

Damian Hartmann

Text and Data Mining and Copyright in Switzerland and the European Union

Die Schweiz und die Europäische Union haben neue Urheberrechtsschranken geschaffen, um den Einsatz von Text und Data Mining vor urheberrechtlichen Hindernissen zu bewahren. In diesem Beitrag werden die neuen urheberrechtlichen Schrankenbestimmungen beleuchtet sowie die Situation in der Schweiz mit der Situation in der Europäischen Union verglichen.

La Suisse et l'Union européenne ont imposé de nouvelles restrictions au droit d'auteur afin de préserver l'utilisation du Text and Data Mining des obstacles liés au droit d'auteur. Le présent article examine ces nouvelles restrictions et compare la situation en Suisse à celle de l'Union européenne.

I. Introduction

1. Problem background
2. Structure

II. Text and data mining

1. Definitions
2. How it works
3. Connections to copyright

III. Legal situation in Switzerland

1. Copyright protection
2. Relevant copyright barriers for text and data mining

IV. Legal situation in the European Union

1. Copyright protection
2. Relevant copyright barriers for text and data mining

V. Controversies and comparison of text and data mining provisions

1. Controversies in Switzerland
2. Controversies in the European Union
3. Comparison

VI. Conclusion

I. Introduction

«Text and Data Mining (TDM) serves as an essential tool to navigate the endless sea of online information, in search of this invaluable treasure that big data might hold for the European economy.»¹

1. Problem background

Text and data mining (TDM) is a highly topical subject. Worldwide, researchers in the fields of medicine, biology, pharmaceuticals, anthropology and criminology, as well as companies in banking, forestry, fashion and marketing² are exploiting the potential of diverse TDM applications to discover and profit from new knowledge from the immense

amount of text and data in our data-driven world. This potential and the associated high value of TDM is illustrated by the opening quote.

One of the biggest challenges that TDM users face is complying with copyright in the text, data and databases used, which may be affected by TDM processes. Due to legal uncertainties in the copyright field, various regulators have decided to include a new exception provision in copyright law. In the European Union, this was done with the enactment of the Digital Single Market Directive (Directive 2019/790/EU)³ and in Switzerland with the implementation of a new provision into the national Copyright Act (CopA). Both the EU and Swiss regulations are still recent, so – as far as can be seen – the courts have not yet had the opportunity to comment on these new provisions.

The aim of this article is to discuss and evaluate the new TDM provisions in Switzerland and the EU in a comparative legal manner, after providing an overview of the copyright issues associated with TDM applications in Switzerland and the EU.

Due to various points of contact, TDM can also be classified in larger thematic complexes, for example, in the areas of artificial intelligence, machine-learning, big data or (open) innovation. These thematic complexes are not addressed here since this article focuses specifically on TDM applications. From a legal point of view, TDM is relevant for different areas of law such as copyright law, data protection law, contract law and unfair competition law. However, this article is limited to the treatment of issues arising from copyright law.

DAMIAN HARTMANN, M.A. HSG in Law and Economics, Trainee Lawyer, St. Gallen.

- 1 C. GEIGER/O. BULAYENKO/G. FROSIO, Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data? Legal Analysis and Policy Recommendations, *International Review of Intellectual Property and Competition Law* 49/2018, 816.
- 2 E. ROSATI, Copyright as an obstacle or an enabler? A European perspective on text and data mining and its role in the development of AI creativity, *Asia Pacific Law Review* 27/2019, 201 f.
- 3 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (OJ L 130, 17.5.2019, p. 92–125).

2. Structure

This article is structured as follows: It starts with the topic of text and data mining (II). Various definitions of TDM are presented, followed by a description of how TDM works. The three steps in which TDM usually takes place (access to content, extraction and copying of data, and pre-processing and discovery of hidden knowledge) are shown in particular. In addition, the points of connection relevant to copyright law are presented. In a subsequent step, the legal situation in Switzerland is examined (III). Here, copyright protection of text and databases is presented and the Swiss copyright barriers relevant for TDM are highlighted. After that, the paper deals with the legal situation in the European Union (IV). Again, copyright protection in general and for databases in particular is shown, followed by a presentation of the European copyright limitations and exceptions relevant for TDM. Finally, the controversies of the Swiss and the European TDM provisions are discussed, and a comparison is drawn between the Swiss TDM provision and the two European TDM provisions (V). The focus lies hereby on Art. 24d CopA for Switzerland and on Arts. 3 and 4 of Directive 2019/790/EU for the European Union.

II. Text and data mining

1. Definitions

Data mining originates from computer science, where it became established in the late 1980s. From a computer science perspective, data mining is defined as «a set of mechanisms and techniques, realized in software, to extract hidden information from data».⁴ Text mining is a type of data mining in which the data mining activity is applied to text.⁵ Data mining is based on statistics and machine learning technology. Therefore, data mining is not a separate technology, but rather an application of machine learning and statistics.⁶

Somewhat less technical than the definition from a computer science perspective is that of the Swiss Federal Council. According to the Federal Council, TDM is a research tool for the electronic evaluation of large amounts of text and data, whereby cross-references in the large amounts of text and data can be quickly identified by using various statistical and mathematical methods.⁷ A legally binding definition of TDM does not exist in Swiss copyright law.

A similar description of TDM forms the basis of the Digital Single Market Directive. According to Recital 8 of the aforementioned directive, TDM comprises new processes with which information available in digital form, such as text and data, can be automatically evaluated using computers. In this way, large amounts of information can be processed and findings can be obtained.⁸ According to the binding definition of the EU, TDM means any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations.⁹

2. How it works

A broad spectrum of different TDM techniques and methods exists.¹⁰ Accordingly, there are different ways to represent the different steps of TDM.¹¹ For the purposes pursued here, it is convenient to divide the TDM process as follows: (a) accessing data and text; (b) extracting and copying data and text; and (c) discovering knowledge.¹² These three steps are briefly outlined below.

a) Accessing text and data

In a first step, access to text and data is needed. Text and data can derive from different sources. One possibility is to use «crawling» to automatically search websites and create individual data collections.¹³ It is possible to use unstructured data from public websites or social media, but structured data from databases of professional database operators also come into consideration.¹⁴ In addition to such individually compiled text and data collections, existing databases can also be used for TDM.¹⁵ The core activity of TDM takes place only in the following second and third steps. Nevertheless, access to text and data is a prerequisite for TDM to operate at all.

