Copyright obstacles in data mining

18.09.2019

Key takeaways

- **Data mining simplifies the structuring and analysis of large amounts of data.**
- **Without the author's consent, data mining of copyright-protected works is only permitted under current Swiss law for private or (under certain circumstances) internal use by juridical persons.**
- **According to revised Swiss law, data mining shall be permitted for copyright-protected works if it is carried out for scientific research purposes.**

Introduction

“Data is the new oil, information the new gold.”

The volume of available, high dimensional data and information has increased rapidly in recent years. Whether in industry, medicine or research, there is no doubt that data is rising in value – and growing in significance.

But not all that glitters is gold. Given the immense data volumes, which can be unstructured, there are significant challenges in terms of managing, recording, storing and analysing individual data points. While potentially meaningful data was selected manually in the past, the sheer volume of available data is a barrier to this approach in most cases today.

But data mining can help. This method of intelligent data analysis and interpretation uses computer-assisted technology and algorithms to structure existing data and analyse it to “mine” new knowledge and insights. Data mining can also use artificial intelligence to make future projections. Data mining algorithms can formulate hypotheses about possible correlations, patterns or trends, and simultaneously assess their probability.

Pestalozzi Rechtsanwälte AG
Feldeggstrasse 4
CH-8008 Zürich
T +41 44 217 91 11
zrh@pestalozzilaw.com

Pestalozzi Avocats SA
Cours de Rive 13
CH-1204 Genève
T +41 22 999 96 00
gva@pestalozzilaw.com

pestalozzilaw.com

Even though data mining is already widely used and appreciated for its capabilities, it is important not to lose sight of the legal risks associated with automated data analysis and works protected by copyright law.

This legal update provides an overview of data mining, the copyright hotspots and the legislative measures under way in response to current uncertainty.

What is data mining?

Data mining refers to a process of automated, software-based examination, analysis and structuring of big data with the aim of identifying new patterns and connections within existing data sets. Data mining seeks to extract previously unknown information from this data. Its primary aim, then, is to deliver new insights and knowledge.

Data mining works with data of any kind (e.g. videos, text or images). If the data set relates exclusively to unstructured or weakly structured text data, the general term used is ‘text mining’.

How does data mining work?

Although there are many technical realisation models for data mining, the process essentially follows the two main steps described in the following.

The first step involves preparing the foundation for any data mining process: the data set. The data set is created by scanning the available digital data and extracting the relevant data points based on defined selection criteria. The tools used for this process are often referred to as crawling and scraping software, which downloads and continuously and permanently stores the relevant data. If necessary, the data set is then transformed into a format that the data mining software can process.

The second step consists of the data mining itself; the data mining software is applied to the prepared data set. The application holds the data from the prepared data set temporarily in the random-access memory while it searches for relevant information. The data is subsequently overwritten with new data and thus erased.

What is the copyright issue with data mining?

The data mining processes requires data to be stored permanently (for preparation of the data set) and temporarily (for analysis of the data set). In technical terms, this means that data has to be copied or replicated, which can lead to copyright law issues when using data copied from copyrighted works in data mining processes.

According to the legal definition, works enjoy copyright protection if they are literary and artistic intellectual creations with an individual character (Art. 2 (1) Federal Act on Copyright and Related Rights, CopA). This definition is rarely met for raw data, which can thus hardly be protected by copyright. In contrast, texts with literary, scientific and/or technical content, as well as images and audio-visual works are often protected by copyright. The same is true of data collections if the selection and structure of the data have an individual character.

In relation to such works, the author has the exclusive right to decide whether, when and how his work is used (Art. 10 (1) CopA). This exclusive right also includes the author's right to produce copies of the work (Art. 10 (2) lit. a CopA; so-called right to copy) and, with it, the right to enable perception of the work by human senses independent of the original.

As it of no consequence whether the copy is a permanent or temporary one, in the data mining process both the storage of data for the data set and for the analysis are exclusive rights reserved for the author of the work.

Unauthorised reproduction of a copyrighted work by a third party during the data mining process fundamentally constitutes a breach of copyright, which may result in consequences under both civil and criminal law.

Does data mining breach copyright law per se?

The author's right under copyright law to prohibit copying as described above is not absolute. Swiss copyright law contains provisos limiting the author’s rights to exclusivity for works already published if certain criteria are met (Art. 19 et seq. CopA).

In connection with data mining, where the data set to be analysed contains copyrighted works, it is important to first consider the proviso applicable when using the work for personal use and when making a temporary copy. The applicability of provisos needs to be assessed separately for each of the two main data mining process steps (cf. section 3).

Proviso for private use

The law allows published works to be used privately without the consent of the copyright holder. The definition of private use includes any personal use of a work or use within a circle of persons closely connected to each other (Art. 19 (1) lit. a CopA).

