Copyright and provenance

Some practical problems

John Mark Ockerbloom
Principles of Provenance workshop
June 26, 2007
Our resources are going digital

Google Is Adding Major Libraries to Its Database

By JOHN MARKOFF and EDWARD WYATT
Published: December 14, 2004

Google, the operator of the world's most popular Internet search service, plans to announce an agreement today with some of the nation's leading research libraries and Oxford University to begin converting their holdings into digital files that would be freely searchable over the Web.

It may be only a step on a long road toward the long-predicted global virtual library. But the collaboration of Google and research institutions that also include Harvard, the University of Michigan, Stanford and the New York Public Library is a major stride in an ambitious Internet effort by various parties. The goal is to expand the Web beyond its current valuable, if eclectic, body of material and create a digital card catalog and searchable library for the world's books, scholarly papers and special collections.

John Ockerbloom

UNIVERSITY OF PENNSYLVANIA LIBRARY

June 26, 2007
But not without a fight

Publishers sue Google over book search project

By Aline Gilbert
Staff Writer, CNET News.com
Published: October 19, 2005, 9:37 AM PDT

The Association of American Publishers has filed a lawsuit against Google, alleging the Internet company's plans to scan and digitally distribute the text of major library collections would violate copyright protections.

The group filed suit after lengthy discussions with Google's management about the company's Print Library Project broke down, the AAP said on Wednesday.

As part of the project, Google is working to scan all or parts of the book collections of the University of Michigan, Harvard University, Stanford
What can I do with a work?

- Often simply represented as a binary value “restricted access” / “no restrictions”
- But really derives from a series of copyright questions
  - What copyrights apply to this work?
    » Which are still in force? Which have expired?
  - Who holds the copyrights?
    » And how can they be contacted?
  - What permissions apply for use?
    » To specific parties (first serial rights to a magazine..)
    » General (GPL, Creative Commons…)
    » Granted by law (compulsory license, fair use, preservation, orphan works…)
Google’s conservative algorithm for showing more than snippets

Boolean can_show_full_without_permission
     (client, publication_place, publication_date) {
    If (client in US) {
        If ((publication_place in US) {
            return true if (publication_date < 1923);
        } else {
            return true if (publication_date < 1909);
        }
    } else {
        return true if (publication_date < 1864);
    }
    return false;
}
A portion of a copyright expiration flowchart

FLOWCHART FOR DETERMINING WHEN U.S. COPYRIGHTS IN FIXED WORKS EXPIRE

EFFECT OF COPYRIGHT NOTICE
Published before March 1, 1989. Publication without copyright notice caused works to enter public domain subject to certain exceptions.
Published on or after March 1, 1989. Copyright notice not required.

Yes: Expire
No

Yes

Published before January 1, 1950?

Published before January 1, 1948?

Published before January 1, 1947?

Published before January 1, 1946?

Published before January 1, 1945?

Published before January 1, 1944?

Published before January 1, 1943?

Published before January 1, 1942?

Published before January 1, 1941?

Published before January 1, 1940?

Published before January 1, 1939?

Published before January 1, 1938?

Published before January 1, 1937?

Published before January 1, 1936?

Published before January 1, 1935?

Published before January 1, 1934?

Published before January 1, 1933?

Published before January 1, 1932?

Published before January 1, 1931?

Published before January 1, 1930?

Published before January 1, 1929?

Published before January 1, 1928?

Published before January 1, 1927?

Published before January 1, 1926?

Published before January 1, 1925?

Published before January 1, 1924?

Published before January 1, 1923?

Yes

Expires later of:
1. December 31, 2047, or
2. Date of expiration that applies to work as created (fixed) before January 1, 1978 (see bottom right of chart).

No

Published before January 1, 1979?

Published before January 1, 1978?

Published before January 1, 1977?

Published before January 1, 1976?

Published before January 1, 1975?

Published before January 1, 1974?

Published before January 1, 1973?

Published before January 1, 1972?

Published before January 1, 1971?

Published before January 1, 1970?

Published before January 1, 1969?

Published before January 1, 1968?

Published before January 1, 1967?

Published before January 1, 1966?

Published before January 1, 1965?

Published before January 1, 1964?

Expires later of:
1. December 31, 2032, or
2. Date of expiration that applies to work as created (fixed) before January 1, 1978 (see bottom right of chart).

No

Published before January 1, 1963?

Published before January 1, 1962?

Published before January 1, 1961?

Published before January 1, 1960?

Published before January 1, 1959?

Published before January 1, 1958?

Published before January 1, 1957?

Published before January 1, 1956?

Published before January 1, 1955?

Published before January 1, 1954?

Published before January 1, 1953?

Published before January 1, 1952?

Published before January 1, 1951?

Published before January 1, 1950?

Published before January 1, 1949?

Published before January 1, 1948?

Published before January 1, 1947?

Published before January 1, 1946?

Published before January 1, 1945?

Published before January 1, 1944?

Published before January 1, 1943?

Published before January 1, 1942?

Published before January 1, 1941?

Published before January 1, 1940?

Published before January 1, 1939?

Published before January 1, 1938?

Published before January 1, 1937?