b) Extracting and copying of text and data

As soon as a data set becomes available, the actual TDM process begins. In order for the collected data or the database to be analyzed, the data in their entirety or at least portions thereof must be copied or extracted and stored permanently or temporarily, depending on the specific TDM tech-

4 On the whole F. COENEN, Data mining: past, present and future, *The Knowledge Engineering Review* 26/2011, 25.

5 COENEN (Fn. 4), 28; S. BRÄNDLI, Data Mining als Forschungsmethode: Die Probleme des Grabens nach Datengold, in: S. Brändli/R. Haragama/R. Schister/A. Tamò (Hg.), *Mensch und Maschine – Symbiose oder Parasitismus?*, Bern 2015, 44; I. A. STAMATOUDI, Text and Data Mining, in: I. A. Stamatoudi (Ed.), *New developments in EU and international copyright law*, Alphen on the Rhine 2016, 254; M. TRUYENS/P. VAN EECHE, Legal aspects of text mining, *Computer Law and Security Review* 30/2014, 153 f. According to BRÄNDLI (Fn. 5), 44 and COENEN (Fn. 4), 28 f. data obtained from images, videos or graphs can also be used for data mining, which is then referred to as «image mining» or «graph mining» in the sense of further subcategories of data mining.

6 COENEN (Fn. 4), 26.

7 BBl 2018, 690.

8 On the whole Recital 8 of Directive 2019/790/EU.

9 Art. 2(2) of Directive 2019/790/EU.

10 STAMATOUDI (Fn. 5), 260.

11 See EP, PE 604.941, 5, STAMATOUDI (Fn. 5), 260 and G. SPINDLER, Text und Data Mining – urheber- und datenschutzrechtliche Fragen, *Gewerblicher Rechtsschutz und Urheberrecht*, 2016, 1112 for a four-step description of TDM and BRÄNDLI (Fn. 5), 44 f. for a different three-step description of TDM.

12 This three-step description of the TDM process is introduced by ROSATI (Fn. 2), 204.

13 TRUYENS/VAN EECHE (Fn. 5), 165; BRÄNDLI (Fn. 5), 45.

14 TRUYENS/VAN EECHE (Fn. 5), 154 f. See also BBl 2018, 627.

15 GEIGER/BULAYENKO/FROSIO (Fn. 1), 817.

nique.¹⁶ The extracted or copied collection of data and text used for a TDM analysis is called a «corpus».¹⁷ There are TDM applications in which neither data nor words are copied or extracted, and TDM applications where only a marginal part, e.g. two to three words, are copied. However, such TDM applications tend to be rare.¹⁸ Usually, a TDM application relies on some type of copying.¹⁹

c) Pre-processing and discovering hidden knowledge

Depending on what kind of data has been compiled and what goal is being pursued through the TDM analysis, the data must be prepared for the analysis. For instance, it may be necessary to first structure the data, delete superfluous data, or correct erroneous data (so-called «data clean-up»). Furthermore, depending on the TDM technique used, text must be transformed into a format suitable for computers (so-called «normalization»). These actions represent the preparatory actions for the subsequent analysis.²⁰

Once these steps have been completed, the data are ready for analysis. Depending on the TDM application, different analysis results are created. Typical results of a TDM analysis are pattern findings, clustering and classification or summaries.²¹ However, TDM not only allows the identification of previously unknown relationships and patterns, but also statements about future developments.²²

3. Connections to copyright

From a copyright point of view, two central sets of questions arise with regard to TDM, which concern the object of protection on the one hand and the content of copyright on the other hand.²³

Regarding the object of protection, data and text used for TDM must first be qualified under copyright law. TDM is only relevant with respect to copyright law if the underlying data and text enjoy copyright protection at all. Since there is no internationally harmonized world copyright law,²⁴ the copyright protection of data and text concretely used for TDM is basically determined by the applicable national copyright law. In determining the material protected by copyright, the question of the copyright protection of databases is of particular interest. Of importance hereby are the requirements of the TRIPS Agreement regarding copyright protection of data collections,²⁵ since both Switzerland and all EU Member States are parties to the TRIPS Agreement.²⁶

If copyright-protected data and text are used during TDM, the question that arises in the area of the content of copyright is to what extent TDM interferes with the rights of the copyright holder. The focus here is on the temporary and permanent copies that are regularly produced during the TDM process.²⁷

In the following two sections,²⁸ these two sets of questions are examined for Switzerland and the European Union.

III. Legal situation in Switzerland

1. Copyright protection

a) Text

The Swiss Copyright Act protects all literary and artistic works that are intellectual creations and have an individual character.²⁹ Art. 2 CopA contains a non-exhaustive list of categories of works,³⁰ which includes in particular (scientific) linguistic works.³¹ A linguistic work exists if a conceptual content is expressed by means of language. According to jurisprudence, the requirements concerning the criterion of individual character are low for linguistic works.³² In the case of scientific linguistic works, the individual character may

16 R. DUCATO/A. STROWEL, Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to «Machine Legibility», *International Review of Intellectual Property and Competition Law* 50/2019, 658.

17 SPINDLER (Fn. 11), 1112; TRUYENS/VAN EECHE (Fn. 5), 154. See also DUCATO/STROWEL (Fn. 16), 667 f.

18 On the whole STAMATOUDI (Fn. 5), 261 f. See also ROSATI (Fn. 2), 206.

19 EP, PE 604.941, 6.

20 On the whole ROSATI (Fn. 2), 209. See also STAMATOUDI (Fn. 5), 260, according to who the preparatory transformation of the extracted data on the one hand and the copying of the data (loading of the data) on the other can be temporally interchanged. If the transformation takes place first, the data is temporarily stored for this purpose and only permanently stored (loaded) after the transformation. This process is called ETL (extract, transform and load). If the data is first stored (loaded) and then transformed, this is referred to as ELT (extract, load and transform). The respective process is determined by the data to be analyzed.

21 COENEN (Fn. 4), 26 ff.

22 ROSATI (Fn. 2), 209. See also TRUYENS/VAN EECHE (Fn. 5), 153, who mention text clustering, concept extraction, sentiment analysis and entity relation modelling as possible applications of text mining.

23 S. BRÄNDLI, Die Flexibilität urheberrechtlicher Schrankensysteme. Eine rechtsvergleichende Untersuchung am Beispiel digitaler Herausforderungen, Bern 2017, N 551.

24 R. M. HILTY, Urheberrecht, 2. Aufl., Bern 2020, N 107.

25 According to Art. 10(2) TRIPS Agreement «compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself».

26 J. A. L. STERLING/T. COOK, European Union, in: J. A. L. Sterling (Ed.), *Sterling on world copyright law*, 5th ed., London 2018, para. 28.18.

