

Licence to Mine? Ein Überblick über Rahmenbedingungen von Text and Data Mining und den aktuellen Stand der Diskussion

Christian Winterhalter

*Universitätsbibliothek der Humboldt-Universität zu Berlin,
christian.winterhalter@ub.hu-berlin.de*

Abstract

Der Artikel gibt einen Überblick über die Möglichkeiten der Anwendung von Text and Data Mining (TDM) und ähnlichen Verfahren auf der Grundlage bestehender Regelungen in Lizenzverträgen zu kostenpflichtigen elektronischen Ressourcen, die Debatte über zusätzliche Lizenzen für TDM am Beispiel von Elseviers TDM Policy und den Stand der Diskussion über die Einführung von Schrankenregelungen im Urheberrecht für TDM zu nichtkommerziellen wissenschaftlichen Zwecken.

The article gives a survey about the potential application of text and data mining (TDM) or similar techniques on the basis of given licence agreements for subscription-based electronic resources. It also resumes the debate about the supplemental licence amendments for TDM that has arisen over the introduction of Elsevier's TDM Policy. Finally, it describes the current discussions about the possible implementation of copyright exemptions for TDM within the context of non-commercial scientific research.

1 Einleitung

Text Mining, Data Mining, Informationsextraktion, automatische Indexierung etc. – so sehr die genannten Verfahren Unterschiede im Detail und in ihren disziplinären Hintergründen aufweisen, etwa hinsichtlich der Strukturiertheit ihres Analysegegenstands (Text vs. Daten) oder ihres Konzepts (Klassifizierung der Elemente eines Korpus nach vorliegenden Kategorien vs. Erkennung von Mustern und gegebenenfalls nachträgliche Kategorienbildung), so sehr ist ihnen doch ein Aspekt gemein: ein weitgehend maschinell basierter Ansatz zur Verarbeitung grösserer Datenmengen, häufig verbunden mit dem Ziel der sukzessiven Automatisierung. Gemein ist den Verfahren auch, dass sie mit zunehmendem Umfang der analysierten oder bearbeiteten Daten (z.B. durch die Analyse von Volltexten statt nur ihrer Metadaten) in der Regel besser und die Analyseergebnisse genauer werden. Idealerweise besteht bei diesen Verfahren ein vollumfänglicher und dauerhafter Zugriff auf den zu analysierende Textkorpus oder das Datenset.

Maschinelle Verarbeitung, Umfang und (Zwischen-)Speicherung der zu analysierenden Texte/Daten sowie der Weg des Zugriffs auf die Daten – diese Punkte reichen an dieser Stelle bereits als grundlegende Leitfragen für die Prüfung aus, inwiefern derartige Verfahren im

Rahmen bestehender Lizenzen für den Zugriff auf elektronische Ressourcen zum Einsatz kommen können. Dabei wird der Fokus auf die Diskussion über und die Regelungen zu Text and Data Mining (nachfolgend TDM) gelegt, das stellvertretend für verschiedene maschinelle Verfahren gelten kann. Eine informationswissenschaftliche Perspektive, deren Interesse auf dem Einsatz von Verfahren zum Zwecke der automatischen Indexierung/Sacherschließung liegt, unterscheidet sich in ihren Bedarfen zunächst nicht grundlegend von denjenigen anderer Wissenschaftsdisziplinen, obschon die jeweiligen Folgeinteressen teilweise divergieren mögen (Anwendungsinteresse, Nachnutzbarkeit etc.).¹

2 Regelungen in aktuellen Lizenzverträgen

Betrachtet man gängige Passagen aus Lizenzverträgen zu subskriptionspflichtigen elektronischen Produkten (Zeitschriften, eBooks, Datenbanken), finden sich zumeist nahezu gleichlautende Passagen wieder, die den Einsatz der erwähnten Verfahren regeln:

