The challenge of data ownership in open science

Non per Profitto: La Scienza Aperta e il ruolo dell’Università

Trento, 25 Ottobre 2018

Dr. Thomas Margoni
Senior Lecturer in Intellectual Property and Internet Law
Director of the IP LLM Programme
School of Law – CREATe
University of Glasgow
Is data owned?

Copyright theory (and sometimes copyright law) says: no
E.g.: principles, facts, data as such, etc are not protected by copyright law (Arts. 2 WCT, 9(2) TRIPs, but also generally in Berne and most legal traditions).

Other areas of law may say yes, but usually in specifically identified situations, or with limited remedies e.g.:
1) Trade secrets, confidentiality (only if secret and limited remedies)
2) Contracts (enforceability and remedies)
3) Data protection (only qualifying information and scope is protection of private life not investment)
Is data owned?

Copyright theory (and sometimes copyright law) says: no
E.g.: principles, facts, data as such, etc which are not protected by copyright law (Arts. 2 WCT, 9(2) TRIPs, but also generally in Berne and most legal traditions).

Some jurisdictions however offer some sort of protection to data: e.g. EU offers protection against substantial extractions of certain non original databases therefore effectively protecting data (but not created data).

Other jurisdictions (actually the same jurisdiction) have considered creating a new right protecting data producers (not just databases).
Is data owned and why does it matter for OS?

But if we take a closer look, EU Copyright law (doesn’t say anything explicit but) means: YES.

Interplay between Arts. 2 and 5, especially 5(1) InfoSoc requires you to obtain an authorisation for data capturing/extraction. Otherwise we would not need a e.g. TDM exception!

TDM, machine learning and most other AI training activities normally extracts principles, facts, data, correlations, etc, which copyright theory says are not protected, thus the extraction of those unprotected elements from protected works should not need an exception if copyright framework was properly designed.

Main problem with EU copyright design is that it is not properly designed: it harmonises rights broadly (reproduction, redistribution, communication to the public, etc), but does not do the same with exceptions (exhaustive but not mandatory list, narrow interpretation, etc). Main problem with this design is that you can only allow what you are aware of. This is an implicit (or perhaps quite explicit) presumption in favour of protection of investment vs. innovation.
1) **Text and Data mining**: any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations;

2) **Scope**: exception to the right of reproduction;

3) **Beneficiaries**: research organisations with lawful access for research purposes;

4) **Relationship to contracts**: Cannot be limited by contract;

5) **Relationship to technology**: Can be limited by technological measures (integrity measures and TPM)
1) **Text and Data mining**: any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations;

Comment: definition is broad enough to cover current TDM practices.
2) **Scope**: exception to the right of reproduction;

Comment: Problematic. It does not cover rights of redistribution/communication to the public and adaptation (derivative works). It means that all the times that the results of TDM are a copy in part of a protected work (Art. 2 Infosoc as interpreted by CJEU in Infopaq says that even 11 consecutive words can infringe) or when the results can be an adaption (derivative) of the original (thumbnails?) the exception is not available.
3) **Beneficiaries**: research organisations with lawful access for research purposes;

Comment: Problematic. Individuals, micro and SMEs, industry, etc cannot benefit even if acting non commercially. Purposes other than research (e.g. journalism, criticisms, review, etc) are not covered. Why? Potential contrast with fundamental rights?
4) Relationship to contracts: Cannot be limited by contract;

5) Relationship to technology: Can be limited by technological measures (integrity measures and TPM).

Comment: 4) is good. But 5) is contradictory. It creates imbalance and uncertainty with regards to the medium through which a prohibition is expressed. If “exception not available” is expressed in human/legal language (contract) this is not enforceable, but if the same condition is expressed in computer language (DRM or TPM) then it is allowed. 5) basically circumvents 4) in a way that is unreasonable, not proportionate and harmful for consumers.
Data/AI and Copyright Law: the solution

- **Now**: Implement a TDM exception not limited to research organisations for research purposes (i.e. “option 4” of the impact assessment p. 108—109).

  Comment: This will only fix some of the problems identified above, but it could be technically be done in the present draft (although it seems that none of the proposed amendments is in this direction).

- **Tomorrow**: A better drafted EU copyright law clearly marking the boundaries between protection and PD, e.g. through an open and flexible norm that will cover TDM but also future technological advancements.

  Comment: This will allow courts to readily balance investment and innovation needs without having to wait for legislative intervention. The latter has caused a major delay in EU development of TDM/ML/AI and connected technology sectors in comparison to other more innovation oriented jurisdictions (US, Canada, Singapore, Japan, etc).