27 BRÄNDLI (Fn. 23), N 551.

28 See below III for the legal situation in Switzerland and IV for the legal situation in the European Union.

29 R. VON BÜREN/L. DAVID, *SIWR II/1*, 3. Aufl., Basel 2014, N 38. See Art. 2(1) CopA according to which works are literary and artistic intellectual creations with individual character, irrespective of their value or purpose.

30 HILTY (Fn. 24), N 173; VON BÜREN/DAVID (Fn. 29), N 196; E. MARBACH/P. DUCREY/G. WILD, *Immaterialgüter- und Wettbewerbsrecht*, 4. Aufl., Bern 2017, N 255 f.

31 See Art. 2(2) CopA according to which the following categories of works can be distinguished: literary, scientific and other linguistic works; musical works and other acoustic works; works of art, in particular paintings, sculptures and graphic works; works with scientific or technical content such as drawings, plans, maps or three-dimensional representations; works of architecture; works of applied art; photographic, cinematographic and other visual or audio-visual works; choreographic works and works of mime.

32 On the whole HILTY (Fn. 24), N 175 f. See also VON BÜREN/DAVID (Fn. 29), N 202.

result not only from the text, but also from the selection, division and arrangement of the scientific material.³³ Single words are usually part of common speech, which is why they lack individuality and are accordingly not protected by copyright, except in exceptional cases.³⁴ This also applies if they are short word sequences or a title of a copyrighted work.³⁵ Generally, text which is used as a basis for text mining is thus in any case suitable to be protected by copyright.

b) Databases

The copyright protection of a database itself (not of the data contained therein) is governed by the regulation on collected works.³⁶ A more extensive *sui generis* protection right for databases, as exists in the EU, does not exist in Swiss copyright law.³⁷ According to the regulation on collected works, collections are protected in their own right, provided that they are intellectual creations with an individual character with regard to selection or arrangement.³⁸ The protection of works included in a collected work is reserved.³⁹ This means that databases can obtain copyright protection via the provision of collected works, regardless of the copyright protection of the underlying data.⁴⁰

Accordingly, databases which are pure compilations of freely available data, such as a database with the addresses of all doctors in a certain region, are not protectable because neither the selection nor the arrangement has an individual character.⁴¹ In general, such databases, which are designed to be complete, are not protectable.⁴² On the other hand, protection is affirmative for databases that compile statistically significant groups of persons for representative market surveys or clinical trials.⁴³

Compared to the requirements for data compilations under the TRIPS Agreement,⁴⁴ the Swiss regulation on collected works takes the same approach. To the extent that databases are used for analysis in TDM, this means that not only the individual works may be protected by copyright, but additional and independent copyright protection may exist at the level of the database itself.

c) Reproduction

The author has the exclusive right to decide whether, when and how his or her work is used.⁴⁵ The author has the right, in particular to produce copies of the work.⁴⁶ Although the CopA, which is still somewhat antiquated, assumes that copyrighted works are physically copied by using the term «produce copies», the reproduction as a copyright-relevant action must also be observed in the digital world.⁴⁷ Thus, both the permanent and the temporary electronic saving of data on a computer constitutes a reproduction within the meaning of Art. 10(2)(a) CopA, which is generally subject to the consent of the copyright holder.⁴⁸

Since most TDM applications make copies, as mentioned above,⁴⁹ TDM would be dependent on the consent of all copyright holders concerned.

2. Relevant copyright barriers for text and data mining

a) Temporary copies

Due to the far-reaching term of «reproduction»,⁵⁰ the internet would not be useable in a meaningful way if certain limits were not placed on copyright.⁵¹ One of these limits is Art. 24a CopA, which applies to temporary copies. According to this provision, the making of temporary copies of a work is permitted if: (i) they are 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. These conditions must be met cumulatively for the exception of Art. 24a CopA to apply.⁵²

Art. 24a CopA predominantly addresses processes such as browsing and caching,⁵³ but is also relevant in the context of TDM. If temporary copies are created during TDM – which is the case during the actual analysis process – they are covered by Art. 24a CopA; however, the aforementioned provision does not permit the creation of permanent copies,

33 BGE 113 II 306 ff. E. 3a. See also HILTY (Fn. 24), N 179; VON BÜREN/DAVID (Fn. 29), N 255.

34 VON BÜREN/DAVID (Fn. 29), N 203 f.

35 VON BÜREN/DAVID (Fn. 29), N 206 f.; HILTY (Fn. 24), N 236 ff. See Art. 2(4) CopA according to which drafts, titles and parts of works, insofar as they are intellectual creations with an individual character, are also protected.

36 HILTY (Fn. 24), N 249; VON BÜREN/DAVID (Fn. 29), N 256 and 403; MARBACH/DUCREY/WILD (Fn. 30), N 285.

37 HILTY (Fn. 24), N 249; S. BRÜHWILER, *Der Beizug Dritter zu gesetzlich erlaubten Werknutzungen*, Zürich 2021, N 153. See below IV.1.b for the copyright protection of databases in the EU and the *sui generis* protection right.

38 See Art. 4(1) CopA.

39 See Art. 4(2) CopA.

40 VON BÜREN/DAVID (Fn. 29), N 256. According to VON BÜREN/DAVID (Fn. 29), N 403 and HILTY (Fn. 24), N 247 f. typical examples of collected works include newspapers and magazines, yearbooks, complete editions, radio and television broadcasts, encyclopaedias or course materials.

41 VON BÜREN/DAVID (Fn. 29), N 256.

42 HILTY (Fn. 24), N 249.

43 VON BÜREN/DAVID (Fn. 29), N 403.

44 See Fn. 25.

45 Art. 10(1) CopA.

46 Art. 10(2)(a) CopA.

47 HILTY (Fn. 24), N 301. See also MARBACH/DUCREY/WILD (Fn. 30), N 312 according to who the technical nature of the reproduction is meaningless and therefore saving, uploading to and downloading from the internet constitutes a reproduction.

48 HILTY (Fn. 24), N 303; VON BÜREN/DAVID (Fn. 29), N 256.

49 See above II.2.b.

50 See above III.1.c.

51 W. EGLOFF, in: D. Barrelet/W. Egloff (Hg.), *Das neue Urheberrecht. Kommentar zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte*, 4. Aufl., Bern 2020, URG 24a N1; HILTY (Fn. 24), N 472.

