IMPROVING UPTAKE OF TEXT AND DATA MINING IN THE EU

Legal Issues of TDM
The Relevant Questions

Legal issues?

Believe it or not, text and data mining (TDM) can be an unlawful activity. As a miner, you will not always identify them as such, but TDM may be restricted under copyright law, database law and data protection law.
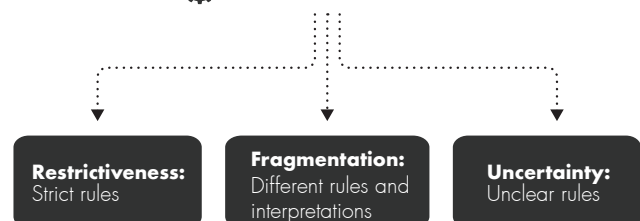
How?

TDM activities often involve the making of (temporary) copies, either on a permanent storage device or in volatile memory, i.e. computer RAM. In successful TDM cases, you would like to publish about it. When you mine protected materials, these acts may be restricted under the law. For example, mining articles, images, or sound may infringe copyrights of others, when they did not consent, or you are mining personal data (names, addresses, IP addresses, genetic data, health records) and must therefore comply with data protection law.

So when is it allowed?

Unfortunately, the law is not always straightforward and laws may differ extensively between European countries. Rules themselves may restrict TDM, but also the broad variety in their national implementations and uncertainty about the scope

HOW REGULATIONS MAY IMPEDE TDM

Restrictiveness: Strict rules

Fragmentation: Different rules and interpretations

Uncertainty: Unclear rules

of the rules may cause potential miners to refrain from using TDM technologies. See the graph on the right.
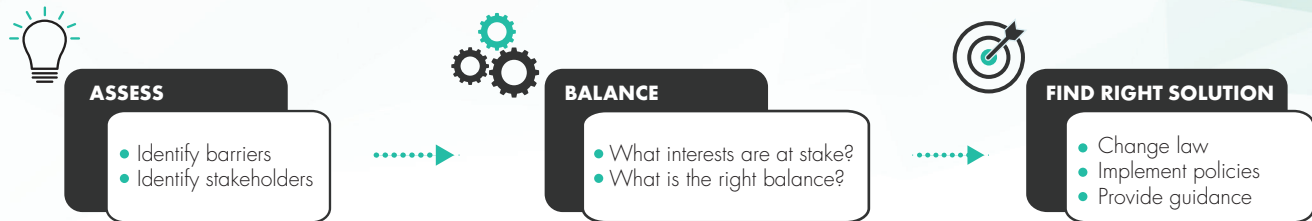
Under copyright law

When you have access to a database with copyrighted works (such as books, articles, music, videos) that you want to mine, merely having that access does not imply that you are allowed to mine. Under current copyright law, mining involves additional acts that require the rightholder's authorisation. Without it, you are restricted to only read, view or listen to the materials.

Under certain conditions, miners may rely on so-called copyright exceptions, which exempt certain uses from the control of the copyright owner. Currently, the United Kingdom has implemented a TDM exception as the first and only in the EU. It allows TDM for non-commercial research purpos-

HOW TO OVERCOME LEGAL BARRIERS?

ASSESS
- Identify barriers
- Identify stakeholders

BALANCE
- What interests are at stake?
- What is the right balance?

FIND RIGHT SOLUTION
- Change law
- Implement policies
- Provide guidance

es when the miner has lawful access to the works. Only in few other countries may TDM be carried out on copyrighted works for the same purpose under a general "research exception".

Under data protection law

Virtually any act carried out on or with personal data triggers the rights and obligations of data protection law. Considering the broad definition of "personal data", it includes any data relating to an identified or identifiable person. When you mine data that relates to individuals, ranging from (IP) addresses to medical records, you should be aware that you need to comply with the obligations and principles of data protection law, such as duties to notify or take security measures, unless you anonymise the data. However, you cannot always be certain that your data is "anonymous", since aggregation with other data sources may render the data to be personal data (again).

So shouldn't the law be changed?!

This question requires very careful and nuanced consideration. The current legal framework does not always promote or ease TDM activities. The fact that the regulations may not entirely fit with the current technological developments does not take away the fact that they are there for (legitimate) reasons and reflect a careful balance between the interest of several stakeholders and society. Under copyright law, a right balance must be found between the (economic) interests of authors and copyright holders on the one hand, and the dissemination of knowledge - such as by means of TDM - on the other hand. Under data protection law, the latter is balanced with the fundamental privacy interests of individuals.

Solutions?

As discussed, one **solution could be to change the law**. Currently, an exception for TDM under copyright law is discussed on a European level. Such an exception will make copyright law less restrictive for TDM carried out under certain circumstances. Its scope will depend on the criteria that the exception will include.

Other solutions may lie in policies: policies may promote licensing by copyright holders to permit TDM, for example by applying an Open Access policy. This may overcome the restrictive nature of copyright laws.

Self-regulation and guidelines can be useful in sectors that deal with personal data, for example, by clearly describing how miners can comply with data protection rules.

Hybrid between Wikipedia copyright symbol (public domain) and EU flag (ineligible for copyright)

Discover more
VISIT OUR COLLECTION

STORIES   PROJECTS   ORGANISATIONS   TOOLS   CHALLENGES