



Elizabeth Howard and Gabriel Ramsey

With some planning, you can use intellectual property laws to adequately protect the wealth of data stored in your databases. Those valuable data may not be copyrightable, but the software used to store them is.

Bioinformatics Databases Questions of Copyright

When people in the biopharmaceutical industry think of intellectual property, we most often think of discoveries, formulations, and processes — the kinds of assets that tend to be guarded as trade secrets or protected by patents. But in recent years, thanks to the emergence of bioinformatics, an increasing number of companies find themselves in possession of extensive collections of sequence information and other data, organized in database formats. These data are commercially valuable, and companies are struggling to find the best ways to use intellectual property laws to protect and leverage them.

The most obvious way to protect database contents is to limit disclosure through contractual restrictions and maintain the information as trade secrets. Licensing agreements prohibiting unauthorized copying of a database have been upheld. However, publicly accessible databases can result from inadvertent or pirated disclosure. There may even be practical, ideological, and competitive reasons for voluntary, public disclosure. The question of intellectual property protection for publicly disclosed information therefore arises. This discussion considers the applicability of copyright protection to published, informational contents of bioinformatics databases and ultimately proposes a mechanism to achieve such protection.

The Easy Questions

Database content, ideas, and facts. The Copyright Act is designed to protect “expression” created by an “author,” but expressly denies protection to any “idea, procedure, process, system, method of operation, concept, principle, or discovery” (1). Accordingly, copyright does not protect scientific discoveries or discovered “facts.” So much information in emerging bioinformatics databases, such as GenBank, Data Bank of Japan (DDBJ), and European Molecular Biology Laboratory (EMBL) databases, for

example, particularly pure research results and sequence data, are not protectable.

Copyright can be used to protect journal articles or other prose or graphic expressions of experimental results, but the pure results contained in such articles, again, cannot be protected. “The author of a scientific article published in a professional journal is certainly not entitled to a monopoly of the ideas presented therein” (2). Indeed, a particular expression of findings is protectable only to the extent that the words or images used are not the only way to express the results. This can be problematic in scientific contexts because many theories and facts “lend themselves to a very limited manner of expression” (3). So, graphical representations of genomic data incorporated into databases (such as graphics representing a particular genome) may be more amenable to copyright protection than text. Arguably the decision to represent facts in a particular visual form is an exercise in creativity.

The inability to copyright facts is a serious limitation, and it has inspired some ingenious potential solutions. One recent commentator, for example, has suggested that DNA sequences be converted to music files such as MP3, so that copyright protection might be achieved through the musical expression of genetic content (4). Perhaps not as frivolous as it seems, this argument demonstrates the difficulty of protecting natural facts. Moreover, it remains an open issue whether nonnatural sequences might constitute works of authorship worthy of copyright protection.

The Big Question

Database protection. Although copyright protection for individual pieces of data within bioinformatics databases is limited, “compilations” of such data can constitute copyrightable subject matter (5). Such compilations may be protectable when an author exercises the requisite creativity in choosing which facts to include and in what order to place or arrange them.

Gabriel Ramsey is an associate and **Elizabeth Howard** is an intellectual property partner at Orrick, Herrington & Sutcliffe, LLP, Silicon Valley Office, 1000 Marsh Road, Menlo Park, CA 94025, 650.614.7316, fax 650.614.7400, ehoward@orrick.com, www.orrick.com

In the bioinformatics context, the collection of data in discrete database records returned by search queries are potentially protectable compilations. Database structures can be much more dynamic, with constantly changing records. Also complicating this discussion, and beyond the scope of this article, is the question of whether the physical or logical organization of the computer database itself is copyrightable.

There are two problems, however, with copyright protection for bioinformatics database records. First, courts have not wholly grappled with compilation "creativity," and thus reliance on this theory may be burdened by considerable risk. Indeed, "creativity" in this context usually amounts to a matter of characterization. If a compilation is derived from a process characterized as "thoughtful" selection, protection can be afforded. However, selection or arrangement characterized as "obvious," "typical" or "routine" would be unprotected. The Supreme Court has observed that although originality "does not require that facts be presented in an innovative or surprising way," it is equally true that "the selection and arrangement of facts cannot be so mechanical or routine as to require no creativity whatsoever" (6).

A rough test for assessing creativity is to consider the extent that external factors dictate selection or organization: "The creative spark is missing" where "external factors so dictate selection that any person composing a compilation . . . would necessarily select the same categories of information" (7). If industry standards or conventions require a particular categorization of data, those categories will not be protectable. This can be particularly problematic in scientific fields in which a premium is placed on consistency and "creative" selection and organization of data can lack utility. Further, in tension with database copyright protection is the denial of protection for any "procedure," "process," "system" or "method of operation." Database structure or organization can derive from purely functional needs and can be viewed in functional terms, rather than as the result of "creative authorship."