52 EGLOFF (Fn. 51), URG 24a N 3; VON BÜREN/DAVID (Fn. 29), N 867.

53 EGLOFF (Fn. 51), URG 24a N 4; HILTY (Fn. 24), N 471. See also MARBACH/DUCREY/WILD (Fn. 30), N 367.

which is why the provision does not cover all mechanisms of reproduction in the various forms of TDM applications.⁵⁴

b) Private personal use

According to Art. 19(1)(a) CopA, published works may be used for private use, whereby private use means any personal use of a work or use within a circle of persons closely connected to each other, such as relatives or friends. Private personal use does not differentiate concerning the lawfulness of access to the works, so that works made accessible unlawfully can also be used for private personal use as long as they have been published.⁵⁵ Moreover, private personal use covers all types of use of the work.⁵⁶

With regard to TDM, under Art. 19(1)(a) CopA, a private researcher can thus create both a permanent data set and permanent and temporary copies that are created during the TDM process.⁵⁷ This means that if TDM is conducted as private research in the sense of Art. 19(1)(a) CopA, it is completely possible to do so without violating copyright law.⁵⁸ It should be noted, however, that private researchers must perform all actions themselves. If the work is outsourced to a third party, there is already a restriction in that the third party may not make a complete or extensive reproduction of commercially available works.⁵⁹ However, TDM often takes place with the help of a third party and is regularly not limited to a private group of persons within the meaning of Art. 19(1)(a) CopA.⁶⁰ Furthermore, the exception in Art. 19(1)(a) CopA is not sufficient to justify commercial TDM⁶¹ and it is exclusively applicable to natural persons.⁶² Thus, in practice, many TDM activities often do not fall under the Art. 19(1)(a) CopA exception.

c) Use of works for the purpose of scientific research

It has been shown that the existing barriers of copyright law do not completely cover various TDM applications, so that there is a residual risk that copyright infringements are committed by means of TDM. TDM has great significance in practice and is indispensable in a modern, data-driven economy.⁶³ Furthermore, since research, which is an area where TDM frequently occurs, is of great importance for Switzerland, the existing uncertainties are to be eliminated by a new regulation in the CopA – the so-called *science barrier*.⁶⁴ To accomplish this, on 1 April 2020, a new provision entered into force with Art. 24d CopA.

According to Art. 24d CopA it is permissible for the purposes of scientific research to reproduce a work if the copying is due to the use of a technical process and if the works to be copied can be lawfully accessed.⁶⁵ Upon conclusion of the scientific research, the copies made in accordance with Art. 24d CopA may be retained for archiving and backup purposes.⁶⁶ The new exception in Art. 24d CopA supplements the already existing copyright barriers, which is why it applies cumulatively to the latter.⁶⁷ The individual elements of the *science barrier* will be explained in the following paragraphs.

First, it should be noted that Art. 24d CopA does not permit all uses of a work within the meaning of Art. 10 CopA, but only reproduction. However, the right of reproduction is comprehensive, so that a complete work may also be reproduced.⁶⁸ The reproductions may be stored for archiving and backup purposes and thus do not have to be destroyed. This storage option is useful because it allows research results to be reviewed and the research project to be repeated.⁶⁹

The central essence of Art. 24d CopA lies in the purpose limitation to scientific research. The reproduction of copyrighted works is only permitted if it is done for the purpose of scientific research.⁷⁰ According to the Federal Council, scientific research is understood to be the systematic search for new knowledge within and across different scientific disciplines. It includes both basic and applied research. In the case of a research project that serves not only scientific but also other purposes, scientific research must remain the main purpose of the project for Art. 24d CopA to apply.⁷¹ No distinction is made between commercial and non-commercial research because the distinction would be too difficult to make. Therefore, Art. 24d CopA is intended to allow the reproduction of a work for the purpose of scientific re-

54 M. ISLER, Wissenschaftsschranke (Art. 24d URG), in: P. Mosimann (Hg.), Das revidierte Urheberrecht. Die wesentlichen Neuerungen – eine Standortbestimmung, Basel 2020, N 228; BRÜHWILER (Fn. 37), N 159; BRÄNDLI (Fn. 23), N 562. See also BBl 2018, 628.

55 BGE 145 III 72 ff. E. 2.2.2 (the Swiss leading case for streaming on piracy platforms); HILTY (Fn. 24), N 456; D. BARRELET/W. EGLOFF, in: D. Barrelet/W. Egloff (Hg.), Das neue Urheberrecht. Kommentar zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, 4. Aufl., Bern 2020, URG 19 N 10; VON BÜREN/DAVID (Fn. 29), N 770 f.

56 HILTY (Fn. 24), N 455; VON BÜREN/DAVID (Fn. 29), N 765; MARBACH/DUCREY/WILD (Fn. 30), N 349.

57 BRÄNDLI (Fn. 23), N 554 f.

58 ISLER (Fn. 54), N 234.

59 BRÄNDLI (Fn. 23), N 556. See Art. 19(3)(a) CopA, according to which the complete or substantial copying of a work obtainable commercially is not permitted outside the private sphere defined in Art. 19(1)(a) CopA.

60 BRÜHWILER (Fn. 37), N 262.

61 BRÄNDLI (Fn. 23), N 559.

62 BARRELET/EGLOFF (Fn. 55), URG 19 N 11; VON BÜREN/DAVID (Fn. 29), N 772; MARBACH/DUCREY/WILD (Fn. 30), N 350. See also HILTY (Fn. 24), N 454.

63 HILTY (Fn. 24), N 513.

64 BBl 2018, 627.

65 Art. 24d(1) CopA.

66 Art. 24d(2) CopA. See also Art. 24d(3) CopA, according to which Art. 24d CopA does not apply to the copying of computer programs.

67 BBl 2018, 629 f.; W. EGLOFF, in: D. Barrelet/W. Egloff (Hg.), Das neue Urheberrecht. Kommentar zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, 4. Aufl., Bern 2020, URG 24d N 3; BRÜHWILER (Fn. 37), N 266; ISLER (Fn. 54), N 234.

