openM1N7ED

Dr. Thomas Margoni
Assistant Professor of Law, University of Stirling, UK

Open Science: A legal perspective

Nostra res agitur: la scienza aperta come questione sociale. AISA,22-23 October 2015

Open Scholarship
A legal perspective

- Why TDM

- What legal barriers to TDM

- Exceptions to legal barriers

- Limits of exceptions

- Licences

- Legislative reform

- Policy choices

UNIVERSITY *of* STIRLING

openM1N7ED
Open Mining Infrastructure for Text & Data

thomas.margoni@stir.ac.uk

Why TDM

- Over 1.5 million scientific publications every year

- ca 50 million as of 2010.

- reading and analysing them is beyond human capacities

- Text and data mining (TDM) is a powerful tool for discovering value in data, by analysing structured and unstructured datasets and content and to discover concepts and entities in the world, patterns they may follow and relations they engage in, and on this basis annotate, index, classify and visualise such content.

thomas.margoni@stir.ac.uk

Why TDM

- However, *"text and data mining remains a fragmented set of tools"* and *"access is more than just being able to download something; in some cases the user (or more likely their institution) may need to pay [and negotiate] four different costs to enable the materials to be mined – traditional access (reading) costs, the right to copy, the right to digitise and then the right to text mine"*

(JISC, 2012)

thomas.margoni@stir.ac.uk

What legal barriers to unrestricted TDM

- **Copyright and rights related to copyright (e.g. Sui generis database right (SGDR))**

  - These rights usually restrict the reproduction (copy) and distribution of protected works or databases with substantial investment

  - Problem: reproduction is defined very broadly by EU law (any temporary or permanent copy of the whole or **part** of a work, etc); SGDR restricts copies of **substantial parts** and repeated copies of insubstantial parts

  - Therefore any TDM (or any other act) which requires any temporary copy of the original work or DB or part thereof infringes protected works and DB

What legal barriers to unrestricted TDM

- **Privacy/data protection**
    - Protects personal data (e.g. databases containing names, addresses, age, sex, etc).
    - One of the most important elements is the concept of **consent:** data subject can give consent for treatment of his/her data (e.g. in a DB). But such consent needs to be specific for a purpose. Consent cannot be given for any type of use (like e.g. copyright licences). Therefore, all data subjects may have to give their consent for every new use, something difficult to foresee in a OA environment
- **PSI**
    - Public Sector Information legislation is based on a different paradigm than other approaches (e.g. U.S. where works of Federal Government are not protected in the U.S.). PSI 2013 has a "open by default" approach but copyright and other similar rights and privacy are object of specific exclusion and therefore PB are under no obligation to make them accessible and/or reusable

UNIVERSITY *of* STIRLING

openM1N7ED
Open Mining Infrastructure for Text & Data

thomas.margoni@stir.ac.uk

What legal barriers to unrestricted TDM

– **Contracts/terms of use**

- Even when no rights exist on a specific BD (because there is no originality, no substantial investment, no personal data, etc) terms of use of data provider may restrict use and redistribution of DB. This limitation is based on a contractual relationship but is still an enforceable obligation (although there are differences)

UNIVERSITY *of* STIRLING

openM1N7ED
Open Mining Infrastructure for Text & Data

Exceptions to legal barriers: from "free uses" to "exceptions"

- **Copyright and rights related to copyright**
  - Exception and limitations to copyright (ELC), fair dealing, fair use. ELC are limited (in EU max 1 mandatory plus 20 at discretion of MS)
    - For TDM only possible one is exception for research and teaching. Problem is that it is not uniformly implemented in all MS and that it is usually limited to partial copies. It is also limited to non commercial activities and only for illustration for teaching and research. Fair dealing (e.g UK) is a broader standard but not as much as fair use (US).
    - Recently, UK introduced a limitation to copyright and related rights for acts of TDM for non commercial purposes and for DB legally accessed.

thomas.margoni@stir.ac.uk

UNIVERSITY of STIRLING

openM1N7ED
Open Mining Infrastructure for Text & Data

Exceptions to legal barriers: from "free uses" to "exceptions"

- **Privacy/data protection**
  - Anonymisation of data (removal of personal data) but this is time/money consuming and may reduce the usefulness of DB

- **PSI**
  - PSI legislation does not affect FoA (Freedom of Access) legislation which is MS power. But if MS empower FoA legislation then PSI "reusable by default" rule applies. However, limitation regarding copyright&C. and privacy still applies

- **Contracts/terms of use**
  - These are private agreements so there are no real exceptions. However, certain regulations (antitrust, abusive clauses, consumer protection) could under certain circumstances invalidate specific terms. This is however a case per case issue and does not seem to constitute a sound course of action.

