All for One and One for All

As scientific instruments become ever more powerful, from orbiting observatories to genome-sequencing machines, they are making their fields data-rich but analysis-poor. Ground-based telescopes in digital sky surveys are currently pouring several hundred terabytes ($10^{12}$ bytes) of data per year into dozens of archives, enough to keep astronomers busy for decades. The four satellites of NASA’s Earth Observing System currently beam down 1000 terabytes annually, far more than earth scientists can hope to calibrate and analyze. And looming on the horizon is the Large Hadron Collider, the world’s largest physics experiment, now under construction at CERN, Europe’s particle physics lab near Geneva. Soon after it comes online in 2007, each of the five detectors will be spewing out several petabytes ($10^{15}$ bytes) of data—about a million DVDs’ worth—every year.

These and similar outpourings of information are overwhelming the available computing power. Few researchers have access to the powerful supercomputers that could make inroads into such vast data sets, so they are trying to be more creative. Some are parceling big computing jobs into small work packages and distributing them to underused computers on the Internet. With this strategy, insurmountable tasks may soon become manageable.

One approach to such “distributed computing” was pioneered by computer scientists working with SETI, the Search for Extraterrestrial Intelligence. The phenomenally successful SETI@home program now makes use of the idle computer time of millions of ordinary computer users, working as a screen saver to quietly crunch away at radio-signal data from deep space. As John Bohannon describes on p. 810, the same screensaver technique is now being used by a wide array of researchers studying everything from climate change to gravitational waves and protein folding. Bohannon also delves into the strange tribal world (p. 812) of the “crunchers”: computer enthusiasts whose goal is to become the most prolific processors of data for various screen-saver research projects. And on p. 813, Mark Buchanan samples a piece of computer navel gazing: a distributed computing project to study the geography of the Internet itself.

Another way of distributing both data and computing power, known as grid computing, taps the Internet to put petabyte processing on every researcher’s desktop. On p. 814, Foster highlights the development of a lingua franca of grid computing: a set of standardized interfaces and protocols that permits researchers to work across the Web. Hey and Trefethen (p. 818) describe the U.K.-based e-Science program to design plug-and-play grid technologies for a range of disciplines. And Buetow (p. 822) outlines the ways in which cyberinfrastructure can weld together the vastly different styles of biomedical research.

For all the excitement, however, there are disturbing trends in the directions being taken by funding agencies that have historically been involved with driving the Internet revolution. In their Editorial (p. 757), Lazowska and Patterson consider how downsizing and short-term thinking threaten to derail the next generation of information innovation.

—Daniel Clery and David Voss
All for One and One for All
Daniel Clery and David Voss (May 5, 2005)
Science 308 (5723), 809. [doi: 10.1126/science.308.5723.809]

Editor's Summary

This copy is for your personal, non-commercial use only.

Article Tools
Visit the online version of this article to access the personalization and article tools:
http://science.sciencemag.org/content/308/5723/809

Permissions
Obtain information about reproducing this article:
http://www.sciencemag.org/about/permissions.dtl