68 On the whole EGLOFF (Fn. 67), URG 24d N 4.

69 On the whole BBl 2018, 629; EGLOFF (Fn. 67), URG 24d N 8; ISLER (Fn. 54), N 225.

70 EGLOFF (Fn. 67), URG 24d N 5.

71 On the whole BBl 2018, 628; ISLER (Fn. 54), N 223.

search, regardless of who is conducting the research and how that research is funded.⁷²

Furthermore, the reproduction must be the result of a technical process. Although this technology-neutral formulation has been chosen, it is clear that Art. 24d CopA is primarily directed at TDM processes.⁷³

Finally, there must be lawful access to the copyrighted works. This is to be understood – analogous to the copyright exception for private personal use – as meaning that the user of the work has lawfully obtained access to the works, but not that the copy of the work used was lawfully created.⁷⁴

IV. Legal situation in the European Union

1. Copyright protection

a) General

Principally, each EU Member State has its own copyright laws. In detail, these differ considerably, not least because some Member States follow the common law copyright system and others the continental European author's right system. Although it would be useful for the creation of a single market, the considerable differences which exist make a uniform copyright law at the Union level impossible.⁷⁵ Nevertheless, the copyright laws of the EU Member States are harmonized on a selective basis by a total of four regulations and twelve directives.⁷⁶ These EU legal acts are of great importance for the global development of copyright law since they are binding for a large number of different countries.⁷⁷

In relation to TDM, this means that the copyright protection of text and data is assessed from the perspective of the respective applicable national copyright laws. On the European level, the following three legal acts are in the foreground regarding TDM: Directive 96/9/EC («Database Directive»),⁷⁸ Directive 2001/29/EC («Information Society Directive»)⁷⁹ and the above-mentioned Directive 2019/790/EU («Digital Single Market Directive»)⁸⁰

b) Databases

Databases⁸¹ are protected in the EU in two ways. On the one hand, databases are protected as such by copyright law – in accordance with the TRIPS Agreement – provided that they constitute an intellectual creation by virtue of the selection or arrangement of their contents.⁸² On the other hand, the EU grants a *sui generis* right to the producer of a database, by which the extraction and further use of the database or parts thereof can be prohibited, provided that a substantial investment was required in producing the database.⁸³ In both cases, the copyright of databases and the *sui generis* right, Member States may provide an exception for scientific research, as long as the source is indicated and a non-commercial purpose is pursued.⁸⁴

TDM can thus encompass actions that interfere with both the copyright of databases and the *sui generis* right. The focus here is again on the reproductions that occur dur-

ing the TDM process.⁸⁵ Due to the various manifestations of TDM, it remains unclear to what extent TDM interferes with the copyright of databases and the *sui generis* right and thus requires the consent of the right holder, including consideration of the respective exception provisions.⁸⁶

72 BBl 2018, 629; ISLER (Fn. 54), N 223. See BBl 2018, 629, according to which it is unclear whether, for example, a government-funded research project at a university that produces a commercially exploitable result would have to be considered as research with a commercial purpose or with a non-commercial purpose. Another example of the ambiguity would be a research project conducted by a non-profit organization that is funded by a private commercial enterprise.

73 ISLER (Fn. 54), N 221.

74 On the whole EGLOFF (Fn. 67), URG 24d N 7.

75 On the whole HILTY (Fn. 24), N 142 ff.; STERLING/COOK (Fn. 26), para. 28H.01 ff. See J. A. L. STERLING/T. COOK, General Overview, in: J. A. L. Sterling (Ed.), *Sterling on world copyright law*, 5th ed., London 2018, para. 1.79 ff. for the differences between the copyright and the author's right system.

76 See HILTY (Fn. 24), N 143 and STERLING/COOK (Fn. 26), para. 28.30 for a list of all EU regulations and directives in the field of copyright.

77 STERLING/COOK (Fn. 26), para. 28.21.

78 Directive (EC) 1996/9 of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (OJ L 77, 27.3.1996, p. 20–28).

79 Directive (EC) 2001/29 of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (OJ L 167, 22.6.2001, p. 10–19).

80 See Fn. 3.

81 See Art. 1(2) of Directive 96/6/EC, according to which for the purposes of Directive 96/6/EC, «database» shall mean a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means».

82 See Art. 3(1) of Directive 96/6/EC, according to which «databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright», and Art. 3(2) of Directive 96/6/EC, according to which «the copyright protection of databases shall not extend to their contents and shall be without prejudice to any rights subsisting in those contents themselves». See fn. 25 for Art. 10(2) TRIPS Agreement.

83 See Art. 7(1) of Directive 96/6/EC, according to which «Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database».

84 Regarding the copyright of databases see Art. 6(2)(2) of Directive 96/6/EC, according to which «Member States shall have the option of providing for limitations on the rights set out in Article 5 of Directive 96/6/EC in cases where there is use for the sole purpose of scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved». For the *sui generis* right see Art. 9(b) of Directive 96/6/EC, according to which «Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorization of its maker, extract or re-utilize a substantial part of its contents in the case of extraction for the purpose of scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved».

85 On the whole Recitals 8 and 9 of Directive 2019/790/EU.

86 GEIGER/BULAYENKO/FROSIO (Fn. 1), 825 ff.; STAMATOUIDI (Fn. 5), 280 f.; DUCATO/STROWEL (Fn. 16), 658 ff.

2. Relevant copyright barriers for text and data mining

a) Temporary copies in the Information Society Directive

According to the Information Society Directive, the author has the right to authorize or prohibit reproductions and partial reproductions in any form.⁸⁷ Exceptions to this are temporary, technically necessary reproductions that occur during transmission in a network or during legitimate use and have no economic value of their own.⁸⁸

Similar to the situation with Art. 24a CopA in Switzerland, the exception for temporary copies in the Information Society Directive is often insufficient in cases of TDM.⁸⁹

b) Exceptions and limitations in the Digital Single Market Directive

In the EU the legal uncertainty surrounding TDM was also seen as a threat to the competitive position of the Union.⁹⁰ A new exception has therefore been created in the area of scientific research («text and data mining for the purposes of scientific research»)⁹¹ Since TDM can also play an important role for private and public actors outside of scientific research, e.g. to make complex decisions which require large amounts of data analysis or to develop new technologies, the Digital Single Market Directive also creates an exception for private and public entities («exception or limitation for text and data mining»). This exception is intended to encourage innovation in the private sector.⁹² Both exemptions were to be implemented by Member States by 7 June 2021.⁹³