Anbieter	Vertragsauszug
Elsevier B.V., Standard Agreement 2014	1.3 Subscriber and its Authorized Users may not: abridge, modify, translate or create any derivative work based on the Subscribed Products, except to the extent necessary to make them perceptible on a computer screen to Authorized Users; [...] use any robots, spiders, crawlers or other automated downloading programs, algorithms or devices to continuously and automatically search, scrape, extract, deep link, index or disrupt the working of the Subscribed Products; ²
Springer	2. Licensee and its Authorized Users shall not: update, change, revise, adapt, modify, translate, transform or create any derivative work of the Content, [...] systematically download any Content, use routines designed to continuously and automatically search and index the Content (full text and meta data), such as web-crawling or spider programs or engage in any activity likely to burden Springer's Websites ³
Wiley	C.1.c Authorized Users may search, view, and browse Licensed Electronic Products using the interface provided by Wiley. Automated searching, robotic searching and decompilation are strictly prohibited. ⁴

¹Die Frage der Nachnutzung derart gewonnener Ergebnisse zum Zwecke der Anreicherung eigener Dienste z.B. mit Sacherschließungselementen (Repositorien, Fachportale etc.) würde eine gesonderte Betrachtung erfordern. Eines der Kernprobleme dürfte in diesen Fällen jedoch in der öffentlich zugänglichen Verwendung der Ergebnisse im Rahmen eigener Dienstleistungen liegen, die teilweise in Konkurrenz zu den Produkten der Anbieter stehen könnten. Dabei handelt es sich um eine Form der Nutzung der Ergebnisse, welche vielfach ausgeschlossen oder eingeschränkt ist.

²https://www.elsevier.com/__data/assets/pdf_file/0008/90935/Sample_Elsevier_BV_Subscription_Agreement-Academic-Government_022514.pdf

³<http://www.springer.com/gp/librarians/terms-and-conditions-for-online-products-of-springer/18520>

⁴<http://onlinelibrary.wiley.com/licenseAgreement>

American Physical Society (APS)	Prohibitions on Certain Use: 6.B. Systematic or programmatic downloading of the Licensed Materials (for example, downloading entire journal issues), [...] and/or the systematic making of print or electronic copies is prohibited. Downloading portions of the Licensed Materials for the purpose of creating systematic and persistent local copies (not including transient, dynamic caches of individually requested material) is prohibited. [...] D. In the event that abusive usage of the online access threatens the integrity and effectiveness of the Publisher's online distribution system, the Publisher has the right to discontinue access immediately and will inform the subscriber of the abuse within one business day. Online access will be restored as soon as practical after the abuse is corrected. ⁵
American Chemical Society (ACS)	6. Prohibited Uses [...] b. Authorized Users may not modify, alter, or create derivative works of the materials contained in the ACS Products without prior written permission from ACS. Indexing – by human or machines means, aggregating, data mining, peer-to-peer (or similar) file-sharing are all prohibited uses unless an institution concludes a specific, separate license with ACS to do so. [...] d. Licensee acknowledges that ACS may prevent Licensee and its Authorized Users from using, implementing, or authorizing use of any computerized or automated tool or application to search, index, test, or otherwise obtain information from ACS Products (including without limitation any “spidering” or web crawler application) that has a detrimental impact on the use of the services under this License. ⁶

Die Formulierungen aus den exemplarisch ausgewählten Standardlizenzverträgen grosser Anbieter unterscheiden sich eher im Detail; der Schwerpunkt wird einmal mehr auf die verschiedenen Möglichkeiten einer automatisierten Suche und des Downloads sowie die potentiellen Auswirkungen auf die Stabilität der Anbietersysteme gelegt, dann wieder mehr auf die bearbeitende Abwandlung von Werken oder auf die Speicherung lizenzierten Materials bzw. Teilen davon.

