community Grid resources are focused on research that has the greatest impact and to support work that might otherwise be bypassed in favour of more commercial projects.”

All WCG projects run on the Berkeley Open Infrastructure for Network Computing (BOINC; <http://boinc.berkeley.edu>), a software platform for volunteer computing created by the University of California at Berkeley (CA, USA). BOINC was originally developed to support the SETI@home project—which seeks to detect intelligent life on other planets by analysing radio signal data—but has since been modified to accommodate research projects in a range of disciplines, from molecular biology to climate dynamics (Anderson, 2003; Anderson & Reid, 2009).

Inspired by the success of the first large projects SETI@home and Folding@home, launched in 1999 and 2000, respectively, volunteer computing has since thrived, allowing people to download, install and run software dedicated to an expanding range of research problems, thus contributing to scientific endeavour at next-to-no cost to themselves. “Routes to meaningful engagement are often difficult to identify, match, and integrate into our lives,” commented Dwayne Spradlin, President and Chief Executive Officer of InnoCentive (Waltham, MA, USA)—a company that aims to solve ‘big problems’ by connecting companies, academic institutions, public sector and non-profit organizations with the world’s brightest minds. “All too often, we are left with financial donations as the only currency for participation. [...] Lending our intellectual means in support of the efforts and challenges about which we care most may be the most fulfilling.”

The project FightAIDS@home, for example, is utilizing the collective computing power of its volunteers to “assist fundamental research [...] building on our growing knowledge of the structural biology of AIDS” (<http://fightaidsathome.scripps.edu/>) in order to design more effective anti-HIV

therapies. FightAIDS@home began with the help of a small start-up company called Entropia back in 2000—it is thus the first biomedical application to be supported by Internet distributed computing. In late 2005, the project switched to the IBM-backed (Armonk, NY, USA) World Community Grid (WCG; see Sidebar A) platform. “That was a great opportunity, since IBM had the wherewithal to not only host the project on much larger servers, but more importantly had the public relations organization to get the word out about FightAIDS@home to a much larger audience,” commented project leader Arthur Olson, Professor and Director of the Molecular Graphics Laboratory at the Scripps Research Institute (La Jolla, CA, USA). “Today there are over 1,000,000 processors signed up.”

Olson explained that the resources made available through WCG allowed his group to explore computationally intensive

When it comes to science, mass collaborations that take advantage of the ‘wiki’ world have been running for about a decade in the form of volunteer computing...

approaches to docking large libraries of compounds against large panels of HIV mutant proteases, searching for chinks in the armour of HIV that could be exploited by drug candidates—research that the group has been able to publish (Chang *et al*, 2007). “Our initial results have shown that there are a set of ‘spanning’ mutants that appear to characterize the mutational space that the HIV protease can explore and still be viable,” he said. “We are currently using new methods that allow us to represent the dynamic character of these protein targets in the docking process, giving us a better picture of how potential drug molecules may interact with the HIV protease” (Fig 1).

Another project, Help Defeat Cancer (HDC; <http://pleiad.umdj.edu/IBM/index.html>), ran on the WCG platform between July 2006 and April 2007, using the huge voluntary computational power for the high-throughput analysis of digitally imaged cancer tissue microarrays—in which separate tissue cores are combined for multiplex histological analysis. By using the idle time of volunteer’s computers, the project aimed to characterize protein expression patterns that could be used to reliably classify subtypes and stages of disease progression in breast, colon, head and neck cancer. The research team hoped that using an automated approach would allow for a more consistent evaluation of expression patterns, avoiding the error-prone ‘by eye’ evaluation of samples. “We think this is a really great example of how volunteer computing projects can speed results for researchers and have a real impact,” commented Robin Willner, Vice President of the IBM Global Community Initiatives.

Building on the results from the HDC project, the team, based mainly at The Cancer Institute of New Jersey (CINJ; New Brunswick, NJ, USA), recently received a US\$2.5 million grant from the US National Institutes of Health (NIH, Bethesda, MA, USA). The central objective of the new project is to build an expanded support system for the analysis and classification of imaged cancer specimens with improved accuracy, allowing researchers, physicians and scientists to prescribe tailor-made treatments for various cancers, based on how cancers with similar protein expression signatures have reacted to treatments in the past. “[The] World Community Grid enabled us to validate our imaging

and pattern recognition algorithms and establish a reference library of expression signatures for more than 100,000 digitally imaged tissue samples,” commented David J. Foran, Director of the Center for Biomedical Imaging & Informatics of CINJ and Principal Investigator of HDC, in a press release (IBM, 2008). “The overarching goal of the new NIH grant is to expand the library to include signatures for a wider range of disorders and make it, along with the decision support technology, available to the research and clinical communities as grid-enabled deployable software.”

