

Comment

Who owns the data?

Gregory A Petsko

Address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.
E-mail: petsko@brandeis.edu

Published: 31 March 2005

Genome Biology 2005, **6**:107 (doi:10.1186/gb-2005-6-4-107)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2005/6/4/107>

© 2005 BioMed Central Ltd

Besides an astronomical amount of sequence data and a lot of useful technology, perhaps the most significant legacy of the genomics revolution has been an insatiable appetite for data. This hunger was part of the reason that the privately funded human genome project at Celera Corporation released its sequence information sooner than intellectual property considerations would have made desirable (competition from the publicly funded human genome sequence project was the other part). The same hunger motivated the US National Institutes of Health (NIH), the National Science Foundation, and the Howard Hughes Medical Institute to require that structural biologists funded by those agencies deposit their atomic coordinates into a public database in a timely manner. But this flood of information hasn't curbed the appetite at all. Like Cleopatra in Enobarus's marvelous description from Shakespeare's *Antony and Cleopatra*, it seems genomics makes hungry where most she satisfies.

Of course, this desire wars with another fundamental human appetite: that for money. Much of modern life science is driven by the longing to make a profit. It fuels the biotechnology and pharmaceutical industries. It underlies the choice of research problems in many academic laboratories. And at its heart is the concept of property, of ownership, both of ideas and of data. This concept would seem to be perpetually opposed to that of free, publicly available sequences, structures and technologies.

Historically, the battlefield on which this conflict was fought was the courtroom, where scientists and corporations would engage in Talmudic-style disputes over dates in notebooks, interpretations of patents, and other claims to priority. In the immediate post-World War II era these arguments tended to be over technology developed by physicists, chemists and engineers. Biologists didn't join the fray until after 1980: in part there was no biotechnology industry until about then, but it was largely because most academic biology was publicly funded, in the US by the NIH. That would seem to make the results of such research public property.

The Bayh-Dole Act, passed by the US Congress in 1980 and named for its co-sponsors Senators Birch Bayh and Robert Dole, changed all that. The Act provided recipients of federal research and development funds with the right to retain ownership of their patents. It did even more: it charged them with the responsibility of ensuring commercial use of inventions created with federal financial support. While it is technically possible for a university to have different policies regarding the patenting and licensing of inventions which were not developed as a result of federally funded research, in general the universities' interest in maintaining the flexibility to draw research funds from multiple sources, including the federal government, and the desire to avoid applying conflicting policies, have led to most of them having a single policy that is consistent with the Act. The underlying tenet of the Bayh-Dole Act is that federally funded inventions should be licensed for commercial development in the public interest. That principle is now reflected in virtually all university policies in the US, whether or not the invention is federally funded.

Since the Bayh-Dole Act permits universities, other non-profit organizations such as teaching hospitals, and, in most cases, commercial federal contractors to retain title to inventions that are conceived or first reduced to practice in the performance of a federal grant, contract, or cooperative agreement (in exchange for certain obligations on the part of the contractor), it immediately created a huge economic incentive for academic biologists to start their own companies or to become involved with existing ones. Bayh-Dole was directly responsible for the explosive growth of the biotechnology industry in the 1980s. It also created the culture of intellectual property that underlies that industry. For over twenty years, the answer to the question "Who owns the data?", according to the Bayh-Dole Act, has been "the scientist who collected it and the organization for which he or she was working at the time". Since raw facts could not be property (you may patent a mousetrap, but not data on mice; you may copyright an article, but not the data on

which it is based - although the patenting of gene sequences is a blow to this tradition), this answer led to a culture in which data were hoarded, often to be published only after the application itself was developed.

This answer is now being challenged by a new one, driven by the cultural change genomics is creating in the life sciences - a culture of public databases and open access. The first area of modern biology to reel under the challenge has been the scientific journal publishing industry. Some journals, such as *Science*, are published by not-for-profit scientific societies (which derive a hefty chunk of their operating expenses from the subscriptions); more, like *Nature*, are revenue-generators of for-profit publishing houses. About ten years ago, a group of scientists headed by Nobel Laureate Harold Varmus, then Director of the NIH, began to argue that it was unfair to ask other scientists, who are after all members of the public, to pay to read the results of research that had been publicly funded. They quickly found allies in patients' advocacy groups, who believe advances in medicine would come about more quickly if everyone had equal access to discoveries. Despite considerable skepticism by many scientists - and much gnashing of teeth from publishers - about five years ago the first 'Open Access' journals began appearing. Their business model is that authors of papers appearing therein must pay a fee for the privilege (peer review is still required for acceptance), but in return, all rights to the material in the paper remain with the author and anyone can access the full text and any supplemental information free of charge forever. Scientists in developing countries, in particular, benefit greatly from such a policy, since many journal subscriptions, online or in print form, are beyond their means.

