Changing the rules?

The agreement between Celera and *Science* magazine concerning Celera's publication of its human genome sequence is upsetting many researchers in bioinformatics

Science writers have for some years been moving away from their historically awestruck 'ain't-science-grand' reporting of research developments. Instead, they are more often inspecting science with a critical eye, particularly policy matters. While they are at it, they are filling an information gap, performing a job that the scientific press cannot—or will not—do. The result, however, is that disputes which might in the past have been kept in the scientific family are now broadcast

The larger world has begun to learn that the rules of science are being revised in ways that some find unsettling, and that science is at something of a loss about what to do

much more widely. The larger world has begun to learn that the rules of science are being revised in ways that some find unsettling, and that science is at something of a loss about what to do.

The latest episode in the discord between science and commerce spilled into public view in the *Los Angeles Times* in early December, 2000. The case in point is the decision by the journal *Science* to allow special—unprecedented, some charge—rules governing the data underlying a paper that it wanted very much: Celera Genomics's account of its once-derided but highly successful shotgun sequencing of the human genome. The agreement between the company and the journal will let Celera keep the sequence data on the company's own computers, rather than deposit it in one of the public databanks as is usual. In practice, this has been GenBank at the National Institutes of Health, the European Bioinformatics Institute in Hinxton,

UK and the DNA Data Bank of Japan, designated as official and cost-free repositories at a 1995 meeting where the NIH and the Wellcome Trust drafted principles for the Human Genome Project. The three databanks have melded so seamlessly that, to the user, they appear as one.

In return for keeping the data on their own computers, *Science* says, Celera will permit free access to its human genome sequences for academic researchers. 'Academic users may access it, do searches, download segments up to one megabase, publish their results, and seek intellectual property protection,' *Science* said in a statement dated December 6, 2000. The statement went on, 'Longer downloads, up to and including the whole genome, are allowed but require a more formal agreement, signed by an institutional representative, not to redistribute the data.' Commercial users are to be permitted free access too, on condition that they sign a Material Transfer Agreement vowing not to commercialize their results or redistribute the sequence. Or they can sign up for a fee which goes as high as \$15 million for giant pharmaceutical companies.

This arrangement between Celera and *Science* was greeted with horror, chiefly at the notion that the sequences will not be nestled in the public databanks alongside the sequences generated by the publicly-funded international consortium

This arrangement was greeted with horror in Europe and the USA, chiefly at the notion that the sequences will not be nestled in the public databanks alongside the sequences generated by the publicly-funded international consortium. 'They have started the 'balkanisation' of genomic data,' Michael Ashburner, joint head of the EBI, said. '[The sequence data] is absolutely useless for serious bio-

informatics if it's not in one location.' *Science* has responded tartly that its rules for authors have never specified GenBank. Donald Kennedy, *Science's* editor, acknowledged in an interview that the results would be better for academic researchers if the sequences resided there. 'But Celera couldn't do that because US copyright law doesn't protect these databases,' he said.

The disconcerting intricacies of this internecine fracas are many. One is the fact that the public consortium withdrew its papers from *Science* and submitted them to *Nature*, *Science's* great rival for Supreme Ruler of journal publishing. The public consortium and the private company still managed to keep last June's joint vow to publish simultaneously, however. Both journals released their papers online, and free on February 11. Formal publication came later that week, in *Nature's* issue dated February 15 and *Science's* dated February 16. The timing coincided brilliantly with the big annual meeting of *Science's* publisher, the American Association for the Advancement of Science (AAAS), which began the same day. AAAS flaunted its coup, distributing

thousands of free copies of the issue in tote bags emblazoned 'Celera.'

According to Ashburner, however, *Science* might still end up with the short end of the stick because of the deal it cut with the company. 'People will stop publishing genomics papers in *Science*,' he predicted.

Another grotesquerie is that Celera made use of some of the public data in assembling its own sequences. Indeed,

argues Eric Lander, who heads the public project's lead sequencing centre at the Whitehead Institute for Biomedical Research in Cambridge, MA, Celera could not have strung its 'shotgunned' jumble of sequences together in proper order without guidance from the public genome map. The irony is that NIH researchers are now lobbying the agency to buy them \$15 000 individual subscriptions to Celera's Discovery System, the company's much-admired set of tools for analysing genome data. Several academic subscribers—the Max Planck Gesellschaft in Germany, the Karolinska Institute in Stockholm, the University of Tokyo, and a number of US universities—have already signed up.

Still another oddity is that, in a move that could be captioned, 'I trust you, honey, but I cut the cards,' *Science* says that to guarantee access, it will hold a copy of Celera's genome data 'in escrow.'