Auf der anderen Seite haben einzelne Verlage wie Oxford University Press (OUP) – nicht zuletzt vor dem Hintergrund der UK Copyright Exemption⁷, die TDM explizit für nichtkommerzielle wissenschaftliche Zwecke zulässt – inzwischen ihre Verträge so angepasst, dass dort die angeführten Restriktionen nicht mehr vorkommen, keine expliziten Regelungen zu TDM enthalten sind oder eine explizite Erlaubnis dafür erforderlich wäre.⁸ Ausserhalb der vertraglichen Regelungen wird jedoch darum gebeten, TDM-Vorhaben im Vorfeld abzustimmen:

„Oxford University Press recognizes the research benefit of Text and Data Mining (TDM) across a variety of research fields. As such, we are happy to accommodate TDM for non-commercial use. Although researchers are not required to request permission for non-commercial text-mining, OUP is happy to offer consultation

⁵https://librarians.aps.org/site_license

⁶<http://pubs.acs.org/userimages/ContentEditor/1224170901223/multisite.pdf>

⁷<https://www.gov.uk/guidance/exceptions-to-copyright#text-and-data-mining-for-non-commercial-research>

⁸<http://www.oxfordjournals.org/resource/image/institutional-site-licence-agreement-v2-8-7-16.pdf>

with a technical project manager to assist in planning your project, including avoidance of any technical safeguards triggers OUP has in place to protect the stability and security of our websites.“⁹

Ohne auf die vielen Nuancen eingehen zu können, die im Zwischenbereich dieser entgegenstehenden Positionen angesiedelt sind, lässt sich jedoch feststellen, dass auf Grundlage der bisher üblichen Klauseln in den Lizenzverträgen in einer grossen Zahl der Fälle nach wie vor keine oder nur sehr eingeschränkte Möglichkeiten für TDM oder ähnliche Verfahren bestehen. Werden derartige Verfahren entgegen den vertraglichen Regelungen angewandt, kommt es regelmässig zu meist automatisierten Meldungen der Anbieter über Verstösse oder vermutete Verstösse gegen die Lizenzbedingungen, die in der Regel mit der Sperrung des IP-Bereichs einhergehen, der als Ausgangspunkt für einen Verstoß identifiziert wurde. Wie sensibel die Systeme der Anbieter teilweise eingestellt sind, kann man immer wieder an Vorfällen beobachten, die sich z.B. auf eine Nutzung der „Find-Full-Text-Option“ in gängigen Literaturverwaltungsprogrammen zurückführen lassen.

Nach dem Befund der vertraglichen und technischen Schwierigkeiten, die bei einer Vielzahl von lizenzierten Ressourcen bestehen, läge natürlich die Frage nahe, ob sich die bestehenden prohibitiven Regelungen in dem Sinne modifizieren oder ergänzen lassen, dass entweder keine explizit formulierten Beschränkungen mehr enthalten sind (wie im Falle von OUP), oder aber durch Neuverträge oder Nebenabreden in Amendments/Addenda zu bestehenden Verträgen die Möglichkeit von TDM vorgesehen werden. Doch unabhängig von der praktischen Umsetzbarkeit, deren eine Seite einen hohen Aufwand, und deren andere Seite die spezifische Konstellation der Kräfteverhältnisse in Verhandlungen darstellt: Ist es überhaupt sinnvoll, die Frage in dieser Form zu stellen und den „Lizenzweg“ einzuschlagen?

3 Licence to Mine? Der Weg über Lizenzen

Prinzipiell wäre ein derartiger Weg bereitet, denn fast alle der oben genannten Verlage bieten inzwischen eine Standardklausel zur Regelung von TDM an. So enthält der oben zitierte Vertrag von Elsevier einen explizit TDM gewidmeten Passus, der – abweichend von den zitierten untersagten Nutzungsarten – TDM als Möglichkeit zulässt, jedoch bestimmte Begrenzungen im Hinblick auf den technischen Zugang und die Nutzung der Ergebnisse geltend macht. Die nachfolgenden Vertragspassagen (Stand des Sample Agreement: 25.02.2014) stellen die konkrete Umsetzung der Policy dar, die Elsevier Ende Januar 2014 veröffentlichte.¹⁰ Unter Punkt 1.3 „Authorized Uses“ heisst es:

„Subscriber may:

- access the text and data mining service online via an API at <http://www.developers.elsevier.com> to continuously and automatically extract and index and/or process information from the Subscribed Products to which the Subscriber separately subscribes and load and integrate the results (the ‘TDM Output’) on the Subscriber’s text-mining system for access and use by Authorized Users; and