The Supreme Court held that telephone "white pages" employing an alphabetical organization lack sufficient creativity

because such organization ". . . is not only unoriginal, it is practically inevitable" (8). Similarly, organization of database sequence information by functional categories or keywords such as "gene name," "protein name," "author names," "organism names," or other widely used and obviously functional categories may lack the requisite creativity for copyright. (See, for example, www.ncbi.nlm.nih.gov/blast/ — standard nucleotide-nucleotide search results incorporating largely functional fields and raw sequence data.)

A category is more likely to be sufficiently creative if it is only one of numerous possible options (9). In a case involving dental procedures, taxonomy of procedures were creative where they "could be classified . . . in any of a dozen different ways." Similarly, if the selection of data populating a given record is more easily characterized as a matter of "discretion," "personal taste," or "judgment," protection is more likely (10).

For example, if data is carefully selected to construct a database of narrow relevance (such as to answer a particular question) requisite creativity can exist. However, courts would likely be inclined to deny protection if the database owner simply incorporated all relevant data. Further, the "decisions" to incorporate certain information must be motivated by discretion or judgment, rather than scientific necessity. Indeed, it has been observed that "creativity inheres in making nonobvious choices from among more than a few options" (10).

The second problem with database copyright protection is that copying data from the database is not infringement if the selection and arrangement in the new work is not substantially similar to the original work. That is, although some protection is afforded against wholesale copying of a substantial part of the database, the database owner is left with little recourse against rearrangement in noninfringing formats or against uses of individual pieces of information.

In response to the difficulties of copyright protection and perceived inequity, several stand-alone database protection regimes have been contemplated in the United States and abroad. Such frameworks would replace the originality requirement of copyright with some form of protection for the effort and investment in accumulation of data. The

most dramatic provision is the European Union's 1996 "Directive on the Legal Protection of Databases," which applies copyright-type protection to certain compilations of data regardless of creative organization (11). So protection for bioinformatics databases in the European Union extends further than in the United States.

Several broad statutes, modeled on the European regime, have been proposed and rejected by Congress. These provisions have been defeated in the United States largely because of the Constitutional need to carefully balance free speech values and proprietary rights. While this balance is built into the United States copyright regime, there has been substantial disagreement regarding how the balance would work in database protection statutes which, effectively, extend copyright protection.

The Bottom Line

Protection of software tools for search and analysis.

Although database records are potentially protectable through the copyright "compilation" concept, the protection is narrow. However there remains a viable copyright theory, requiring careful attention to what is valuable about collections of bioinformatics data. Database users have different goals. For instance, some wish to locate particular sequence or other data, and others are looking for patterns, such as related sequences and matching sequence domains.

Each such goal is advanced by effective search, retrieval, correlation, and analysis of database contents. Those functions are carried out by software which, for instance, might predict the structure of proteins or correlate a vast quantity of experimental results or locate a single relevant sequence. Software is protectable by copyright (12).

Copyright protection of such software can operate as de facto protection of the structure and organization of the database itself in a way that copyright law cannot achieve directly. That is, a computerized database is not a static structure, like a telephone book. Rather, the "organization and structure" of such a database is dependent upon the bounds of functionality of particular software accessing information. That is particularly true of an enormously

Continued on page 64

Bioinformatics Databases continued from page 48

large bioinformatics database — a veritable “tsunami of information” (13).

Public disclosure of an enormous amount of raw, unprotectable data, in some unorganized format, perhaps does not pose a threat of free riders or of loss of investment, if in fact the database owner also controls the most effective software applications to “use” the database. So to the extent that copyright fails in its protection of content, owners of bioinformatics databases may ultimately realize a return on their investment through copyrighted forms of access.

References

- (1) 17 U.S.C. § 102(b).
- (2) *Alexander v Irving Trust Company*, 132 F. Supp. 364, 368 (SD, NY 1995).
- (3) *Silva v MacLaine*, 697 F. Supp. 1423, 1428 (ED Mich. 1988).
- (4) Willem P.C. Stemmer, “How to Publish DNA Sequences with Copyright Protection,” *Nature Biotechnology* 20, 217 (March 2002).
- (5) 17 U.S.C. § 103.
- (6) *Feist Publications, Inc. v Rural Telephone Service Company, Inc.*, 499 U.S. 340, 362 (1991).
- (7) *Matthew Bender and Company v West Publishing Company*, 158 F.3d 674, 682 (2d Cir. 1998)
- (8) *Feist Publications, Inc. v Rural Telephone Service Company, Inc.*, 499 U.S. 340, 363 (1991).
- (9) *American Dental Association. v Delta Dental Plans Association*, 126 F.3d 977, 979 (7th Cir. 1997)
- (10) *Matthew Bender and Company v West Publishing Company*, 158 F.3d 674, 682 (2d Cir. 1998)
- (11) *Kregos v Associated Press*, 937 F.2d 700, 704 (2d Cir. 1991).
- (12) Council Directive 96/9, 1996 O.J. (L77/20).
- (13) Computer Software Copyright Act of 1980, H.R. Rep. No. 96-1307, Part 2, 96th Cong., 2d Sess. (1980), p. 19. **BPI**