

nature genetics

volume 23 no. 1

september 1999

The database and the law

Information is both a commodity and a source of knowledge. Its dual function is central to the question of how best to regulate its dissemination, how to ensure that both innovation and free competition are encouraged. Current legislation that protects intellectual property does so through patent and copyright law—and, in Europe, through the European Directive on the Legal Protection of Databases.

With a couple of bills that prescribe the way in which content of databases may be used under consideration by the Congress of the United States (not to mention the much-discussed e-BioSci server proposed by the director of the National Institutes of Health), the control of factual information is the focus of much debate. The translation of either bill into law will affect the way in which the information contained in databases is used by geneticists (and others) for research and education. As such, it

comes as no surprise that publishers and organizations such as the National Academy of Sciences and the American Medical Association have taken positions.

With printed material, the 'first sale' doctrine of copyright law allows a copy of a work to be passed from hand to hand without permission of the copyright owner. (A hard copy of this issue of *Nature Genetics*, for example, is likely to pass through the hands of at least six people.) The dissemination of digital information, on the other hand, is more problematic, because it often involves copying the information.

Neither copyright law nor patent law protects the factual information of databases. Copyright law protects works of original expression—it protects the way in which information is expressed, rather than the facts themselves. In fact, it decrees that information should remain free of legal constraints. This allows, for example, an investigator to develop and publish an algorithm that makes use of previously published data, without fear of legal action. Patent law protects ideas that meet four criteria, one of which is nonobviousness.

A collection of factual information, such as a telephone directory or a database of all *Saccharomyces cerevisiae* gene sequences is neither original nor unobvious. Unless there is something that is arguably unique in the way in which the information is selected or presented, the collection cannot be protected by copyright law. And even if it can, copyright does not extend to the facts, data and other forms of unbundled information that a database contains.

The database protection bills H.R. 1858 (The Consumer and Investors Access to Information Act) and H.R. 354 (The Collections of Information Antipiracy Act)

"[d]ata are the building blocks of knowledge and the seeds of discovery. They challenge us to develop new concepts, theories and models to make sense of the patterns we see in them . . . They also are the foundation of sensible public policy in our democracy."

—Joshua Lederberg

(Testimony during Hearings on H.R. 345 before the House Subcommittee on Courts and Intellectual Property of the House Committee of the Judiciary, 106th Congress, 1999).

represent efforts to enact new, hybrid intellectual property rights that protect the content of databases, in that they impose restrictions on the way in which content can be used. They are closely related to The European Directive which was passed in 1996 and is now enacted in approximately half of the member states of the European Union. The Directive protects against extraction or reutilization of the whole or any substantial part, evaluated quantitatively or qualitatively, of a database that is the product of substantial investment—colloquially, this is known as ‘sweat of the brow’ investment (as opposed to the sort that is guided by original, creative insight). How one assesses whether a particular set of gene sequences is a substantial part of a database, qualitatively speaking, may be hard to define. The latest version of H.R. 354 tips its hat to the ‘fair-use’ provisions of copyright law in that it exempts “additional, reasonable uses” by educational, scientific and research organizations, but it limits this to “an individual act of use or extraction of information done for” specified purposes. This may place the burden of defense on the researcher.

Coupled with the Uniform Computer Information Transactions Act (UCITA; formerly known as Article 2B of the Uniform Commercial Code; ref. 1), H.R. 354 threatens to tilt the balance that is maintained by copyright law in the direction of protectionism and away from free competition. UCITA clarifies the enforceability of licensing terms for electronic databases and thus may embolden database producers to use very restrictive terms in their license agreements, limiting the downstream use of the information in a database unless additional fees are paid. It also enforces license terms that are not accessible to the purchaser prior to his/her payment of the licensing fee, by enforcing shrink-wrap and click-wrap licenses. (These are standard-form electronic licenses.) UCITA is currently supported by major software producers, such as Microsoft Inc. and International Business Machines Inc., and opposed by many commercial users of computer information and academic libraries.

Not surprisingly, the different forms of legislation have drawn criticism. UCITA was dropped by the American Law Institute (which participated in its drafting) once it became clear that the bill was unlikely to be embraced by all states. It should be noted though, that because UCITA allows a licensor to specify which law applies to the agreement, its adoption by a single state may make it applicable to computer information transactions nationwide. Paul Uhlir (of the National Research Council) and Jerome Reichman (of the Vanderbilt University School of Law) note that H.R. 354, if implemented, may hinder the creation of complex databases—for scientific and educational purposes—that incorporate data obtained from previously-existing databases². (H.R. 1858 is less limiting in this respect.) Especially alarming is a prediction by Reichman, who feels that enactment of the proposed legislation will lead to restricted access and that obtaining information in the future may become akin to “navigating the waterways of the Middle Ages, with a tax levied every which way you go”. He also thinks that the biogenetics community may be one of the hardest hit by a combination of a database law and UCITA.

Clearly, present legal manoeuvres aim to influence the way in which information in databases is transacted: who will benefit from it; how they will do so; and at what cost. The implications of the database protection bills should not go unnoted by the scientific community, nor should those of UCITA. Whereas H.R. 354 seems unlikely to be passed during the current sitting of Congress, it has powerful backers, with the likes of Elsevier Reed Publishing Inc., Bloomberg, the American Medical Association and The McGraw Hill Companies, Inc. lobbying in its favour. Celera Genomics Inc. takes no formal position on the legislation, although, according to its Director of Policy Planning, Paul Gilman, it is “concerned” at the prospect of database duplication. As well it might. Legislation that protects investment is appropriate. It is also appropriate, however, that legislation benefits society as a whole; in this respect, H.R. 354 and UCITA do not seem to fit the bill.



1. Reichman, J.H. & Franklin, J.A. *Univ. Penn. L. Rev.* **147**, 875–970 (1999).
2. Reichman, J.H. & Uhlir, J.A. *Univ. Berkeley Tech. L. Rev.* **14**, 793–838 (1999).