



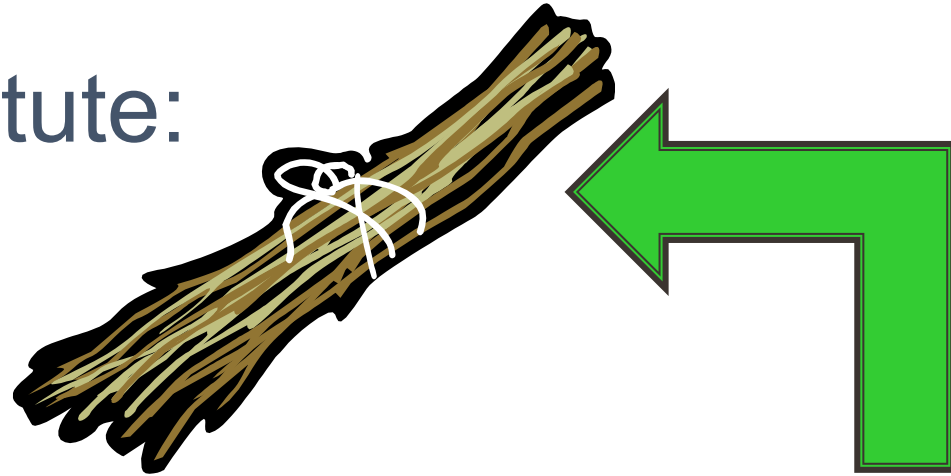
TEXT AND DATA MINING ON BOTH SIDES OF THE ATLANTIC

@KyleKCourtney
#ICEPOPS2019
Edinburgh, Scotland



Fair Use interpreted as a Right

Preamble to fair use statute:



“Notwithstanding the provisions of **section 106 ...the fair use of a copyrighted work...for purposes such as **criticism, comment, news reporting, teaching, scholarship, or research**” is not an infringement of copyright.**

Fair Use Factors

Four Factors of Fair Use

- (1) the purpose and character of the use;
- (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 upon the potential market for or value of the copyrighted work.

Development of *Transformative Fair Use*

- (1) Does the use **transform the material**, by using it for a **different purpose**?
- (2) Was the **amount taken appropriate** to the new purpose?