And on 3 February, NIH announced that as of 1 May this year it expects that all research papers resulting from research it funds will be deposited into an open-access electronic archive that will be maintained by the US National Library of Medicine (which currently runs the PubMed journal database and PubMed Central full-text archive, within a year of their appearing in any journal. Current estimates are that over one third of all highly cited papers in the life sciences report the results of NIH-sponsored research, so the policy is likely to have a big impact almost immediately, even though there is no active enforcement. If the existing open-access journals like *PLoS Biology*, *Journal of Biology*, and this journal (which makes all refereed research articles freely available online but charges a subscription price for access to other content, such as my Comment columns - which are worth every penny) are able to stay in business by, for example, charging authors rather than subscribers, and if they start to attract top-flight papers, the closed-access journals will come under severe financial pressure to adopt a similar business model. In any case, given the new NIH policy, it would seem that for much of their content, closed-access journals will only have a year - and maybe eventually a lot less than that - to make their profits. The Wellcome

Trust in the UK is also a big supporter of Open Access, and is considering establishing a joint archive of papers with the US National Library of Medicine. Where Wellcome goes, the UK Medical Research Council is likely to follow. Add in Germany, France and Japan and most of the literature will be covered.

Even more intriguing is the advent of open-access technology. Here there is a model from outside biology: so-called 'open-source' software. Programs developed under the open-source concept have their source code freely available to users, with the restriction that any improvements made by anyone must be offered to the user community free of charge. A variation of this model levies a cost to commercial users while allowing academics and other non-profit groups to obtain the code free of charge. The first example, the Linux operating system (named after its inventor, Linus Torvalds, who is popularly credited with the open-source model), has proven so successful that it is making Bill Gates and Microsoft nervous about the future of their closed-source, very much for-profit Windows operating system. Open-source software has begun to have a big impact in structural biology, where programs like Coot, PyMol, Phenix and so on are making high-quality crystallographic computing available to all.

And now this idea is being applied to biotechnology. Early in 2005 an exploratory project called Science Commons was launched. The mission of Science Commons - an offshoot of Creative Commons, which provides less restrictive copyright licenses to authors - is to develop open licenses for technologies. As a model, it could do worse than look to a remarkable new concept developed by CAMBIA, a non-profit biotech research group affiliated with Charles Stuart University in Canberra, Australia. In a paper published, ironically, in the closed-access journal *Nature* on 10 February (Broothaerts *et al.*, Gene transfer to plants by diverse species of bacteria, *Nature* 2005 **433**:629-633), researchers at CAMBIA report a breakthrough in biotechnology by successfully transferring foreign genes to plants using several bacteria other than the usual *Agrobacterium tumefaciens* (At). They introduced a specially modified Ti plasmid into *Rhizobium*, *Sinorhizobium* and *Mesorhizobium* - all organisms closely related to At - and showed that the transformed strains could be used to express foreign genes from the plasmid in tobacco, rice and *Arabidopsis*. Integration of the inserted segment into the plant genomes was also confirmed. The work is exciting because many plants, especially crop plants, are resistant to gene transfer by At. But it's also noteworthy because of what CAMBIA is doing with it.

CAMBIA has applied for a patent on the technology, which they call TransBacter™. But they are offering this technology as an 'open-source' alternative to At technology, which is controlled by Monsanto, the large agricultural firm that holds the relevant patents. CAMBIA calls its license concept

BIOS - Biological Innovation for Open Society. The way it works is simple. Others may commercialize products based on the procedure. But any improvements in the technology must be shared freely, to the benefit of all users. The intent is that researchers in poor countries especially, where agricultural research is very important, will thus have open access to a method that may help their efforts. There's a website, Bioforge [<https://www.bioforge.net/>], to help biotech researchers collaborate on this and other developments (among them new reporter/marker genes and microarray-style genotyping technologies). There are several levels of projects, some open only to BIOS licensees, some open to all and some open at intermediate levels. Joining a project enables the participants to see, use, and deposit information that will not necessarily be available in the public domain. It will allow them to share their improvements with other members of the protected commons community of BioForge. In order to join a project, organizations and individuals must agree to the community norms about confidential sharing of improvements and biosafety data, and must provide information on their institutional affiliation and policies that may apply to sharing of data. Access to certain projects may require a legal commitment to the sharing of improvements in return for being able to obtain the benefit of the technology and improvements.

For humanitarian efforts and work on crops that are of limited interest in developed countries, CAMBIA's model promises to be truly revolutionary. It doesn't do away with the incentive to invent, or to develop, but it makes the information needed to do such things available to everyone. If there is an untapped reservoir of creativity in the Third World, an idea such as this might unleash it. It will be interesting to see whether the concept catches on, as open-source software clearly has. No one wants to see the financial incentives that have fueled the biotechnology explosion removed. But companies can clearly live within the open-source model - IBM does, for example (open-source software even contributes to its revenues, since among other things IBM makes much of its money by selling services to people who have open-source software and need help). CAMBIA, by the way, was funded by the Rockefeller Foundation, Horticulture Australia, and Rural Industries R&D Corporation, so in a sense its work represents a triumph of the Bayh-Dole concept. It remains to be seen whether the pharmaceutical industry, which in my opinion would benefit greatly from increased sharing of ideas and information, could find an open-source model it could live with. But if scientific publishing and software development are any indication, this is not an idea that's going to go away any time soon.

Who owns the data? Increasingly, at least for some things, the answer is starting to be nobody. Or everybody.