

# Copyright Protection for Bibliographic, Numeric, Factual, and Textual Databases

---

JEROME K. MILLER

## Introduction

COPYRIGHT PROTECTION FOR computer programs was specifically excluded from the Copyright Revision Act of 1976 to permit the National Commission of the New Technological Uses of Copyrighted Works (CONTU) to complete its study of the issue. The commission's final report was issued in 1978 and included a recommendation for the revision of Section 117 of the Copyright Act, governing copyright protection for computer programs.<sup>1</sup> Congress accepted CONTU's recommendation and an amendment implementing the recommended changes was attached to the 1980 Patent Revision Act; the copyright portion of the act reads:

Sec. 10(a). Section 101 of title 17 of the United States Code is amended to add at the end thereof the following new language:

A "computer program" is a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result.

(b) Section 117 of Title 17 of the United States Code is amended to read as follows:

§117. Limitations on exclusive rights: Computer programs

Notwithstanding the provisions of §106, it is not an infringement for the owner of a copy of a computer program to make or authorize the making of another copy or adaptation of that computer program provided:

---

Jerome K. Miller, former Assistant Professor, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign and currently President, Copyright Information Services, Champaign, Illinois.

(1) that such a new copy or adaptation is created as an essential step in the utilization of the computer program in conjunction with a machine and that it is used in no other manner, or

(2) that such new copy or adaptation is for archival purposes only and that all archival copies are destroyed in the event that continued possession of the computer program should cease to be rightful.

Any exact copies prepared in accordance with the provisions of this section may be leased, sold, or otherwise transferred, along with the copy from which such copies were prepared, only as part of the lease, sale, or other transfer of all rights in the program. Adaptations so prepared may be transferred only with the authorization of the copyright owner.<sup>2</sup>

The issue of copyright protection for computer programs has been treated in a number of law review articles and papers presented at learned conferences.<sup>3</sup> The court challenges to copyright protection for computer programs center on video displays and the ROM (read only memory) microchips containing programs for arcade-type games.<sup>4</sup> To simplify greatly, the courts found that if the creators of the chips and visual displays fulfill the requirements for registration, deposits, and copyright notices, the copyrights are valid and may be defended in the courts, although many details remain unclear and must be resolved through legislation or litigation. Although copyright protection for computer programs and arcade-type games has been treated by the courts and the scholarly literature, little attention has been given to copyright protection for databases.<sup>5</sup>

### **Copyright Protection for Databases**

Copyright protection for databases falls outside the protection for computer programs included in the new Section 117, as described previously. Databases receive their protection, or lack of protection, under the older provisions in the copyright act, including the provisions on public domain, compilations, derivative works, and original works.<sup>6</sup>

#### *Public Domain Materials*

There appear to be three types of public domain materials. The first, and most obvious, are works of the U.S. government which are not eligible for copyright protection. Section 105 of the copyright act states:

Copyright protection...is not available for any work of the United States Government, but the United States Government is not precluded from receiving and holding copyrights transferred to it by assignment, bequest, or otherwise.<sup>7</sup>

## *Copyright Protection for Databases*

The House Report which accompanied the copyright law states:

The basic premise of Section 105...is the same as that of Section 8 of the present law—that works produced for the U.S. Government by its officers and employees should not be subject to copyright. The provision applies the principle equally to unpublished and published works.<sup>8</sup>

Under this provision, the great volumes of data produced by the Bureau of the Census, the Department of Housing and Urban Development, the Department of Commerce, and others are in the public domain and may be freely used by all, including the database creators and vendors. Section 105 also places judicial decisions, executive documents, and legislative reports in the public domain where they also may be freely used by all, including database creators and vendors.