The provision on TDM for the purposes of scientific research provides that research organizations and cultural heritage institutions may reproduce and extract works and databases protected by the Database Directive and the Information Society Directive, and to which they have lawful access in order to conduct TDM for scientific research purposes.⁹⁴ This exception provision cannot be limited by contract.⁹⁵ Scientific research includes both the natural sciences and the humanities. Universities, higher education institutions, libraries and research institutes are considered research organizations, regardless of their legal form. However, the research organizations must not be profit-oriented or they must act in a state-recognized mission, which is characterized, for example, by state funding or a public law contract. On the other hand, organizations are not considered to be research organizations if they are subject to the determining influence of commercial enterprises, e.g. through structural circumstances such as in their capacity as shareholders, and could thereby obtain preferential access to the research results.⁹⁶ Content that is made available on an open access basis, content for whose access there is a contractual basis between the research institution and the rights holders, and content that is freely available on the internet qualifies as lawful access.⁹⁷

The provision on exception and limitation for TDM provides that for reproductions and extractions of works and databases protected by the Database Directive and the Information Society Directive made in the context of TDM,

an exception or limitation is made to the rights granted in those Directives, provided that lawful access to the works exists and rightholders do not place machine-readable conditions of use on their works.⁹⁸ Here, lawful access is understood to mean that content published on the internet is considered lawfully accessible as long as the rightholder does not prohibit TDM. A reservation may be made by the

87 See Art. 2(a) of Directive 2001/29/EC, according to which «Member States shall provide for the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part for authors, of their works».

88 See Art. 5(1) of Directive 2001/29/EC, according to which «temporary acts of reproduction referred to in Article 2 of Directive 2001/29/EC, which are transient or incidental and an integral and essential part of a technological process and whose sole purpose is to enable a transmission in a network between third parties by an intermediary, or a lawful use of a work or other subject-matter to be made, and which have no independent economic significance, shall be exempted from the reproduction right provided for in Article 2 of Directive 2001/29/EC».

89 GEIGER/BULAYENKO/FROSIO (Fn. 1), 822; STAMATOUDI (Fn. 5), 270 f.; DUCATO/STROWEL (Fn. 16), 658. See above III.2.a.

90 Recital 10 of Directive 2019/790/EU. See also ROSATI (Fn. 2), 210.

91 Recital 11 of Directive 2019/790/EU.

92 Recital 18 of Directive 2019/790/EU.

93 See Art. 29(1) of Directive 2019/790/EU, according to which «Member States shall bring into force the laws, regulations and administrative provisions necessary to comply with this Directive by 7 June 2021». See DUCATO/STROWEL (Fn. 16), 665 for an overview of the evolutions along the legislative process of the TDM exceptions in the Digital Single Market Directive.

94 See Art. 3 of Directive 2019/790/EU, according to which: «(1.) Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of Directive 2019/790/EU for reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access. (2.) Copies of works or other subject matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results. (3.) Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective. (4.) Member States shall encourage rightholders, research organisations and cultural heritage institutions to define commonly agreed best practices concerning the application of the obligation and of the measures referred to in paragraphs 2 and 3 respectively».

95 See Art. 7 of Directive 2019/790/EU, according to which «any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 of Directive 2019/790/EU shall be unenforceable».

96 On the whole Recital 12 of Directive 2019/790/EU.

97 Recital 14 of Directive 2019/790/EU.

98 See Art. 4 of Directive 2019/790/EU, according to which: «(1.) Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of Directive 2019/790/EU for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining. (2.) Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining. (3.) The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online. (4.) This Article shall not affect the application of Article 3 of Directive 2019/790/EU».

right holder, on the one hand, by machine-readable means and, on the other hand, by a contractual agreement, but also by unilateral declaration.⁹⁹

Obviously, the two exception provisions can overlap. Therefore, it is provided that the exception provision «TDM for the purposes of scientific research» (Art. 3 of Directive 2019/790/EU) is *lex specialis* to the exception provision «exception or limitation for TDM» (Art. 4 of Directive 2019/790/EU).¹⁰⁰

V. Controversies and comparison of text and data mining provisions

1. Controversies in Switzerland

The Swiss TDM provision was only slightly controversial in Switzerland, in contrast to the other copyright law amendments addressed in the Message from November 2017.¹⁰¹ Nevertheless, some differences of opinion between the various stakeholders emerged during the consultation process, which can be summarized as follows. Consumers fully agreed with the draft. Among producers and culture creators, some welcomed the draft on the condition that the publication of copyrighted works would not be allowed. Others opposed the draft outright or supported it only provided there would be no licensing option available for TDM users. Remuneration was supported by producers and culture creators. The users and the cantons agreed with the draft but criticized the remuneration obligation – which was still provided for in the draft but was ultimately deleted – because this would have led to multiple remuneration. Some users and cantons also wanted the TDM provision not to be limited by the criterion of scientific purpose. Opinions were divided among copyright holders. Some supported the TDM provision or even wanted to extend it, some demanded a limitation of the TDM provision with the criterion of licensing possibility, some supported the limitation with the criterion of scientific purpose, and some rejected the TDM provision altogether. The majority of the political parties supported the TDM provision but rejected the remuneration obligation. Some parties wanted to waive the scientific purpose criterion.¹⁰²

A typical picture emerges here for copyright protection. While consumers and users argue for an exception to copyright, culture creators and producers are in favour of maintaining far-reaching copyright protection. This is intuitive, since copyright holders want to exploit their exclusive right to the greatest extent possible and therefore seek strong copyright protection, while users and consumers are interested in the most unhindered use of works possible. With Art. 24d CopA, a compromise was worked out insofar that both the needs of researchers were implemented with the TDM exception, but at the same time a restriction to the scientific area was made in favour of copyright holders. Ultimately, it can be said that the winners are researchers and copyright holders, while private individuals and companies using TDM for non-scientific purposes are the losers.

In Swiss doctrine, the TDM provision (Art. 24d CopA) in its final form has been welcomed.¹⁰³ Despite the praise, however, there is also a central criticism of the new TDM provision: the question arises whether it was really necessary to insert the criterion of scientific purpose, since the TDM provision already requires lawful access to copyrighted works. The requirement of lawful access leads to the fact that in many cases copyright holders already have to be compensated for access to the work (e.g. a fee for the use of a database or an electronic library of a scientific publisher). From an economic point of view, it would therefore also have been possible not to tie Art. 24d CopA to the condition of scientific purpose, because access is already compensated in most cases and therefore there is no economic interest in maintaining copyright protection outside of scientific research (related to TDM).¹⁰⁴ However, waiving the criterion of scientific purpose would obviously have gone too far in view of the various interest groups involved, which is why this criterion was retained. In this respect, Art. 24d CopA cannot be considered as the result of a purely economic analysis, but of a political weighing of interests.

Eventually, in view of the various interests involved in the legislative process and the legally flawless design of Art. 24d CopA, the Swiss TDM provision can be described as successful.

