

THE LEGAL FRAMEWORK FOR TDM

Professor Charles Oppenheim
c.oppenheim@btinternet.com

COPYING IS OFTEN INVOLVED IN TDM

- PDFs, the *lingua franca* of academic journal articles, are not machine readable
- For TDM purposes, they must be transferred into a different digital form
- That form is often custom and specific to the research question being asked and the most appropriate tools to answer that question
- So there is a need to copy/adapt the original PDF



COPYRIGHT

- Gives the owner the right to authorise, or to refuse to authorise, any of the so-called restricted acts, including: copying; adapting; redisseminating all, or a “substantial” part, of a copyright work
- Substantial does not mean “most of”, but rather “what is important”
- If someone does such restricted acts without permission, they have infringed the right and can be sued
- However, there are exceptions to copyright, whereby someone CAN copy, etc., without having to ask for permission or pay fees
- Only two countries (UK recently being one of them) have a specific exception for TDM in their laws – more on that later
- In the absence of such an exception in a country’s national law, researchers much ask for permission from the copyright owners. Generally, the copyright owners are publishers, because authors have (foolishly) assigned their copyright to them

WHAT ABOUT DATABASES?

- Two types of protection
 - Copyright as a literary work “if, and only if, by reason of the selection or arrangement of the contents of the database, the database constitutes the author’s own intellectual creation.”
 - “database right” - see next slide
 - protects both electronic and printed collections
- A database can enjoy both database right AND copyright, or might just enjoy one of those rights, or (rarely) may not be protected at all

DATABASE RIGHT

- A database is defined as:
“a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means”
- To get database right, its creation must involve:
“substantial investment in obtaining, verifying or presenting the contents of the database” – N.B. this means many databases do NOT get database right. Collections of data which just “fall out” of a piece of research do not get the right.
- The object of protection is to:
“prevent extraction and/or re-utilisation of the whole or a substantial part, evaluated qualitatively or quantitatively, of the contents of that database.”

DATABASE RIGHT

- Protection for 15 years - renewable if database is substantially transformed
- Right is infringed if contents are extracted and re-utilised without permission or licence
- Thus TDM without permission on a database is potentially infringement, but:
 - lawful users may access and use the database
 - lawful users may extract and re-utilise *insubstantial* parts (in spite of any contractual term to the contrary)
 - lawful users may extract (but not re-utilise) *substantial* parts for the purposes of illustration for non-commercial teaching or research, provided acknowledgement is made

THE NEW UK LAW

- Came into force in June 2014
- Specific exception to copyright for TDM
- From the official guidance to the new exception issued by HMG: “Text and data mining usually requires copying of the work to be analysed. An exception to copyright exists which allows researchers to make copies of any copyright material for the purpose of computational analysis if they already have the right to read the work (that is, they have ‘lawful access’ to the work). This exception only permits the making of copies for the purpose of text and data mining for non-commercial research. Publishers and content providers will be able to apply reasonable measures to maintain their network security or stability but these measures should not prevent or unreasonably restrict researcher’s ability to text and data mine. Contract terms that stop researchers making copies to carry out text and data mining will be unenforceable.”
- Does not apply to database right. Interesting problem if a particular database enjoys both copyright and database right!
- UK researchers do not have to ask for permission, pay fees, etc., to do such TDM
- What is, or is not “non-commercial”? Not always clear! The question must be asked at the time the TDM was undertaken, so unexpected commercial benefits at the end of the project as OK, so long as at the time the intent was non-commercial
- “Lawful access” usually means licensed content, whether OA or a subscription to the materials, but also includes lawful access to printed works held in, say, a library

THE (PAST) PROBLEMS OF APPROACHING PUBLISHERS FOR LICENCES FOR TDM

- Many publishers wanted unreasonably high fees and/or placed restrictions on what could be done with their materials after TDM, and/or required researchers to use its API, and/or took an extremely long time to decide how to respond to a TDM request
- TDM researchers had to approach multiple publishers, each of whom had different attitudes, conditions, and speed of response to such requests.
- This was very costly to a researcher, and had significant impact upon the take up of TDM, as well as inhibiting academics from sharing the outputs of their TDM research
- These problems were inhibiting the take-up of TDM, thereby limiting the potential benefits this technology enables.
- Also explains why so many TDM experiments are limited to OA materials

PUBLISHER TDM LICENCE INITIATIVES GENERALLY DO NOT HELP

- Publishers have started offering their own TDM licences and policies
- Their licences often impose unfair (and in the case of the UK, unenforceable) constraints on researchers' freedom to exploit TDM, e.g., requiring users to employ publisher's API, putting unnecessary restrictions on how much can be copied, or how fast it can be copied.
- Why “unenforceable”? Because, as noted earlier, UK law specifically states that any contract or licence term that prevents anyone from doing TDM in the manner prescribed in the new exception shall be deemed null and void.
- Really need a test case on these attempted restrictions.
- Springer and Royal Society offer generous TDM provisions.
- So why are so many publishers offering restrictive licences in the UK?
Maybe they hope licensees are ignorant of the strength of the new law, or the publishers in fact don't know about it. So they are either deliberately misleading, or ignorant

EU SITUATION

- Under EU law, countries in the EU are able to introduce exceptions for non-commercial TDM research.
- However, so far only the UK has taken advantage of this. The Republic of Ireland is considering such a change to its national law.
- The EU is considering introducing a new Directive that would impose a copyright exception (and maybe also a database right exception) on Member States for TDM. The new Commissioner responsible for copyright law seems to have this as a priority, but that is no guarantee that anything will happen.
- In any case, even if it does happen, Member States will take their time to implement any such new law.
- So, far the time being, the UK has a key advantage over other Member States so we can lead TDM activities in the UK, with organisations in other countries as partners.

ELSEWHERE

- Outside of Europe, only Japan has introduced such exceptions. USA “fair use” may allow TDM, but fair use law is notoriously ambiguous and publishers are litigious.
- **There needs to be an international treaty requiring all countries to include an exception for TDM in their national laws** but that won't happen for decades, if at all
- I rest my case, m'lud

SOME USEFUL RESOURCES/ ACKNOWLEDGEMENT

- JISC report on benefits of TDM - D. McDonald and U. Kelly, *The value and benefits of text mining* (2012),
<http://www.jisc.ac.uk/reports/value-and-benefits-of-text-mining>
- Official guidance on the new UK copyright exception for TDM -
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/315014/copyright-guidance-research.pdf
- Excellent general overview of the change to UK law and its implications -
<http://copyrightuser.org/topics/text-and-data-mining/> - provides link to the precise wording in the law
- Image shown in this presentation is from Wikipedia - Creative Commons CC BY licence