Although lending your computer’s idle time to the service of science is certainly a worthy cause, some computer owners might feel that the projects on offer are rather unexciting—at least from a layperson’s point of view. Foldit (Fig 2), a brand new online protein folding ‘game’ (<http://fold.it>) seeks to address this lack of fun by making voluntary data analysis interactive. The project is essentially an evolution of Rosetta@home, a project set up by David Baker, Professor of Biochemistry at the University of Washington (Seattle, WA, USA), to predict the way in which a given protein folds into its unique three-dimensional shape. After running Rosetta for several years, and spurred on by the feedback received from volunteers, Baker felt that some human intervention could help to solve the folding challenge: “There are too many possibilities for the computer to go through every possible one,” Baker commented in a press release heralding the game’s launch last May (Hickey, 2008). “An approach like Rosetta@home does well on small proteins, but as the protein gets bigger and bigger it gets harder and harder, and the computers often fail. People, using their intuition, might be able to home in on the right answer much more quickly” (Hickey, 2008).

Baker teamed up with Zoran Popović from the Department of Computer Science and Engineering at the university, and together they developed the main framework of the game. Originally, the game used proteins the three-dimensional structures of which were known, but work progressed rapidly to predict the structures of proteins of unknown shape. The game itself involves solving a shape-based puzzle, relying on the participation of those individual ‘players’ with a natural ability to think in three dimensions to figure out which of the many possible protein

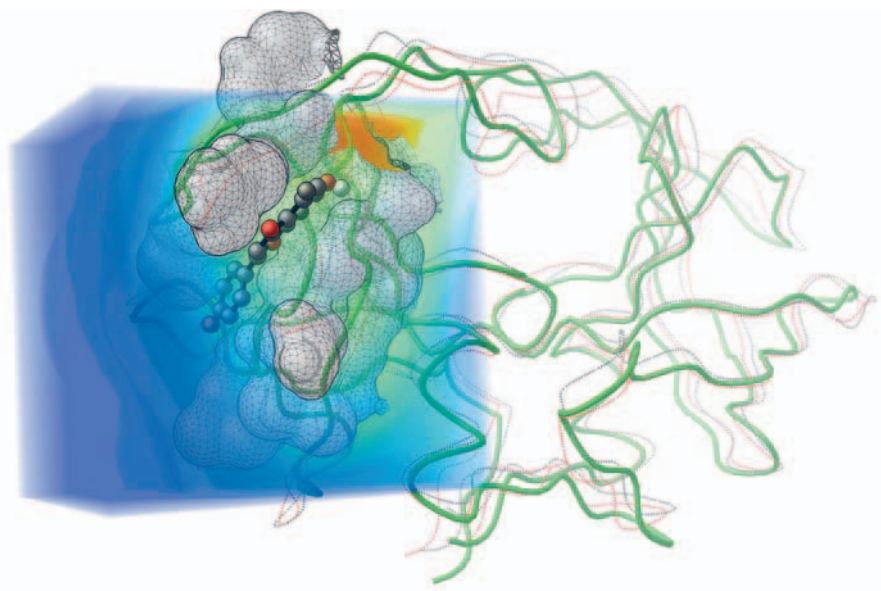


Fig 1 | FightAIDS@home results. The image depicts the HIV protease in three molecular dynamics snapshots, and a compound docking in an exo-site (distant from the active site). The cloud in the cubic volume shows the desolvation grid used to estimate the energetics of the interaction owing to rearrangement of water in the formation of the protein–ligand complex. Credit: Arthur Olson, Scripps Research Institute.

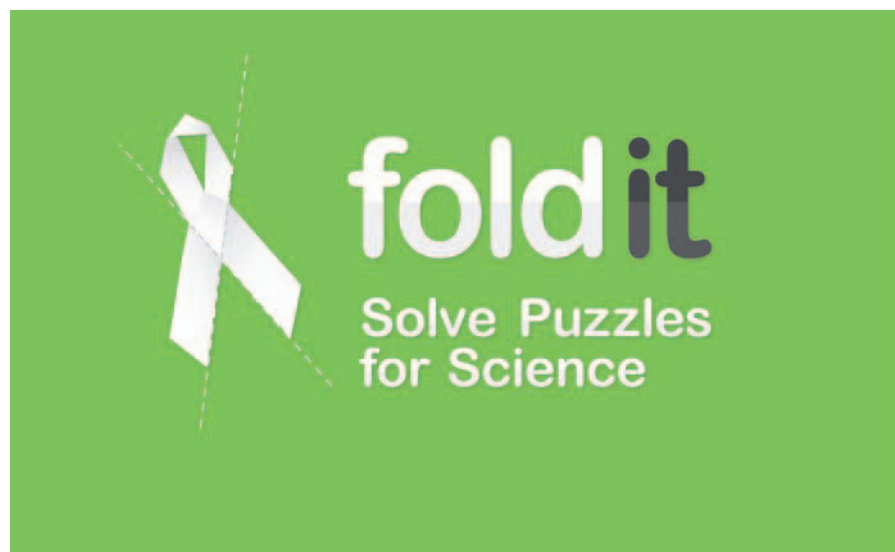


Fig 2 | Foldit’s logo invites people around the world to “solve puzzles for science” by predicting protein folding modes. Credit: University of Washington.

