The right to read is the right to mine: library resources for cross-disciplinary work

Sarah Bull, University of Birmingham
Neil Smyth, University of Nottingham
Data-Asset-Method

Data – Asset – Method: Harnessing the Infinite Archive is an international research network led by the University of Nottingham.
Corpus Protocols

Seth Cayley: Head of Research Solutions, Cengage Learning
Mike Gardner: Web Developer in Web Technologies, UoN
Kat Gupta: Corpus Protocols Researcher, UoN
Michaela Mahlberg: Professor, UoN
Neil Smyth: Librarian, UoN
Stella Wisdom: Digital Curator at the British Library
TIDAL (Times Data Archive Lab)
Mining the Meaning
UK Copyright legislation: 2014

“The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that — (a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and (b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise)”

(The Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014.)
Exceptions to copyright: Research
Gale Leads to Advance Academic Research by Offering Content for Data Mining and Textual Analysis

Grows Big Data Opportunities for the Academic Researcher

Farmington Hills, Mich., November 17, 2014 — Recognizing the benefit of data analysis for the digital humanities field, Gale, part of Cengage Learning, will make available content from its Gale Digital Collections to academic researchers for data mining and textual analysis purposes. Data mining and textual analysis – the process by which text or datasets are crawled by software that recognizes entities, relationships and action – helps researchers draw new conclusions among disparate data and is emerging as an important area of scholarly research.

“Gale is taking an important, industry-leading step by making content available for researchers in this way,” said Frank Menchaca, senior vice president for global product management at Gale. “Data mining coupled with our new curriculum alignment service, in which Gale maps an institution’s library resources to specific areas of faculty research and course focus, is helping our academic customers realize even more value for their investment with Gale.”
Data Mining the Gale Digital Collections

Frequently Asked Questions
View and listen to television and radio news programmes broadcast in the UK since May 2010

Available from twenty-two UK and international news channels, with more programmes added daily

The channels from which we currently select are:


Radio: BBC London, BBC Radio 1, BBC Radio 4, BBC 5 Live, BBC World Service, LBC, talkSport

Advanced word-searching by subtitles available for some television channels

We welcome any feedback on this service. Please contact us broadcastnews@bl.uk
The right to read is the right to mine!

The ContentMine uses machines to liberate 100 000 000 facts from the scientific literature.

http://contentmine.org/
Crossref Text and Data Mining Services

Text and data mining (TDM) is the automatic analysis and extraction of information from large numbers of documents. Researchers are increasingly interested performing text and data mining on scholarly content. This requires automated access to the full-text content of large numbers of articles. Crossref metadata helps researchers get access to this content and enables publishers to provide it.

Crossref Metadata

Crossref maintains the database of DOIs for its 4000+ publisher members. Every DOI has bibliographic metadata associated with it, describing various pieces of information about a piece of content, be that a journal article, book chapter or conference proceeding. The metadata deposited can be expanded to identify where the full text of a piece of content can be found, and this information can then be used by researchers interested in text and data mining.

http://tdmsupport.crossref.org/
Copyright Exceptions - Constraints

- Only applicable in the UK
- Researcher must have lawful access
- Must be for non-commercial purposes
- Work within publisher systems for TDM (technical protection measures)
- Collaborative research projects
- Limited quotation under ‘fair dealing’
- Sufficient acknowledgement
Technological Protection Measures

“It is important to be aware that media, such as DVDs and e-books, are often protected by Technological Protection Measures (TPMs) (also known as copy protection measures or DRM) which prevent unauthorised access or copying.

TPMs can play an important role in enabling copyright owners to offer content to consumers in different ways, as well as preventing piracy. EU and UK law protects the right of copyright owners to use TPMs to protect their works, and circumvention of such technology is illegal”

https://www.gov.uk/exceptions-to-copyright
Publisher licences

- Library signs a web licence for end user access on behalf of University
- For TDM library also has role in ensuring security of licensed content
- Use of the term “snippets” and restrictions on content quotation in research output

Community clarification ongoing based on real world projects
Infrastructure considerations

- Publisher readiness – capacity, service
- Institutional capacity – network, security
- Hard drive or API – no ‘one size fits all’

*Ensuring security of content balanced with enabling innovative research methods*
LERU Statement:
The Right to Read is the Right to Mine

Today, during a breakfast briefing at the European Parliament hosted by MEP Anaredo, the League of European Research Universities (LERU) presented what universities need from the European Union copyright reform. A timely event, after the Waterfall report on the upcoming EU copyright reform. A timely event, after the Waterfall report on the upcoming EU copyright reform. LERU calls upon the EU to ensure that research and innovation is not stifled by excessive copyright restrictions. The current fragmented and opaque EU copyright regime is clearly not a helpful environment for the realisation of the European Research Area.

EU copyright reform: what research universities need from it

As LERU has repeatedly stated, two changes to the current EU copyright regime are of utmost importance for universities:

- mandatory exception for research and education purposes;
- a mandatory exception that will enable users to test and data mine all content to which they have legal access: the right to read is the right to mine.

The current “shopping list” of exceptions and limitations in the InfoSoc Directive, from which Member States can pick and choose which to apply, is not a serious approach for a Union which Member States can pick and choose which to apply, nor is it the right approach for the development of a real European Research Area. Further, neither is it the right approach for the development of research and innovation, the current fragmented and opaque EU copyright regime is clearly not a helpful environment for the realisation of the European Research Area.

The text and data mining copyright exception: benefits and implications for UK higher education

Table of Contents

Introduction .................................................. 1
What is text and data mining (TDM)? ......................... 1
What is the TDM copyright exception? ..................... 2
What does the exception allow me to do? ................... 2
Practical implementation of TDM ......................... 4
Technical protection measures ............................ 4
Sharing the outputs of TDM ............................... 6
Personal data .............................................. 6
Quote exception ........................................... 6
Fair Dealing ................................................. 7
Resources currently in place to support TDM .............. 7
Further information/support ............................ 8
Further advice:

copyright@contacts.bham.ac.uk
copyright@nottingham.ac.uk
The right to read is the right to mine: library resources for cross-disciplinary work

Sarah Bull, University of Birmingham
Neil Smyth, University of Nottingham