UNIVERSITY of STIRLING

openM1N7ED
Open Mining Infrastructure for Text & Data

thomas.margoni@stir.ac.uk

Licences

- Licences are permissions/authorisations (contractually or otherwise based) that allow one or more parties to do perform certain activities.

- Licences (so called esp. in the field of copyright) may be directed to a plurality of subjects and be drafted in standard forms. These are usually called public licences (e.g. CCPL = Creative Commons Public Licence, GPL = General Public Licence, etc)

- In the field of OA, Open Content Licences (e.g. CCPL) are used to grant once and for all a permission to perform acts (copy, redistribute, modify, etc) a work of authorship or other subject matter (e.g. a DB)

- Different type of licences in the OC field. A possible problem is "licence proliferation", i.e. too many (and possibly incompatible) licences. Therefore, in the "open environment" there is a general consensus that new licences should not be created unless really necessary. One of the main goals of the Legal Interoperability WG in OpenMinTeD is to prepare a licence compatibility matrix.

thomas.margoni@stir.ac.uk

Inner limits of licences

- Licences are a powerful instrument but not perfect…

- Examples of problems with licences:

  - "Private ordering tool" i.e. can we entrust a private law tool with a function that should be a matter of public interest (wider access to knowledge)?

  - Licences are a voluntary tool, i.e. only of the owner of Work/DB is willing to grant you access, licences work. If data owner says no, there is no remedy based on contracts that can force him/her to deal with you.

  - Even if DB is willing to employ OA licences, very often there are problems of correct labeling (legal code, metadata, etc) of resources. This is a very serious issue faced in many projects in the OA field.

thomas.margoni@stir.ac.uk

UNIVERSITY *of* STIRLING

openM1N7ED
Open Mining Infrastructure for Text & Data

Legislative reform

- Legislative reform (if properly done) can address many of the limits of licences:

  - It is a "public ordering tool" meaning that the approval process of a piece of legislation goes through the standard legislative procedures with all proper guarantees

  - Legislative reform can be mandatory therefore it can apply even if data owner disagrees (e.g. UK exception cannot be limited contractually).

  - Problem of labeling is partially solved with proper legislative intervention, since all the parties that find themselves in a given situation can perform that specific act under the conditions established in the law (no need of case-by-case analysis). A possible related problem can be identified when legislation establishes specific limitation to e.g. types of use (such as the UK exception that limits its ambit to non commercial acts)

thomas.margoni@stir.ac.uk

Legislative reform

- Legislative reform (if properly done) can address many of the limits of licences:

  - Termination clauses: German and Dutch legislators amended the Copyright Act to establish that authors of scientific works regain their rights to relicense them (e.g. in OA) even if they assigned their copyright to a publisher

  - Other legislators (IT; ES) charged institutions with the task of making sure that researchers publish in OA. Unsatisfactory approach if compared with DE;NL

thomas.margoni@stir.ac.uk

Policy choices

- Every solution has advantages and disadvantages

- Through proper policy choices some of other disadvantages can be fixed.

  - Recommending 1 or a very limited no. of licences which are **compatible** (fixing problem of licence incompatibility)

  - Crucial importance that **data providers, funding agencies, scientific and public institutions** require use of correct licences and subject grant of funding to the **correct implementation** of those licences (fixing problems of "voluntarity" and "labeling")

  - **Influence public debate** so that legislative intervention in the field is appropriate (e.g. definition of right of reproduction, limited amount of ELC, need of a fair use exception, limit of non commercial exception such as in UK).

UNIVERSITY *of* STIRLING

openM1N7ED
Open Mining Infrastructure for Text & Data

thomas.margoni@stir.ac.uk

Questions?

Thank you

thomas.margoni@stir.ac.uk