structures is the best one (Fig 3). Players can be soloists, or can form groups from members endowed with complementary skills, all competing to solve the same puzzle challenges. The next step in the project will be to address protein design, adding new functionality to the game to allow users to design new ‘unnatural’ proteins that could

act in important pharmaceutical or industrial applications. Recently, Baker’s group were even able to use the WCG community to crunch protein data with Rosetta’s *de novo* structure prediction method, gaining insight into the molecular function of a host of *Saccharomyces cerevisiae* proteins (Malmström *et al*, 2007).

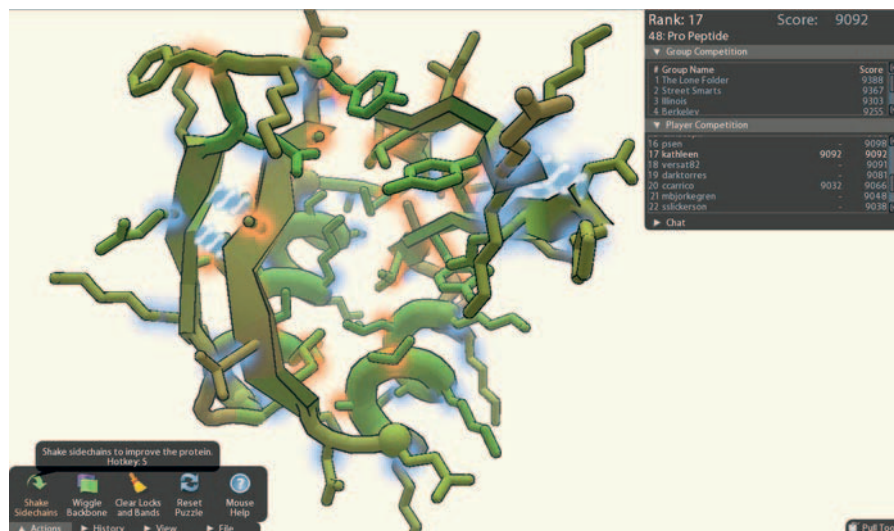


Fig 3 | A screenshot of Foldit’s protein folding software. Players are able to rotate and manipulate the ‘puzzle’ in three dimensions. Their rank, score and the scores of others with whom they are collaborating or competing are shown at the top right of the screen.

Surplus computing power for use in community data-crunching projects is not only available from the central processing units of humble desktop PCs. Gianni De Fabritiis leads the Computational Biochemistry and Biophysics Laboratory at Pompeu Fabra University (Barcelona, Spain), which has built an innovative distributed supercomputing infrastructure composed of graphics processing units (GPUs), which would normally power graphic-intensive computer games. The grid makes use of mainly PlayStation®3 and NVIDIA® graphics cards joined together to deliver high-performance all-atom biomolecular simulations (GPUGRID; www.gpugrid.net). “The main feature of GPUGRID.net is the brute force that it allows in terms of computational power by using accelerator processors, not only as an aggregate, but also as individual volunteered machines,” De Fabritiis commented. “This level of granularity is a fundamental requirement for all-atom molecular simulations as it permits [the] use [of] a wide range of protocols on a very volatile grid” (Giupponi *et al*, 2008). The advantage of using graphics

cards is that “the computational power of graphics hardware is much higher than standard processor[s] due to the large amount of calculations which are required to display realistic three-dimensional images” (www.gpugrid.net).

The distributed infrastructure will allow researchers to compute free energies for protein–ligand and protein–protein interactions, as well as allowing conformational sampling and providing accurate virtual screening for molecules of biomedical interest, De Fabritiis explained. “We are now in a phase where we produce over five microseconds of simulated time per day for systems of the order of 50,000 atoms. [...] The systems being run at the moment include: transmembrane pores, GPCRs [G-protein-coupled receptors] and SH2 domains,” he concluded.

