

# Data for the masses

Initiatives for digital research infrastructure should focus more on making standardized data openly available, and less on developing new portals.

The flood of digital research data that scientists are generating through genomics, sensor and other technologies has made it imperative to create an infrastructure to use, repurpose and preserve those data. Some such efforts are already under way, notably the US National Science Foundation's \$100-million, five-year DataNet programme, and Europe's Alliance for Permanent Access (see Big Data special, <http://tinyurl.com/5hh2rq>). But how should the responsibilities be divided between governments and the private sector?

A series of events in December highlights the complexities of this issue. One was a pioneering move by Amazon to host large scientific data sets for free, starting with GenBank and other widely used sequence and chemical-structure databases

Amazon's move is not altogether altruistic. Although researchers will be able to download the data to their own computers, the company is betting that many will instead use its 'cloud computing' technology, which makes the company's vast server infrastructure available to process the data sets on a pay-as-you-go basis.

Such services could offer immense benefits to research. By giving scientific data a permanent home online, Amazon could help to ease the long-running problem of databases that are abandoned when, for example, funding dries up at the end of a research grant. Its cloud-computing approach could liberate smaller labs from the cost of running data centres of their own. And it should facilitate the sharing of data and analysis tools between widely dispersed research teams.

Also in December, however, came a reminder of the risks of

depending on the public or private sector alone to create such infrastructure: Google announced it was abandoning its plans to host large scientific data sets for free, apparently because of the economic downturn. In November, the European Union, in collaboration with research organizations, libraries and museums, launched the Europeana online digital library as its much-touted alternative to Google Books. Europeana has scanned valuable historical collections, but its computing infrastructure crashed within hours, not clunking back into service until more than a month later. A similar fate met Géoportail, a service created by the French government as a competitor to Google Earth, and launched with great fanfare in 2006 by President Jacques Chirac.

In creating such public offerings, governments address valid concerns that private companies will exercise monopolies over significant cultural and scientific heritage. But their focus on creating their own digital libraries and databases too often means that other, perhaps more important, ways to address such concerns are neglected. Making standardized data openly available to both commercial and not-for-profit organizations, for instance, could spur innovation of superior information services. And to avoid embarrassing crashes, public efforts might well consider partnerships with private firms to tap into the economies of scale and expertise of the Googles and Amazons of the world. ■

**"Making standardized data openly available could spur innovation of superior information services."**

## Culture clash

Samuel Huntington will be remembered for his ideas, controversial as they might have been.

It is my hypothesis that the fundamental source of conflict in this new world will not be primarily ideological or primarily economic. The great divisions among humankind and the dominating source of conflict will be cultural ... The fault lines of civilizations will be the battle lines of the future."

With newspaper headlines dominated by the conflict between Israel and the Palestinians, as well as by the confrontation between Pakistan and India, these early lines in Samuel Huntington's 1993 *Foreign Affairs* essay, 'The Clash of Civilizations', seem more resonant than ever. Huntington, who died on 24 December 2008, aged 81, was a controversial giant among public intellectuals in the United States. He taught at Harvard University for 58 years, founded *Foreign Policy* magazine and worked in the White House under Presi-

dent Jimmy Carter. Yet he attracted strong criticism from scientists.

In the 1980s, for example, members of the US National Academy of Sciences voted twice to deny him membership. At the heart of the dispute was Huntington's use of mathematical notation as a shorthand summary of complex political ideas, such as his use of equations to claim that apartheid South Africa was a "satisfied society". As mathematician and academy member Serge Lang of Yale University in New Haven, Connecticut, declared, this gave "the illusion of science without any of its substance".

That quarrel paled, however, next to the controversy over Huntington's essay. Speculating on conflicts to come in the post-cold-war era, he argued that people ought to be classified as being members of distinct cultures, and asserted that wars can be better understood as conflicts among these 'civilizations', rather than between nations. Most controversial, however, was his suggestion that Chinese and Islamic 'civilizations' posed the biggest potential risks to Western nations.

For many people, including senior policymakers in Europe and the United States, Huntington was saying aloud what they had already