Strengthening the United States’ Database Protection Laws: Balancing Public Access and Private Control

David B. Resnik
The Brody School of Medicine, East Carolina University, Greenville, NC, USA

Keywords: databases, copyrights, intellectual property, sui generis protection, European Union, Feist v. Rural Telephone

ABSTRACT: This paper develops three arguments for increasing the strength of database protection under U.S. law. First, stronger protections would encourage private investment in database development, and private databases have many potential benefits for science and industry. Second, stronger protections would discourage extensive use of private licenses to protect databases and would allow for greater public control over database laws and policies. Third, stronger database protections in the U.S. would harmonize U.S. and E.U. laws and would thus enhance international trade, commerce, and research. The U.S. should therefore follow the European example and develop two tiers of protection for databases: 1) protection for creative databases under copyright law; 2) protection for non-creative databases through a special type of sui generis protection. In order to balance private control of data and public access to data, sui generis protections should define a “fair use” exemption that permits some unauthorized extraction of data for private, educational, and research purposes, provided that such extraction does not adversely impact the economic value of the database.

1. Introduction: The Impact of Electronic Databases on Research

The Internet has created new forms of communication and has greatly enhanced humanity’s ability to access and transmit information. A single person sitting down at his or her personal computer can log on to millions of web sites around the globe and...
download gigabytes of data. The Internet has also had a significant impact on the pace of scientific research and the strategies for conducting research. In the past, a researcher would need to go to the library to browse through old journals to find a particular article, a process that could take days, or even weeks, if the library did not own the journal of interest. Today, he or she can find and download an article in less than a minute using search engines and electronic databases. In the past, the period between submission of an article for publication and its appearance in print could take many months or several years. Today, submission, peer review, and publication can all take place electronically, which decreases significantly the time between submission and publication.1

Perhaps the most important effect that the Internet has had on scientific research is that it now gives researchers access to vast quantities of data that they can search and analyze. Indeed, the precursor to the Internet, the ARPANET, was originally developed as a means of sharing data related to national defense research.2 For example, biologists have available enormous quantities of DNA sequence data for many different genomes, including the human, mouse, rice, corn, fruit fly, and Escherichia coli genomes. They also have access to data on many different biologically significant structures and compounds, including RNA, proteins, hormones, receptors, membranes, cell types, tissues, and x-ray images.3,4 Using the analytical tools developed by the emerging field of bioinformatics, biologists can now employ powerful computers to search (or “mine”) these databases to discover correlations and patterns in the data and to compare different databases.a

This bountiful harvest of information comes at a price, however. It costs a great deal of money to compile, maintain, service, and upgrade electronic databases. Unless society decides to increase its support for public databases, private businesses will need to foot the bill for many of these essential information services. In biomedicine, several companies have developed private databases, including Celera Genomics, Monsanto, Monsanto,

---

a. For example, a scientist can discover a new type of cancer cell by comparing the cell to an existing cell database. If the cell type is unique, he or she can sequence the DNA in the cell and compare that DNA to DNA available in other databases to determine whether that cancer is associated with a particular gene or mutation. If the scientist discovers a malfunctioning gene associated with the cancer, he or she can use the sequence data from that gene to determine the primary structure of the protein expressed by the gene, and he or she can use a protein database to develop a model of that protein’s higher-order structures as well as its functions in cell. Throughout all of his or her steps, the scientist will also benefit from searching publication databases for data, methods, or tools relevant to his or her research, without access to these enormous databases with powerful searching tools, the scientist would have been severely hampered in his or her ability to discover the biochemical basis of this cancerous cell. Thus, electronic databases can greatly increase the pace of research and efficiency of research. Data mining can be an important method in scientific discovery when one has large amounts of data but has not developed a hypothesis. One can search the data to discover patterns, which may suggest a hypothesis. One may then confirm the hypothesis by obtaining additional data. Data mining, which can be characterized as “data-driven” research, differs from the traditional approach to scientific method, which can be characterized as “hypothesis-driven” research. A number of authors have argued that there are epistemological and methodological problems with data mining, but this paper will not explore these issues. For further discussion, see Glymour, C. et al (1997). Statistical themes and lessons for data mining. Data Mining and Discovery 1:1-10.

---