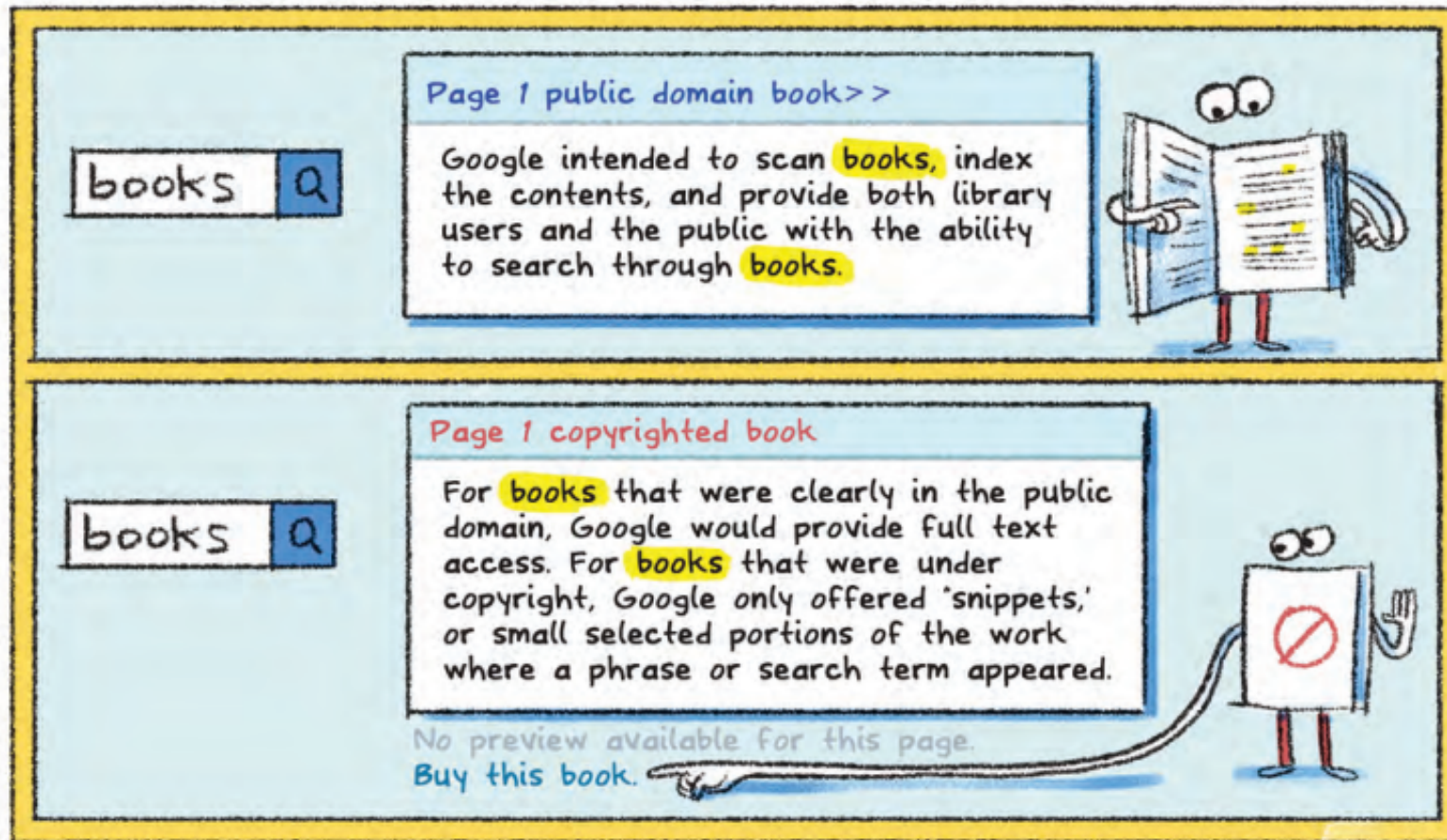
Google Books and HathiTrust Cases

*Transformative Fair
Use applied to text
and data mining?*

****Yes!****



Google Books offered both public domain works and “snippets” of in-copyright works



HATHI
TRUST

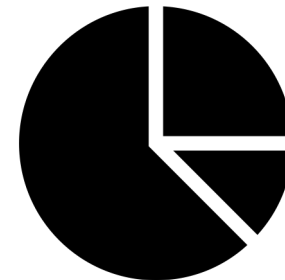
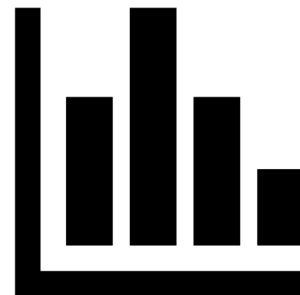
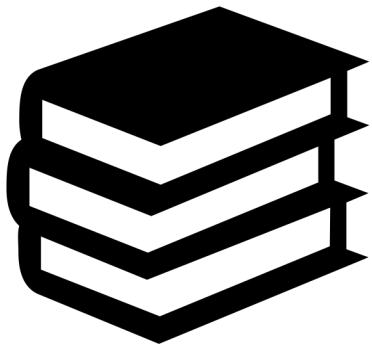
First Factor: Transformiveness

Original purpose?

The book's topic.

New transformative purpose?

Text and data mining for research, not the book's topic!





Commercial use was less significant because of the high transformativeness of the new use

Third Factor: Copying the entire text *was necessary to serve Google Book's new, transformative purpose*



When considering the third factor, the court noted that Google scanned and copied entire works, which generally weighed against fair use.