A second type of public domain materials includes materials which by their nature are not eligible for copyright protection. These may be divided into two groups: (1) older materials in which the copyright has expired, and (2) types of materials which are not eligible for copyright protection. The first of these, materials in which the copyright has expired, include historic data and texts which could be incorporated into numeric, factual, and/or textual databases. At present, few of these materials appear to be available through online databases, but, as these services expand, this may become a consideration. The second group includes items which are specifically excluded from copyright protection for their failure to meet the “originality” test. Section 102 of the copyright act states, “copyright protection subsists...in original works of authorship fixed in any tangible medium of expression....”<sup>9</sup> This statement was taken from the 1909 Copyright Act and is sustained by a number of judicial decisions. The essence of originality is nicely defined in *Doran v. Sunset House Distributing Corporation*:

The requirements for the “originality” necessary to support a copyright are modest. The author must have created the work by his own skill, labor and judgment, contributing something “recognizably his own” to prior treatments of the same subject. However, neither great novelty nor superior artistic quality is required.<sup>10</sup>

The Copyright Office uses the following criteria in rejecting registrations for lack of originality:

### Short Expressions Not Copyrightable

Names, titles, and short phrases or expressions are not...[copyrightable]. The Copyright Office cannot register claims to exclusive rights in brief combinations of words, such as:

- names of products or services;
  - names of businesses, organizations, or groups (including the name of a group of performers);
  - names or pseudonyms of individuals (including a pen name or stage name);
  - titles of works; and
  - catchwords, catch phrases, mottos, slogans, or short advertising expressions.
- This is true even if the name, title, or short phrase is novel, distinctive, or lends itself to a play on words.<sup>11</sup>

This would appear to exclude from copyright protection two types of information commonly found in databases: (1) commonly available data such as the height of the Eiffel Tower or the length of the Amazon River, for which one could not claim originality, and (2) cataloging information. Under the above Copyright Office guidelines, it would appear that a personal or corporate main entry is not eligible for copyright protection and neither is the title of the work. The publisher's name may be a registered trademark but neither the publisher's name, the place of publication, the date of publication, nor the pagination of work appears to be original creations eligible for copyright protection. The only parts of a bibliographic record which appear to be eligible for copyright protection are: (1) the annotation, (2) the classification number, (3) the subject headings or descriptors, and (4) tags and codes and the like. If the classification number, subject headings, or other descriptors are a product of the federal government (e.g., the Library of Congress) then they too are excluded from copyright protection, making the annotation the only part of an individual bibliographic record which is clearly and incontrovertibly eligible for copyright protection. (One should not assume from this that bibliographies and indexes are in the public domain and eligible for unlimited copying, for the organization and arrangement of a database, and especially the program which drives it, are clearly eligible for copyright protection, so wholesale copying of such a database may be culpable, while copying a small number of records from it may be acceptable.<sup>12</sup>) This lack of protection for bibliographic databases undoubtedly contributes to the fact that many vendors do not attempt to obtain copyright protection for their databases and depend on contractual terms to regulate the use and reuse of material taken from them.

The question of eligibility for copyright protection is different from the third situation—the failure to claim copyright protection for databases. This lack of copyright protection for computer programs and databases is of concern to the vendor who has had to rely on contracts to protect computer software.<sup>13</sup> A lack of copyright protection is apparent when one does not see a copyright notice displayed on the screen at sign

## Copyright Protection for Databases

on or sign off or in the opening or closing part of a printout. The failure of the creators or vendors to employ copyright protection may be disadvantageous to them in prosecuting clients who violate the terms of their contracts, but it does not appear to give users any advantages or privileges not contained in the terms of their database contracts. In fact, the contracts employed by database producers and vendors are so powerful they can, and frequently do, prevent users from duplicating public domain materials contained in a database.