Since any use of a work is included in the scope of private use, temporary or permanent copies of a copyrighted work may be made. Natural persons are permitted by law to carry out automated data analysis involving preparation and analysis of a data set in a data mining process exclusively for private purposes. However, the private use proviso does not cover publication of the data mining results, which is only permissible with the permission of the author.

Juridical persons are permitted to use the work privately for internal purposes (Art. 19 (1) lit. c CopA). Consequently, published copyrighted works – with the exception of the (largely) full copying of commercially available works – can be copied, distributed and made accessible for internal information or documentation purposes in enterprises, public administrations, institutions, commissions and similar bodies.

As data mining does not exclusively serve the purpose of internal information (as its specific purpose), it is questionable whether and to what extent the bodies named above would be able to invoke the proviso for internal use. It should also be noted that copying (in order to prepare the data set) for internal purposes is subject to a remuneration obligation towards the author (Art. 20 (2) CopA).

Proviso for temporary copying

It is permissible to make temporary copies of a work if the copies are (i) transient or incidental, (ii) they represent an integral and essential part of a technological process, (iii) their sole purpose is to enable a transmission of the work in a network between third parties by an intermediary or a lawful use of the work, and (iv) they have no independent economic significance (Art 24a CopA).

A temporary copy is generally considered to be such if its life span is limited to the period of time required to enable accurate functioning of the technological process. Storing works protected by copyright for the purpose of creating a data set does not meet this definition of temporary because the works are permanently stored and not deleted upon completion of the data analysis. Thus, the proviso for temporary copying cannot be applied for the creation of a data set.

However, the criteria for the proviso under Art 24a CopA are fulfilled for the data analysis (step 2) itself. This step is exempted from the right under copyright law to prohibit copying and is thus permissible.

Conclusions

In summary, under current Swiss law, data mining of copyright-protected works without consent by the author is only permissible for private or (in certain circumstances) internal use by juridical persons.

If the automated data analysis of copyright-protected works is not intended solely for private or, for juridical persons, internal use, it is not permissible without the permission of the copyright holder because the creation of a data set as the basis for the data mining process is not covered by any proviso under copyright law.

Obtaining the required permission frequently proves a difficult, costly and time-consuming task in practice. It is often unclear whether, and if so where, copyright-protected works are located within the data set to be analysed, and who owns the rights. Under current law, there is therefore always an immanent risk of performing data mining without the required permission of any copyright holders and thus in breach of the law.

What changes will the revised Copyright Act introduce?

Given the major significance of research for Switzerland and the legal uncertainty associated with data mining in the context of research activity, the revised Copyright Act will introduce a new proviso for scientific applications.

Accordingly, in future it shall be permissible to copy a work for scientific research purposes if copying is a necessary part of the technological process and the works to be copied can be accessed legally (Art. 24d (a) draft CopA). In addition, it shall also be permissible to store the copy made in the process described above for archiving or backup purposes once the scientific research activity has been completed (Art 24d (2) draft CopA). The so-called science proviso does not apply to copying of computer programs/software (Art. 24d (3) draft CopA).

For the science proviso to apply, the (primary) purpose needs to be scientific research, which is described as the systematic search for new findings within one or more scientific disciplines. Basic research and applied research are both included. In contrast to EU law, the new proviso shall also cover scientific research for commercial purposes as restriction to non-commercial research only would lead to irresolvable definition issues.

Furthermore, the science proviso shall apply for both temporary and permanent copies (including entire works), provided these copies are required for the application of a technological process (e.g. algorithms) for research purposes. Finally, it shall be permissible to distribute the findings of the research, provided they can be considered separate from the original works.

Since the science proviso only applies if the copyrighted works were accessed legally (e.g. assumed to be the case if the works were acquired legally or are freely accessible on the internet), the new proviso does not require explicit consent from the author to make a technical copy of the work. The author does not enjoy a reserved right of use like the one provided by EU law. Furthermore, no remuneration is owed; in contrast to the proviso for internal use by juridical persons, the science proviso does not provide for remuneration.

Outlook

The new science proviso responds to the currently inadequate situation that data mining of copyrighted works for research purposes requires all the necessary permissions to be obtained in the form of licenses, and remuneration to be paid.

Although the science proviso addresses existing uncertainty and inconvenience in relation to data mining for research purposes, it does not respond to other questions around automated data analysis beyond the field of research.

For applications other than private or internal use, therefore, it is still necessary to obtain all required consent and pay the corresponding remuneration if data mining is performed for a reason other than research.

It is also important to note the time line: the revision of the Copyright Act remains in the process of parliamentary consultation; the new regulations can thus be expected to enter into force in 2020 at the earliest.

Contributing authors: Lorenza Ferrari (Partner), Nando Lappert (Associate), Luca Bossard (Junior Associate)

Michèle Burnier



Partner
Attorney at law

Pestalozzi Attorneys at Law Ltd
Cours de Rive 13
1204 Geneva
Switzerland
T +41 22 999 96 00
michele.burnier@pestalozzilaw.com