Published before January 1, 1936?

Published before January 1, 1935?

Published before January 1, 1934?

Published before January 1, 1933?

Published before January 1, 1932?

Published before January 1, 1931?

Published before January 1, 1930?

Published before January 1, 1929?

Published before January 1, 1928?

Published before January 1, 1927?

Published before January 1, 1926?

Published before January 1, 1925?

Published before January 1, 1924?

Published before January 1, 1923?

Yes: Expire
No

Created (fixed) before January 1, 1979?

Created (fixed) before January 1, 1978?

Created (fixed) before January 1, 1977?

Created (fixed) before January 1, 1976?

Created (fixed) before January 1, 1975?

Created (fixed) before January 1, 1974?

Created (fixed) before January 1, 1973?

Created (fixed) before January 1, 1972?

Created (fixed) before January 1, 1971?

Created (fixed) before January 1, 1970?

Created (fixed) before January 1, 1969?

Created (fixed) before January 1, 1968?

Created (fixed) before January 1, 1967?

Created (fixed) before January 1, 1966?

Created (fixed) before January 1, 1965?

Created (fixed) before January 1, 1964?

Created (fixed) before January 1, 1963?

Created (fixed) before January 1, 1962?

Created (fixed) before January 1, 1961?

Created (fixed) before January 1, 1960?

Created (fixed) before January 1, 1959?

Created (fixed) before January 1, 1958?

Created (fixed) before January 1, 1957?

Created (fixed) before January 1, 1956?

Created (fixed) before January 1, 1955?

Created (fixed) before January 1, 1954?

Created (fixed) before January 1, 1953?

Created (fixed) before January 1, 1952?

Created (fixed) before January 1, 1951?

Created (fixed) before January 1, 1950?

Created (fixed) before January 1, 1949?

Created (fixed) before January 1, 1948?

Created (fixed) before January 1, 1947?

Created (fixed) before January 1, 1946?

Created (fixed) before January 1, 1945?

Created (fixed) before January 1, 1944?

Created (fixed) before January 1, 1943?

Created (fixed) before January 1, 1942?

Created (fixed) before January 1, 1941?

Created (fixed) before January 1, 1940?

Created (fixed) before January 1, 1939?

Created (fixed) before January 1, 1938?

Created (fixed) before January 1, 1937?

Created (fixed) before January 1, 1936?

Created (fixed) before January 1, 1935?

Created (fixed) before January 1, 1934?

Created (fixed) before January 1, 1933?

Created (fixed) before January 1, 1932?

Created (fixed) before January 1, 1931?

Created (fixed) before January 1, 1930?

Created (fixed) before January 1, 1929?

Created (fixed) before January 1, 1928?

Created (fixed) before January 1, 1927?

Created (fixed) before January 1, 1926?

Created (fixed) before January 1, 1925?

Created (fixed) before January 1, 1924?

Created (fixed) before January 1, 1923?

Yes: Expire
No

Work for hire, anonymous, or pseudonymous?
Costs to produce digital works

• Digitization costs
  – $0.01 - $0.10/page at scale
  – $5-$50 for a 500-page book

• Copyright clearance costs
  – Books are relatively simple; magazines, recorded music, film, typically involve lots of rightsholders
An optimization problem in multiple dimensions

- Maximize value of collection
  - (include as many valuable resources as you can, with broadest rights possible)
- Maximize throughput of rights-clearing
  - (to build a large collection fast)

- Minimize cost of rights-clearing
- Minimize risk of legal penalties
  - (which can be very large in worst case)
Some provenance questions in copyright determination

• Provenance of work
  – Who created the work? (And when did they live?)
  – When and where was the work first published?
    » (inside and outside the US)
  – Does the work include or derive from other copyrighted works? Which ones?

• Provenance of rights
  – Was there a copyright notice? What did it say?
  – Is this a work for hire? For whom?
  – Was the work registered for copyright? When? Renewed?
  – Was copyright transferred? If so, when and to whom?
  – Record of rights permission grants?

• Determination can work from copyright knowledge bases or from works, involve negative as well as positive assertions with various degrees of certainty
An example
Copyright registrations

• Rightsholders filled out registration forms
• Form copies in registration books in DC
  – The earliest saved artifact
• Form data summarized in card files in DC
  – What Copyright Office staff search
• Card files summarized in Catalog of Copyright Entries in various libraries
  – What folks outside DC tend to search
• Digitizations, databases produced
  – Page Image index at Penn; Stanford’s Determinator database
• Digitization cost depends on where you start
Some recent developments

• OCLC proposed project to associate copyright-related assertions with WorldCat entries in global union catalog

• Proposals to exempt “orphan works” from infringement liabilities
  – But need to conduct “reasonably diligent search”; and document that the search was done
What do we need, and how can provenance research help?

• Theoretical foundations for evaluating sufficiently reliable information chains
  – What facts support our copyright determinations, and what is their provenance?

• Practical issues of scaling up evaluation
  – (preferably with automation)

• Common representation for copyright assertions and searches, and their provenance
  – Accommodating distributed contributions, evaluation

• Simple, cheap methods of storing and querying this information