2. Controversies in the European Union

The design of the European TDM provisions gave rise to criticism. First, however, the positive aspects should be considered. One of these positive aspects is the harmonizing effect of Art. 3 of Directive 2019/790/EU. Since this exception is mandatory, it creates a uniform regulation for researchers in the EU and thus promotes large, cross-border and Europe-wide research projects. Moreover, it is emphasized in particular that Art. 3 of Directive 2019/790/EU cannot be overridden by contractual agreements, but is absolutely mandatory. This ensures that countless consents are not required for the application of TDM, which was a core concern of the TDM discussion.¹⁰⁵

On the other hand, the narrow scope of Art. 3 of Directive 2019/790/EU has been criticized. This provision is limited twice: once by the criterion of scientific purpose, and

99 On the whole Recital 18 of Directive 2019/790/EU.

100 See Art. 4(4) of Directive 2019/790/EU, according to which «Article 4 of Directive 2019/790/EU shall not affect the application of Article 3 of Directive 2019/790/EU».

101 See BBl 2018, 600 for an overview of the various amendment topics in copyright law.

102 On the whole EJPD/IGE, Zwei Abkommen der Weltorganisation für geistiges Eigentum und Änderungen des Urheberrechtsgesetzes, Bericht über das Ergebnis des Vernehmlassungsverfahrens vom 2. Dezember 2016 https://www.ige.ch/fileadmin/user_upload/recht/national/d/urheberrecht/modernisierung_urheberrecht_2015_d/0212_2016_Ergebnisbericht_URG_DE.pdf (Mai 2022), 12 f.

103 ISLER (Fn. 54), N 227; BRÜHWILER (Fn. 37), N 293; P. MOSIMANN/Y. HOSTETTLER, Zur Revision des Urheberrechtsgesetzes, recht 2018, 133.

104 On the whole HILTY (Fn. 24), N 517 f.

105 On the whole GEIGER/BULAYENKO/FROSIO (Fn. 1), 833.

once by the limitation of the exception to research organizations. Apart from the fact that these two criteria give rise to various delimitation issues, the narrow scope of application means that the EU cannot keep up with innovation-promoting countries that work with the flexible instrument of «fair use».¹⁰⁶ Although the EU has created an additional exception with a broader scope of application in Art. 4 of Directive 2019/790/EU, this exception is weak if not ineffective because it allows contractual TDM prohibitions.¹⁰⁷

Furthermore, it is feared that the two TDM provisions of the EU will lead to further fragmentation, since the demarcation between Arts. 3 and 4 of Directive 2019/790/EU is not always easy.¹⁰⁸ However, fragmentation is precisely not desired in European copyright law. A major problem of European copyright law has always been that the various exceptions and limitations to copyright have been structured in different ways. Some are mandatory, some are voluntary, some apply to all protected works (horizontal exceptions), some only to certain works (vertical exceptions) and finally there are exceptions which can be overridden by contract while others take precedence over contractual agreements. Therefore, contractual non-overriding and mandatory exceptions are desirable as they help to achieve harmonization.¹⁰⁹ However, this has only been taken into consideration in Art. 3 of Directive 2019/790/EU, but not in Art. 4.

3. Comparison

The Swiss TDM provision (Art. 24d CopA) differs in several ways from the two European TDM provisions (Arts. 3 and 4 of Directive 2019/790/EU). This is due to the fact that Switzerland has decided for once not to wait for the EU regulation, but to create an exception for TDM as quickly as possible, independently from the EU.¹¹⁰

Unlike the EU, Switzerland refrains from reserving TDM for research institutions only and thus avoids complicated delimitation issues.¹¹¹ Although in Switzerland TDM must also be performed for a scientific purpose, this refers to the method and not the institution that performs TDM.¹¹² The scope of Art. 3 of Directive 2019/790/EU, on the other hand, depends on the narrow definition of research organizations.¹¹³ While the EU has created an additional TDM provision in Art. 4 of Directive 2019/790/EU that is not limited to research organizations or research for scientific purposes, this provision is severely limited by giving copyright holders the ability to contractually prohibit TDM.¹¹⁴ This ability of copyright holders to prohibit TDM can be viewed both positively and negatively. On the positive side, thanks to this opt-out solution, the criterion of scientific purpose can be waived, thus making TDM available for non-scientific research from a copyright perspective. The right to prohibit TDM represents, so to speak, the payoff for extending the TDM exception to the non-scientific field.¹¹⁵ However, the possibility for copyright holders to prohibit TDM can also be seen as a weighty disadvantage compared to the Swiss TDM provision, because a prohibition right has a similar effect as a consent requirement. Con-

sent requirements, however, were precisely what was intended to be abolished in the context of TDM. Even if copyright holders prohibit TDM by contract (e.g. in the General Terms and Conditions), this prohibition is not enforceable under Swiss jurisdiction due to the mandatory nature of Art. 24d CopA. However, authors are free to provide their works with technical protection measures that prevent TDM, since there is no obligation to make works available in a TDM-friendly manner.¹¹⁶ Paradoxically, this means that the owner of a database with non-copyrighted content can enforce a contractual TDM prohibition, whereas this is not possible for a copyright owner because of Art. 24d CopA, which can make the former better off.¹¹⁷

Furthermore, it is positive that Switzerland has chosen a technology-neutral formulation in Art. 24d CopA. This is important because, due to technological developments, it seems restrictive to permit copies for the purpose of scientific research only for TDM and not for other technologies.¹¹⁸ The TDM provisions of the EU, on the other hand, explicitly speak of text and data mining. This is seen as a disadvantage not only because it does not include other technologies but also because it legally establishes that TDM is a copyright-relevant action.¹¹⁹ This is not self-evident because, as men-

106 On the whole GEIGER/BULAYENKO/FROSIO (Fn. 1), 834 f. See also J. A. L. STERLING/S. JACQUES, *Limitations and Exceptions*, in: J. A. L. Sterling (Ed.), *Sterling on world copyright law*, 5th ed., London 2018, para. 10.131 for an explanation of the concept of «fair use». According to them, «fair use» is a limitation of copyright, which occurs in particular in U.S. copyright law. Four criteria are used to assess «fair use»: purpose and character of the use; the nature of the copyrighted work; the amount and substantiality of the portion used; and the effect upon the potential market. See also P. SAMUELSON, *Legally Speaking. Text and Data Mining of In-Copyright Works: Is It Legal?*, *Communications of the Association for Computing Machinery* 64/2021, 20 f. for two examples from the U.S., where the fair use instrument has enabled TDM related applications.