There are many different approaches to harnessing the power of mass collaboration to innovate and solve problems through ‘wikinomics’. As Spradlin noted, “With the convergence of technology—the Internet, social networking, communications—increases in standards of living and education, and a more global awareness than at any time in history, I believe there are now legions of ‘Citizen Innovators’ around the world ready, willing, and able to invest their relevant experience, knowledge, creative talents and hunger for problem solving toward the important

challenges of our time.” His company, InnoCentive, has certainly disclosed one of the most original uses of ‘wiki’ technology. Founded in 2001, InnoCentive taps into the innovation potential of a vast, growing network of some 160,000 InnoCentive Solvers™, to deliver solutions to complex problems in science, technology, business and philanthropy, posted anonymously as ‘challenges’ by private companies, or Seekers™, on the corporate website (www.InnoCentive.com). Individuals can sign up and browse the Open Innovation Marketplace™ for problems that they think they might be able to solve: “Submit the winning solution and earn cash awards from \$5,000 to \$1,000,000,” the website offers. It seems to be a strategy that gets results; in partnership with the Rockefeller Foundation (New York, NY, USA), InnoCentive has hosted a multitude of challenges to help people in impoverished areas of the world, developing solutions related to work on malaria, tuberculosis and energy efficiency, Spradlin remarked. “With a firmly held belief that incentives hold the key to harnessing and focusing the vast collective talent pools available worldwide, we will drive innovation everywhere there is the potential to make a difference. Hence the name: innovation plus incentive, equals InnoCentive,” he said.

Wikinomics has opened the floodgates and turned the stream of innovation into a raging torrent in which anyone with an idea and a computer is free to swim or be merrily transported downstream. Not only are shared efforts and goals pervading the scientific community with the continuous creation and diffusion of new models of data annotation and exchange—such as WikiProteins and WikiPathways (Mons *et al*, 2008; Pico *et al*, 2008)—but also the torrent of information and ideas pours out to the public, allowing worldwide collaboration based around ideals of openness and cooperation. “I will argue that in fact creativity and ingenuity are the most valuable assets lacking in the system today, not access to financial means,” Spradlin commented.

“The implications of this ‘public computing’ paradigm are social as well as scientific,” Anderson wrote in 2003. “It provides a basis for global communities centered around common interests and goals. It creates incentives for the public to learn about current scientific research.” As members of the public

As members of the public can choose to allocate their resources to a specific project, they [...] can democratically influence research policy...

can choose to allocate their resources to a specific project, they have more direct control over the direction of scientific progress and can democratically influence research policy; in this manner, direct engagement could help to reconcile people to science. "We are certainly most thankful for the thousands of people who are willing to donate their unused computer resources to our FightAIDS@home project," Olson said. "[I]t is a unique and effective way to help push the frontiers of biomedical science, and to assist in preventing the spread of HIV/AIDS. The availability of such a large computer resource has broadened our ideas about what is feasible to do in attacking this important problem."

REFERENCES

- Anderson DP (2003) Public computing: reconnecting people to science. Conference on Shared Knowledge and the Web, Madrid, Spain, November 17–19. <http://boinc.berkeley.edu/trac/wiki/BoincPapers>
- Anderson DP, Reid K (2009) Celebrating diversity in volunteer computing. Hawaii International Conference on System Sciences (HICSS-42), Waikoloa, Hawaii, USA, January 5–8. <http://boinc.berkeley.edu/trac/wiki/BoincPapers>
- Chang MW, Lindstrom W, Olson AJ, Belew RK (2007) Analysis of HIV wild-type and mutant structures via *in silico* docking against diverse ligand libraries. *J Chem Inf Model* **47**: 1258–1262
- Giupponi G, Harvey M, De Fabritiis G (2008) The impact of accelerator processors for high-throughput molecular modeling and simulation. *Drug Discov Today* **13**: 1052–1058
- Hickey H (2008) Computer game's high score could earn the Nobel Prize in medicine. Press release, May 8, University of Washington News. <http://uwnews.org/article.asp?articleID=41558>
- IBM (2008) IBM and university researchers to develop research tools to improve cancer patient outcomes. Press release, January 25. <http://www-03.ibm.com/press/us/en/pressrelease/23387.wss>
- Kanter RM (2008) Transforming giants. *Harvard Bus Rev* **86**: 43–52
- Malmström L, Riffle M, Strauss CEM, Chivian D, Davis TN, Bonneau R, Baker D (2007) Superfamily assignments for the yeast proteome through integration of structure prediction with the gene ontology. *PLoS Biol* **5**: e76
- Mons B *et al* (2008) Calling on a million minds for community annotation in WikiProteins. *Genome Biol* **9**: R89
- Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C (2008) WikiPathways: pathway editing for the people. *PLoS Biol* **6**: e184
- Tapscott D, Williams AD (2006) *Wikinomics. How Mass Collaboration Changes Everything*. New York NY, USA: Penguin Group

Andrea Rinaldi

doi:10.1038/embor.2009.79