### Compilations and Derivative Works

Numeric and factual databases, consisting of information gleaned from a variety of sources and arranged in an orderly fashion designed to maximize their utility to potential users, are a mixed type. Some of these facts and data may be in the public domain in which case they remain in the public domain in spite of their incorporation into a copyrighted database. The copyright law defines a *compilation* as:

[A]...work formed by the collection and assembling of preexisting materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship.<sup>14</sup>

Some databases are *derivative works*, which the law defines as:

[A]...work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted. A work consisting of editorial revisions, annotations, elaborations, or other modifications which, as a whole, represent an original work of authorship, is a "derivative work."<sup>15</sup>

The copyright law further states:

The copyright in a compilation or derivative work extends only to the material contributed by the author of such work, as distinguished from the preexisting material employed in the work, and does not imply any exclusive right in the preexisting material. The copyright in such a work is independent of, and does not affect or enlarge the scope, duration, ownership, or subsistence of, any copyright protection in the preexisting material.<sup>16</sup>

Under the terms of the act, such a compilation or collective work does not have to display a separate copyright notice for each contribution but may use a single copyright notice.<sup>17</sup> Thus, the separate contributions as well as the organizational contributions or headnotes

supplied by the developer or vendor of the database are covered by a single copyright notice. This requirement, designed to simplify copyright registration and notices, makes it impossible for the user to determine whether a database is an original work or a compilation of both copyrighted and uncopyrighted materials. Unless the user is able independently to verify that some information in a database is in the public domain, the user has little choice but to respect the copyrights claimed in the compilation.

### **Databases as Original Works**

Although most existing databases are compilations, some may be original works of authorship and as such they receive full protection of the copyright law.<sup>18</sup> Under these terms, the copyright proprietor has the right to reproduce additional copies, to prepare derivative works, to distribute copies of the work to the public by sale, rental, lease, or lending, and to publicly display or perform the work.<sup>19</sup> These copyrights have a duration of life plus fifty years for works of individual authors and seventy-five years for work created by corporate bodies.<sup>20</sup> These exclusive rights provided by Section 106 are modified by the users' rights contained in Section 107—on fair use.

### **Fair Use**

Under the terms of Section 107, the user may reproduce part (rarely all) of the copyrighted work for purposes of criticism, comment, news reporting, teaching, scholarship, or research. Section 107 provides four criteria for evaluating such uses:

1. the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
2. the nature of the copyrighted work;
3. the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
4. the effect of the use on the potential market for or value of the copyrighted work.<sup>21</sup>

There is an extensive literature on the application of the four fair use criteria to the duplication of printed and audiovisual works, especially their application to copying by educators.<sup>22</sup> It appears that the courts have not been called upon to handle infringement cases involving databases, so there is no judicial guidance to assist in the application

## *Copyright Protection for Databases*

of the fair-use guidelines to their duplication. Until the courts begin to provide guidance in this matter, caution is advised. There is little doubt that one may obtain some bibliographic, factual, numeric, or textual data from a database and incorporate it into a book or article which is largely the work of the author. The problem particularly arises when a database user who has access to a minicomputer or smart terminal is able to obtain a substantial body of information from one or more databases, store it in memory, then manipulate the data to create a new work which appears at least superficially to be eligible for copyright protection as an original work.<sup>23</sup> Until the courts indicate to the contrary, one must assume this is an infringement of the copyrights in the works which were duplicated.

Applying fair use to copyrighted databases centers on downloading, that is, copying information from a database onto electronic memory. Downloading can probably be divided into two categories—hardcore and softcore. Softcore downloading is the *temporary* retention of records for the purposes of merging files, purging duplicate records and editing to make the finished product more useful to the end user. Hardcore downloading is retaining records (with or without merging, purging or editing) to avoid paying legitimate user fees. Traditional copyright scholars and most publishers' representatives view most forms of downloading as an infringement. Their objections stem from two common downloading practices—creating local data files and merging or editing files. Merging or editing creates a derivative work, one of the creator's exclusive rights, and it is an infringement, but it is a common practice in information centers and the database creators and vendors don't seem to object to it as long as it does not deprive them of their fees. Hardcore downloading is another matter entirely. Creating a local database to avoid paying database service fees is a clearcut and unequivocal infringement of the copyrights and it undoubtedly violates the terms of the user's contract. Some database producers and vendors are beginning to offer annual downloading licenses, and libraries and information centers that want to download to create a local database should obtain downloading licenses before creating local files.

### **Copyright Protection for Locally Produced Databases**

Securing copyright protection for locally produced databases is reasonably simple so long as three points are covered: notice, registration and deposit.

