

Text and Data Mining (TDM): Copyright and Licensing Issues

Lesley Ellen Harris, Copyrightlaws.com

SLA Conference, Philadelphia, 12 June 2016



Agenda

- Copyright versus licensing
- U.S. and Canadian copyright legislation
- Fair use and fair dealing
- Digital licensing agreements
- Best practices

TDM

Definition:

- copying whole works
- extracting from existing content such as text, numbers, images, video
- curating and analyzing

Questions:

- is this allowed under copyright laws?
 - is this allowed under your license?
 - is this a permanent or temporary copy? Are you making a copy or using unprotectable information? Are you keeping a copy of the articles for their TDM analysis?
-

Copyright versus Licensing

- Copyright sets out exclusive rights subject to limitations
- Copyright is federal legislation
- Contracts/licenses are between the parties that sign them
- Contracts are negotiable
- Contracts are state or provincial

U.S. and Canadian Copyright Legislation

- No specific provisions or statutory exceptions for TDM

Fair Use (U.S.)

- Does it allow TDM for noncommercial research? What about commercial use?
- Can be overridden by license
- See “Fair Use in Text and Data Mining: ARL Publishes Issue Brief” by the Association of Research Libraries at <http://www.arl.org/news/arl-news/3643>

	Favoring Fair Use	Not Favoring Fair Use
Purpose	Nonprofit educational use, transformative (e.g., if results not public available)	Commercial, profiting from use
Nature	Factual, published	Fiction, unpublished
Amount	Small amount, not central amount, proportionate for purpose	Large amount, central to original
Effect	No effect on marketplace	Could replace sale of original, could impact marketplace

Fair Dealing (Canada)

- Does it allow TDM for noncommercial research?
What about commercial use?
- Can be overridden by license



Determining Fairness in Canada

- Purpose: research, private study, education, parody or satire, criticism, review or news reporting
 - Character of copying: including whether it involves single for a legitimate purpose or multiple copies
 - Amount: quantity and quality
 - Alternatives: could a non-infringing copy be used?
 - Nature of the work: is the original published or unpublished?
 - Effect of copying: will the copy compete with the commercial market of the original work?
-

What is a Digital Licensing Agreement?

- A license is permission
- No transfer of ownership
- Permission required
- Permitted uses are negotiated
- Use of content is according to terms and conditions

Primer on Digital Licensing Agreements

- Parties, licensor and licensee, description of the content
- Definitions
- Permitted Uses
- Non-Permitted Uses
- Authorized Users
- Warranty/Indemnity

Is TDM in Your Digital Licensing Agreements?

- Read terms and conditions of each license
- Is TDM permitted?
- Is TDM prohibited?
- Is the license silent on TDM?
- Is fair use/dealing permitted or prohibited?
- Can the license be interpreted to allow TDM?
- Can you ask the vendor?

Databases that Support TDM

- University of Southern California (USC), most of its database licenses do not allow TDM and they are continuing to work with vendors to include TDM in their licenses
 - Table re licenses with Gale, JSTOR, L-N, Proquest, ScienceDirect and more
 - <http://libguides.usc.edu/textmining/databases>
-

Best Practices for Negotiating Licensing Agreements with TDM

- Know your needs
- Review definitions and all clauses in your licenses
- Communicate your needs
- Negotiate
- Only sign a license that works for you
- Routinely ask for TDM rights in your licenses

Definition of TDM

***Text and Data Mining:** The process of deriving information from texts or databases, often by exploring patterns and trends across large databases of content.*

- What definition would work for you?
- Wording for a clause?

Liblicense Model Agreement 2014

- » *Text and Data Mining.* Authorized Users may use the Licensed Materials to perform and engage in text and/or data mining activities for academic research, scholarship, and other educational purposes, utilize and share the results of text and/or data mining in their scholarly work, and make the results available for use by others, so long as the purpose is not to create a product for use by third parties that would substitute for the Licensed Materials. Licensor will cooperate with Licensee and Authorized Users as reasonably necessary in making the Licensed Materials available in a manner and form most useful to the Authorized User. If Licensee or Authorized Users request the Licensor to deliver or otherwise prepare copies of the Licensed Materials for text and data mining purposes, any fees charged by Licensor shall be solely for preparing and delivering such copies on a time and materials basis.
-

Liblicense

- See <http://liblicense.crl.edu>

Discussion Points

- Should legislation address TDM?
- Should TDM be addressed in licenses?
- Industry-wide Best Practices for TDM in licensed content?
- How can we communicate awareness of the issues internally and to the public?