107 ROSATI (Fn. 2), 215.

108 C. SGANGA, *A new era for EU copyright exceptions and limitations? Judicial flexibility and legislative discretion in the aftermath of the Directive on Copyright in the Digital Single Market and the trio of the Grand Chamber of the European Court of Justice*, *Europäische Rechtsakademie Forum* 21/2020, 328.

109 On the whole SGANGA (Fn. 108), 315 f.

110 BBL 2018, 628.

111 ISLER (Fn. 54), N 232; HILTY (Fn. 24), N 514; MOSIMANN/HOSTETTLER (Fn. 103), 133.

112 HILTY (Fn. 24), N 514.

113 See Art. 2(1) of Directive 2019/790/EU, according to which: «research organisation» means a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research: (a) on a not-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognised by a Member State; in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organisation».

114 HILTY (Fn. 24), N 515.

115 On the whole BRÜHWILER (Fn. 37), N 269. See also BRÜHWILER (Fn. 37), N 277, according to who the restriction of the Swiss TDM provision to research for scientific purposes is a major shortcoming.

116 On the whole ISLER (Fn. 54), N 230.

117 ISLER (Fn. 54), N 231.

118 HILTY (Fn. 24), N 522; BRÜHWILER (Fn. 37), N 267.

119 ROSATI (Fn. 2), 215; DUCATO/STROWEL (Fn. 16), 667.

tioned at the beginning, not every TDM application is relevant from a copyright law perspective.¹²⁰ It is true that there is now no longer any legal uncertainty as to whether or not TDM is covered by the already existing exceptions. However, the question now arises whether TDM applications, which would not have raised copyright issues before the new TDM provisions came into force, now also fall under the new TDM provisions. New legal uncertainty has thus been created especially in the area of non-copyrighted content.¹²¹

Besides, the Swiss TDM provision only mentions the reproduction of works, whereas the TDM provisions of the EU additionally explicitly mention extraction. However, this difference is not relevant insofar as the concept of extraction primarily addresses the *sui generis* right for databases, which does not exist in Switzerland.¹²²

Finally, both the Swiss and the European TDM provisions are a step into the right direction by solving copyright problems in the field of scientific research. The attempt of the EU to create a TDM exception outside of scientific research with Art. 4 of Directive 2019/790/EU was also progressive. Since this attempt can be considered a failure due to the possibility of a contractual prohibition, the Swiss TDM solution can be considered more successful overall. This is because the scope of Art. 24d CopA is broader than that of Art. 3 of Directive 2019/790/EU. For cross-border research projects, this means that the Swiss TDM regulation represents a smaller hurdle than the European TDM regulation and is therefore more attractive for innovation. Although this may appear to be an advantage from a Swiss perspective, identical TDM regulations in Switzerland and the EU would have been desirable to avoid regional differences.

VI. Conclusion

TDM can be defined as a variety of electronic applications that can be used to analyze large amounts of text and data and discover hidden knowledge. Although the processes of the various TDM applications differ, TDM usually takes place in three steps; in the first step, access to text and data (from various sources) is obtained; in the second step, these text and data are extracted or copied and a corpus is created;

in the third step, the content of this corpus is prepared and finally analyzed. What is relevant in terms of copyright is, on the one hand, that protected works may be used and, on the other hand, that temporary and permanent copies of the content may be created during the TDM process.

In Switzerland, text and data in the form of intellectual works of individual character are protected by copyright. Databases are protected in accordance with the TRIPS Agreement provision on data collections. Additional protection exists for databases in the EU with the *sui generis* right. In Switzerland as well as in the EU, there are various exception and limitation provisions that partially cover TDM. However, only the newly introduced TDM provisions – Art. 24d CopA for Switzerland and Arts. 3 and 4 of Directive 2019/790/EU for the EU – are able to fully cover TDM.

Since the Swiss and the European TDM provisions were enacted independently from each other, there are some differences. Due to the various interests in the legislative process, in Switzerland the scope of the TDM exception was limited by the criterion of scientific purpose. The EU has opted for an even narrower scope of the TDM exception in the area of scientific research by applying the provision only to certain non-profit or governmental research organizations and cultural heritage institutions. While the second European TDM exception provision outside of scientific research was a promising innovative idea, it has been weakened with contractual excludability. Overall, therefore, the Swiss TDM regulation seems to have been more successful than the European one, at least from the perspective of promoting innovation.

The future will tell to what extent the Swiss and European TDM provisions have proven their worth. As a further outlook, it should be mentioned that the question ought to be asked whether isolated exceptions in copyright law are still appropriate for new technologies of the digital world. If one wants to offer the most innovative environment possible, copyright obstacles must be prevented. Perhaps one must even ask the fundamental dogmatic question of the extent to which automatic copyright protection in the digital environment still makes sense. At the same time, the legitimate interests of copyright holders must not be overlooked.

120 See above II.3.

121 ROSATI (Fn. 2), 215; DUCATO/STROWEL (Fn. 16), 667.

122 HILTY (Fn. 24), N 522.

Zusammenfassung

Text und Data Mining wird eingesetzt, um aus grossen Datenmengen neue Erkenntnisse zu gewinnen. Hierbei bestanden regelmässig urheberrechtliche Hindernisse, denen mit den herkömmlichen Urheberrechtsschranken nicht genügend begegnet werden konnte. Die neu geschaffenen Urheberrechtsschranken für Text und Data Mining in der Schweiz und der Europäischen Union sollen hier Abhilfe schaffen. Der Vergleich der schweizerischen und europäischen Urheberrechtsschranken für Text und Data Mining zeigt, dass erhebliche Unterschiede bei der urheberrechtlichen Handhabung von Text und Data Mining bestehen.

Résumé

Le *Text and Data Mining* est utilisé pour obtenir de nouvelles connaissances à partir de grandes quantités de données. Des obstacles liés au droit d'auteur, qui ne pouvaient pas être suffisamment contrôlés par les exceptions traditionnelles au droit d'auteur, se dressaient régulièrement face à cette pratique. Les nouvelles restrictions du droit d'auteur lors de l'utilisation du *Text and Data Mining* en Suisse et dans l'Union européenne devaient y remédier. La comparaison des exceptions au droit d'auteur prévues en droit suisse et en droit européen dans ce contexte montre qu'il existe des différences considérables dans le traitement du droit d'auteur dans le cadre du *Text and Data Mining*.