The copyright notice consists of three parts which normally appear

on a single line: (1) the word "copyright" or the © symbol, (2) the name of the copyright owner, and (3) the year of creation.<sup>24</sup> This would appear as: "Copyright, North-South Data Service, 1982." If the database is revised or expanded, the notice must be updated to include the year of creation and *each* year the database is revised or expanded. In this instance the notice might read: "Copyright, North-South Data Service, 1982, 1983, 1984."

[The] notice of copyright...shall be placed on all publicly distributed copies from which the work can be visually perceived, either directly or with the aid of a machine or device....

The notice shall be affixed to the copies in such manner and location as to give reasonable notice of the claim of copyright. The Register of Copyrights shall prescribe by regulation as examples, specific methods of affixation and positions of the notice on various types of works that will satisfy this requirement. But these specifications shall not be considered exhaustive.<sup>25</sup>

In the case of machine-readable copies, the Copyright Office identifies four suitable locations for the notice:

- (1) A notice embodied in the copies in machine-readable form in such a manner that on visually perceptible printouts it appears either with or near the title, or at the end of the work;
- (2) A notice that is displayed at the user's terminal at sign on;
- (3) A notice that is continuously on terminal display; or
- (4) A [permanently] legible notice reproduced...on a gummed or other label securely affixed to the copies or to a box, reel, cartridge, cassette, or other container used as a permanent receptacle for the copies.<sup>26</sup>

The first three are appropriate for online databases but the fourth format does not appear to meet the "visually perceived" requirement if the user is unlikely to see the box, cartridge or container. (The fourth requirement may be suitable for a database contained on a diskette designed for microcomputers, which will be handled each time it is loaded.)

Registration consists in completing "Form TX: Nondramatic Literary Works," and submitting the completed form, the \$10 registration fee, and the deposit copy to the Copyright Office. ("Literary works," is a catch-all classification for everything except sound recordings, visual arts, and performance materials.) The forms are available free of charge by calling, day or night, 202/287-9100. An answering machine records requests and the forms are mailed promptly, accompanied by a pamphlet explaining the forms and the registration procedures. If the information in the pamphlet and on the forms is unclear,

## Copyright Protection for Databases

help is available from the Copyright Office by calling 202/287-8700 during office hours.<sup>27</sup>

The deposit requirements for machine-readable works are quite different from the requirements for most other works. The Copyright Office does not have the equipment to handle most machine-readable works, so such works deposited at the Copyright Office must be reproduced in printed form or on microform.<sup>28</sup> The deposit requirements for databases depend on (1) whether this is a new or a revised work, and (2) whether it is a single- or multiple-file database. To satisfy the deposit requirement for a new, single-file database, one must prepare a single printout or microform copy of the first and last twenty-five pages or similar units of the database. These pages must include the title, copyright notice and other identifying materials (e.g., address, distributor, sources, online vendor).<sup>29</sup> If an adhesive-label copyright notice is attached to reels, boxes, disks, cassettes, or the like, a copy of this notice must accompany the deposit copy.<sup>30</sup>

When the registration form, the check for \$10 and the deposit copy are ready, these three items, plus the adhesive-label copyright notice, if any, must be packed in a *single* container and mailed to: The Copyright Office, Library of Congress, Washington, D.C. 20559.<sup>31</sup> Registration should be made before release, or within three months after publication or initial online distribution.<sup>32</sup> Either confirmation of the registration or a request for additional information should be returned in about three weeks.

The deposit requirements for multifile and revised databases are complex<sup>33</sup> and it may be advisable to hire an attorney who specializes in copyright matters (most attorneys are unfamiliar with copyright) or a nonattorney, copyright consultant to assist with the process.

## ACKNOWLEDGMENT

The author wishes to acknowledge the assistance of two colleagues, Lawrence W.S. Auld and Carol Tenopir, in preparing this article.