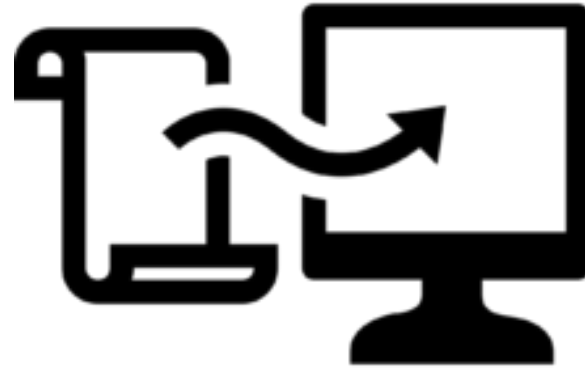
However, in order to allow the user to search millions of books for a word or phrase—it would take years to do that in a library—Google had to scan the entirety of the work.

A transformative work, like a TDM project, “is one that serves a new and different function from the original and is not a substitute for it ... the result of a search is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn...”

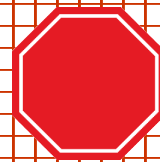
The complete text was necessary to serve the new, transformative purpose.

The court also recognized that “full-work reproduction is critical to the functioning of Google Books,” and that Google limits the amount of text visible to users through their search interface.

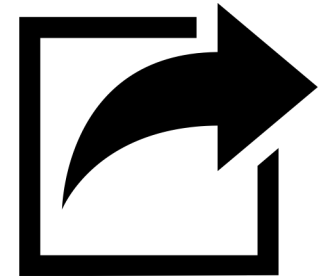
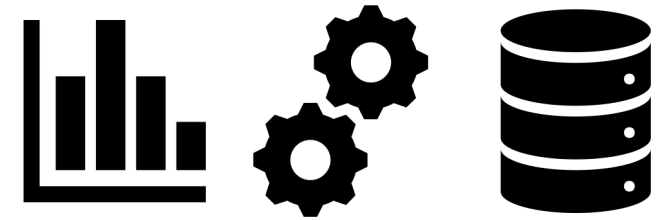
However....



Access License



TDM Project





“The right to read is the right to mine.”

“Anyone who has lawful access to read the literature with their eyes should be able to do so with a machine.”

- Peter Murray-Rust, University of Cambridge

United Kingdom

- Section 29A of the Copyright, Designs and Patents Act 1988 (2014)
- TDM is allowed to be performed providing the following two criteria are satisfied:
 - 1. The TDM practitioner has **lawful access** to the resource.*
 - 2. The analysis is undertaken for the purposes of **non-commercial research**.*

“To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable.”



Lawful Access and Right to Contract

- **Lawful Access:** Guidance published suggests that one has lawful access to the work *if he or she has a right to read it.*
- **Right to Contract:** a contractual clause (e.g. in a license or in terms of service) that prevents or restricts TDM is unenforceable

French TDM Law

- Art. L. 122-5, 10° of the Code de la Propriété Intellectuelle (CPI):
 - *Rightholders cannot forbid "Copies and digital reproductions made from a **lawful source** for the purposes of mining text and data included in or associated with **scientific publications**, for **public research purposes**, excluding all commercial purposes."*

French TDM Law Problems

- No contract override
- Limited to scientific research
- Limited to “research organizations”
- “Lawful access” undefined

European Union: Article 3 of the DSM

- Art. 3 has a mandatory TDM exception which would allow **research organizations** and cultural heritage institutions to text and data mine works to which they have **lawful access** for the purposes of **scientific research**

EU Article 3 Problems

- No contract override
- Limited to scientific research
- Limited to “research organizations”
- “Lawful access” undefined

No TDM Uniformity; Potential Critical Issues

Technical measures should not be used to prevent anyone from exercising TDM rights

API, licensing access, payment, DRM, etc.

Allow UK-like **contract override**

Lawful access = the right to read is the right to mine.

Allow both **commercial and non-commercial** uses

No reason to limit TDM to non-commercial purposes.

Research supports innovation which is something that has commercial use.

No limitation of **TDM purpose** (scientific, public research, etc.)

RISK?

TDM provides invaluable tools for sifting, organizing and analyzing vast amounts of text.

An “essential method of knowledge discovery in the life sciences, the humanities, and linguistics” and a “building block of machine learning and internet search technology.”