

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Copyright, Fair Use, Scholarly Communication, etc.

Libraries at University of Nebraska-Lincoln


8-2017

Metadata and Copyright: Should Institutions License Their Data about Scholarship?

Krista L. Cox

Association of Research Libraries

Follow this and additional works at: <http://digitalcommons.unl.edu/scholcom>

 Part of the [Intellectual Property Law Commons](#), [Scholarly Communication Commons](#), and the [Scholarly Publishing Commons](#)

Cox, Krista L., "Metadata and Copyright: Should Institutions License Their Data about Scholarship?" (2017). *Copyright, Fair Use, Scholarly Communication, etc.*. 59.

<http://digitalcommons.unl.edu/scholcom/59>

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Copyright, Fair Use, Scholarly Communication, etc. by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Issue Brief

ASSOCIATION
OF RESEARCH
LIBRARIES

Metadata and Copyright: Should Institutions License Their Data about Scholarship?

August 2017



Krista L. Cox, Director of Public Policy Initiatives, Association of Research Libraries

Goals in Sharing Metadata

The Association of Research Libraries (ARL) is committed to promoting open scholarship, including making metadata—foundational information about scholarly works—as widely available as possible. Allowing the broadest access to and reuse of metadata advances scholarship, research, discovery, and innovation.

In order to ensure the widest possible uses of metadata, institutions should share their metadata in a way that removes restrictions. However, institutions should also keep in mind that metadata can be shared across borders, a fact that may impact decisions on how to share it.

Is Metadata Copyrightable?

In the United States, copyright protection is granted for creative expression and not for underlying facts, thus raising questions about whether such protection would be afforded to metadata. Metadata includes factual fields in library catalog or other records, generally useful as finding aids, such as the creator’s name, the date of creation, the date of publication, and the title of the work.¹ These short factual pieces of data themselves are not copyrightable because they are short phrases and pure facts.

Compilation of facts, even though aggregation and organization may be useful, is not enough to warrant copyright protection. In *Feist Publications v. Rural Telephone Service*, the Supreme Court confirmed that “sweat of the brow” is insufficient to serve as a basis for copyright protection.² The fact that substantial time or money may be invested in compilation of data or facts—such as

in Rural Telephone Service’s compilation of a phone directory—is irrelevant to the question of copyrightability.³ The Court determined that the creation of an alphabetized list of subscribers did not result in creative expression protected by copyright because it was merely an obvious, non-original compilation.⁴

However a possibility remains that particular compilations of facts, if they exhibit sufficient creativity in their selection, coordination, and arrangement of data, may include enough original expression to arguably warrant copyright protection. Nevertheless, even in these cases, the underlying facts remain free to use. Thus, the metadata for any one work likely would not be copyrightable, but a database of metadata could be. Copyright in the database would not prohibit extraction of individual facts, however.

Some countries, like Australia, similarly reject the “sweat of the brow” doctrine and require original expression for copyright protection. In Canada, database protection may be available for a compilation of the author’s own data or records, but not for data compiled from another source.⁵

While the United States does not have any *sui generis* (or unique) protection for unoriginal databases, other countries do provide such protections. The European Union, for example, has a Database Directive, which provides for 15 years of protection for databases even if they do not reflect protectable expression. The Database Directive’s *sui generis* protection is based on a “sweat of the brow” premise. In the EU, the *sui generis* protection provided under the Database Directive is separate from any other copyright that may be granted and the Directive also requires copyright protection for original databases. Russia similarly affords legal protections for databases. Even aside from such legal protections of databases, some countries may apply the “sweat of the brow” doctrine and afford copyright protection to metadata.

Policies and Community Norms

In the United States, a number of policies and community norms have been established around licensing of metadata, generally with a goal of promoting free reuse and adaptation. Most policies use a Creative Commons Zero (CC0) license or recommend, but do not require, the use of CC0 with attribution.⁶ Many institutions and communities use the CC0 license for metadata to make it clear that the metadata is free to reuse. While these policies—such as the use of a license—may be enforceable, similar community norms act as guidelines without the mandate of compliance.

Some institutions make their data available simply through a CC0 license. For example, the University of Michigan Libraries has made the 1.3 million bibliographic records it created available under CC0.⁷

Many institutions use CC0 but have additional guidelines or recommendations, such as those recommending, but not requiring, attribution. These institutions continue to use CC0 rather than a Creative Commons Attribution (CC-BY) license to ensure that the metadata is widely available, recognizing that attribution may not be possible or practical.