## References

1. *United States Code*. Supplement V, Title 17. "General Revision of Copyright Law." 94th Cong. 2d Sess, 19 Oct. 1976. Washington, D.C.: USGPO, 1981 (hereafter cited as Copyright Act). CONTU completed its work in 1978 and its recommendations appear in its *Final Report of the National Commission on New Technological Uses of Copyrighted Works*. Washington, D.C.: CONTU, 1978).

2. Copyright Act, §117 (the amended version passed as Public Law 96-517, U.S. Congress, Title 35, "Patents." 96th Cong., 2d Sess., 12 Dec. 1980).
3. The most comprehensive treatment appears in Bender, David. *Computer Law: Evidence and Procedure*. New York: Matthew Bender, 1978- (loose-leaf annual), chapt. 4, pp. 4.1-app. 4A-99.
4. Several of these cases are treated in Billings, Roger D., Jr. "Fair Use Under the 1976 Copyright Act: The Legacy of Williams & Wilkins for Librarians" (in this issue of *Library Trends*).
5. The most comprehensive treatment appears in: Keplinger, Michael S. "Computers and Copyright." In *Computer Programs and Data Bases: Perfecting, Protecting and Licensing Proprietary Rights after the 1980 Copyright Amendments*, edited by Daniel T. Brooks and Michael S. Keplinger, pp. 59-73. New York: Law & Business, 1981.
6. The respective sections of the act are treated separately in later sections.
7. Copyright Act, §105.
8. U.S. House of Representatives, *Report No. 94-1476*, §105.
9. Copyright Act, §102.
10. *Doran v. Sunset House Distributing Corporation. Federal Supplement: vol. 197*. St. Paul, Minnesota: West, 1962, p. 944.
11. U.S. Library of Congress. Copyright Office. *Circular R34: Copyright Protection Not Available for Names, Titles, or Short Phrases*. Washington, D.C.: USGPO, 1979.
12. Copyright Act, §117.
13. OCLC has just begun claiming copyright protection for its database. "Good Morning. © OCLC." *American Libraries* 14(Feb. 1983):74.
14. Copyright Act, §101
15. *Ibid.*
16. *Ibid.*, §103(b).
17. *Ibid.*, §104(a).
18. U.S. Congress. House. *Report No. 94-1476: Original Works of Authorship*, sec. 102.
19. Copyright Act, §106.
20. *Ibid.*, §§302(a) and (c).
21. *Ibid.*, §107.
22. Miller, Jerome K. *Applying the New Copyright Law: A Guide for Educators and Librarians*. Chicago: ALA, 1979.
23. Addressing this problem requires a careful application of the laws of copyright, contract, and unfair practices, plus a speedy and reasonably priced permission procedure offered by the database owners and vendors. As smart terminals and personal computers proliferate, the reuse of information obtained from databases is sure to become a controversial issue. It is reliably reported that OCLC's decision to claim copyright protection for its database was prompted by widespread misuses by member libraries who permitted nonmembers to build databases without paying OCLC for the information or services. This appears to be a violation of the OCLC-user contract, but copyright protection for the database may provide additional legal support if OCLC sues infringers.
24. Copyright Act, §401(b).
25. *Ibid.*, §§ 401(a) and (c).
26. U.S. Office of the Federal Register. *Code of Federal Regulations*, title 37, §§201.20 (g), "Works Reproduced in Machine-Readable Copies."
27. U.S. Library of Congress. Copyright Office. *Announcements ML-255*. Washington, D.C.: Copyright Office, 1981.
28. U.S. Office of the Federal Register. *Code of Federal Regulations*, Title 37, §202.20 (c), pt. 2, vol. vii.
29. *Ibid.*, §202.20(c), pt. 2, vol. vii, letter A.
30. *Ibid.*, §202.20(c), pt. 2, vol. vii, letter B.
31. *Ibid.*, §202.3(c), pt. 2.

*Copyright Protection for Databases*

32. Copyright Act, §§411, 412.
33. U.S. Office of the Federal Register. *Code of Federal Regulations*, title 37, §202.20 (c), pt. 2, vol. vii, letter B.

This Page Intentionally Left Blank