That may be an admirably prudent declaration, but it is not clear which planet the escrow is on. Celera's installation in a suburb of the NIH is famous for its stupendous banks of computers that store and crunch its version of the human genome, said to be some 80 terabytes in size.

Much of the argument has focused on specific features of the announced publishing arrangement such as the deposition site. There is a longer-term question, however. Was this a one-time deal, a unique situation, never to be repeated? Is it simply that *Science* was willing to bend the rules just this once for the sake of getting one of the juiciest papers that will ever grace the scientific literature? Or has *Science* forever changed the rules of play for all of science publishing?

Predictably enough, Kennedy declared, '[W]e don't think this ushers in a new era in scientific publication.' He argued that the journal is abiding by its own rules for free access to data. Moreover, he said, *Science* is not leading this parade down the slippery slope toward restricted data access; the new era was ushered in some time ago: '[A]uthors are restricting access to commercial readers, or holding back proprietary data sets, in a number of cases now.' According to Kennedy, recent papers in *Nature*, *Nature Genetics* and the *Proceedings of the National Academy of Sciences* 'have directed readers to websites that offer algorithms for protein-protein interactions (necessary to verify

the conclusions) or microarray data—and say that academic users can access free but commercial users must pay.'

The most concerted opposition has come from bioinformaticists, especially those at the European Bioinformatics

Is it simply that *Science* was willing to bend the rules just this once for the sake of getting one of the juiciest papers that will ever grace the scientific literature?

Institute. EBI's Ewan Birney and others circulated a statement asking their colleagues to protest the deal to *Science*. Sean Eddy, of Washington University, St Louis, who co-authored the statement, reported that it brought a ton of complaints down on the journal. 'Possibly too many. It may have become an irritating bombardment more than the constructive attempt to make a point that Ewan and I were aiming for. It's too early to tell whether it's had an effect. I'm told that *Science* is now attempting to clarify the part that I found most worrying—the failure to distinguish clearly between "redistribution" of Celera data (which is prohibited) and "publication" of results derived from Celera data (which is allowed), since many bioinformatics results involve partial or complete disclosure of the primary DNA sequence data.' Kennedy declared 'I think the final agreements will cover the problem fairly well.'

Even if the final agreement solves the specific bioinformatics problem, Eddy remains an eloquent spokesman for the Platonic idea of free and open dissemination of scientific data. 'I guess I'm disappointed in *Science* and the AAAS. I feel that *Science*/AAAS have, in their enthusiasm to publish the human genome, fallen into the logical fallacy of the false mean—the idea that the solution to two conflicting views should always be a compromise, when in fact sometimes the world really is black and white, and one view is right and the other is wrong,' he said. '[B]ut no compromise needs to be made, and what they're compromising is a core ethical principle of science.' Compromise is unnecessary, he argued, because scientists already have access to Celera's data through paid licenses, and have already published results based on those data. '*Science*'s argument that their deal is the

only possible way of getting us access to the data is disingenuous.'

Nature editor Philip Campbell declined public comment. But an editorial in the genome issue argued that the burden of proof that the journal's policy on genome data access should change lies with the companies. '*Nature* believes that the human genome sequence is not the place for the traditional rules to be broken,' it observed. Andrew Marshall, editor of *Nature Biotechnology*, one of *Nature*'s sister journals, is one of the worriers. While acknowledging that commercial pressures are encroaching on open access to data everywhere, he argued in a January editorial titled 'Slippery slopes?' that in *Science*'s arrangement with Celera, 'more ground appears to have been conceded.' He thinks a precedent has now been set, and that other companies who want the rewards that come with publication in a very classy journal may well insist that the data remain on their own web sites.

Not everyone is in high dudgeon. Many scientists and journal editors alike praise Celera for a degree of open-handedness with its data remarkable for an organization that must keep its stockholders happy. They agree with Kennedy that *Science*'s solution was the best outcome that could be achieved; the important point is that the data will be available free. 'Of course everyone would prefer no constraints at all, but even with the restrictions, it still represents in principle a level of information unprecedented in the annals of science publishing, so I can't fault *Science* for trying to reach a suitable compromise. Would it be better if a private company did not release the fruits of its research? Or just made it available to those willing to pay?' said Kevin Davies, editor-in-chief of *Cell* Press. Davies concedes that he would not be surprised if other commercial groups tried submitting papers with similar restrictions. 'But I highly doubt that any other publishing company will tolerate that imposition.'

Which leaves the ball in the publishers' court.

Tabitha M. Powledge

The author is a freelance science writer and editor near Washington, DC.

Holger Breithaupt contributed to this article.

DOI: 10.1093/embo-reports/kve055