For example, New York Public Library (NYPL) provides: “To the extent that NYPL has a copyright interest in Metadata Records, a Creative Commons CC0 1.0 Universal Public Domain Dedication will apply.”⁸ NYPL’s policy also includes a number of community norms, such as requesting attribution “to the extent it is technologically feasible to do so,” and asking users to make improvements freely available on the same terms as NYPL provides.⁹ The policy further notes that for metadata obtained from the OCLC WorldCat Database, one should act in accordance with the community norms of that database.¹⁰

The Digital Public Library of America (DPLA) releases its metadata under a CC0 license, but encourages users to

follow its community guidelines. DPLA encourages “giving attribution to all the sources of the metadata.”¹¹

Others do not have a firm policy requiring a CC0 license, but nonetheless encourage the use of this public domain dedication. The University of California (UC) Libraries has a metadata policy, which notes that while “the UC Libraries are not obligated to share their metadata, sharing is encouraged.” The policy states that the “Libraries will place the fewest possible restrictions on the reuse of metadata they share. The preferred standard under which to share UC Libraries metadata is CC0... UC Libraries may request appropriate attribution as the source of their shared metadata, to the extent such attribution is technologically feasible, in accordance with community norms. The provision of a CC0 standard has the additional benefit of removing any uncertainty about reuse.”¹²

A number of libraries across the world similarly employ the use of the CC0 public domain dedication to their bibliographic metadata, including the British Library, National Library of Spain, and Swedish National Library, among others.¹³

Recommendation

Institutions have a number of options for how to treat metadata while encouraging widespread use and sharing. Some institutions may determine that no license is necessary to openly share metadata, taking the position that the metadata they produce is not copyrightable. Others might share it using a CC0 license or CC0 with additional guidelines. In determining how to share metadata, institutions should consider possible cross-border implications as well as the potential that what is defined as metadata could grow in breadth and raise additional copyright concerns. Additionally, institutions should take into account the benefit that clear and easy-to-understand policies provide to users. Those wishing to rely on metadata in institutional

databases often desire certainty about their uses and reuses, including whether use of the data across borders is permitted.

Although metadata is arguably not subject to copyright in the United States, to facilitate the use of metadata across borders and avoid later arguments or confusion, some institutions make their metadata available under a CC0 license. They have adopted a simple policy statement noting that, to the extent that there may be copyrightable expression in the metadata, that copyright is being made fully available through a CC0 grant. Such a statement acknowledges that there may not be any copyright protection in the metadata while simultaneously ensuring that any doubt is removed over whether such metadata may be shared or reused, particularly because the existence of a large number of institutional policies relying on a CC0 grant for metadata may cause some users to seek guidance under which the metadata may be used and shared. For those institutions seeking to license metadata, CC0 is a well-known license that is designed to be recognized and useable in almost all countries around the world and may therefore be preferred over general statements of public domain dedication or the use of other licenses.

Endnotes

¹ Longer descriptions, abstracts, or expositions might also be considered by some to be metadata. These longer descriptions may well indeed rise to the level of creative expression and could be protected by copyright, unlike short factual pieces of information.

² *Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340 (1991).

³ *Id.*

⁴ *Id.*

- ⁵ *CCH Canadian Ltd. v. Law Society of Upper Canada* (2004).
- ⁶ The Creative Commons Zero (CC0) license is the most open Creative Commons license. CC0 is a public domain dedication, allowing all further uses of the material. It allows users to use, copy, modify, adapt, and distribute the material. See <https://creativecommons.org/share-your-work/public-domain/cc0/>.
- ⁷ University of Michigan Library, “Open Access Bibliographic Records Available for Download and Use,” modified November 23, 2015, <https://www.lib.umich.edu/open-access-bibliographic-records>.
- ⁸ New York Public Library, “NYPL Policy on Open Bibliographic Metadata,” revised March 2013, <https://www.nypl.org/help/about-nypl/legal-notice/open-metadata>.
- ⁹ *Id.* (“Specifically, NYPL requests that you do not claim any legal right in or impose any legally binding conditions on the Metadata Records or your improvements.”)
- ¹⁰ *Id.*
- ¹¹ Digital Public Library of America, “DPLA Data Use Best Practices,” accessed August 10, 2017, <https://dp.la/info/wp-content/uploads/2013/04/DPLADataUseBestPractices.pdf>.
- ¹² University of California Libraries, “University of California Libraries Metadata Sharing Policy,” revised May 8, 2015, http://libraries.universityofcalifornia.edu/groups/files/sag2/docs/Final_UC_Libraries_Metadata_Sharing_Policy_May8_2015.pdf.
- ¹³ Creative Commons, “CC0 Use for Data,” accessed August 10, 2017, https://wiki.creativecommons.org/wiki/CC0_use_for_data

Association of Research Libraries

21 Dupont Circle, NW
Suite 800
Washington, DC 20036
T 202.296.2296
F 202.872.0884

ARL.org
pubs@arl.org

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