⁹<http://www.oxfordjournals.org/en/help/third-party-data-mining.html>

¹⁰<https://www.elsevier.com/connect/elsevier-updates-text-mining-policy-to-improve-access-for-researchers>

- distribute the TDM Output externally, which may include a few lines of query-dependent text of individual full text articles or book chapters which will be up to a maximum length of 200 characters surrounding and including the text entity matched ('Snippets') or bibliographic metadata. Where Snippets and/or bibliographic metadata are distributed, they should be accompanied by a DOI link that points back to the individual full text article or book chapter. Further the TDM Output should include a Creative Commons proprietary notice in the following form:

'© Some rights reserved. This work is distributed under the terms of the CC-BY-NC Attribution-NonCommercial 4.0, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.'

Text and data mining services online may be accessed by vendors or other third parties retained by the Subscriber only with the express written permission of Elsevier and for the index and/or process information purposes of the subscriber.¹¹

Der anzubringende Urheberrechtshinweis in anderen Vertragsfassungen und auch aktuell auf der Webseite zur TDM Policy lautet abweichend davon: „Further the TDM Output should include a notice in the following form: *'©Some rights reserved. This work permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.'*“¹²

Diese Abweichung dürfte eng mit dem im Vergleich zur Ursprungsversion ebenfalls neuen Hinweis in Zusammenhang stehen, dass Rechte an Bildern und Abbildungen in den Volltexten separat mit den Rechteinhabern geklärt werden müssen.

Die vertragliche Regelung mit der jeweiligen Institution stellt jedoch nur einen Aspekt der Regelung zu TDM dar. Auf dieser Basis können berechtigte Nutzer der Institution über die von Elsevier bereitgestellte Schnittstelle TDM unter Einsatz eines persönlichen API Key durchführen, jedoch nur, wenn sie zuvor dem individuellen Click-Through Agreement zugestimmt und sich auf der Developer Platform registriert haben.¹³

Mit der Implementierung entsprechender TDM-Regelungen in Verträgen vertritt Elsevier einen lizenzbasierten Ansatz zur Regelung von TDM und anderen Verfahren. Dieser Ansatz ist nicht solitär, sondern fügt sich in die gemeinsamen Bestrebungen grosser STM-Verlage ein. Pointiert kommt dieser Ansatz in der gemeinsamen Erklärung „Text and Data Mining for Non-Commercial Research. A Statement by STM Publishers to a Roadmap to Enable Text and Data Mining (TDM) for Non Commercial Scientific Research in the European Union“ aus

¹¹https://www.elsevier.com/__data/assets/pdf_file/0008/90935/Sample_Elsevier_BV_Subscription_Agreement-Academic-Government_022514.pdf. In einem Mustervertrag für den US-Bereich ist der entsprechende Passus aufgrund der unterschiedlichen Rahmenbedingungen für TDM nicht enthalten. Hier finden sich lediglich Passagen, die sich auf die Nutzungen von Daten von Artikeln und deren Einbindung in ein institutionelles Repositorium beziehen, die von Autoren der betreffenden Einrichtung stammen; dieser Passus wird auch auf dem europäischen Markt angewandt: https://www.elsevier.com/__data/assets/pdf_file/0008/90935/Sample_SD_US_Academic_16Jun2016_CLEAN.pdf.

¹²<https://www.elsevier.com/about/company-information/policies/text-and-data-mining/elsevier-tdm-license>

¹³Aktuelle Fassung des TDM Service Agreement vom 21. Juni 2016 unter http://dev.elsevier.com/tdm_service_agreement.html. Dort findet sich auch die Version vom 22. Mai 2015.

dem November 2013 zum Ausdruck. Dort erkennen die unterzeichnenden Verlage prinzipiell den Bedarf der nichtkommerziellen wissenschaftlichen Forschung an und erklären ihre Verpflichtung, TDM zu erleichtern, „*Reaffirming* that licensing is the smart and speedy route to providing access and the rights needed for text and data mining, and for related technologies such as text-to-speech and automated translation services“. Als erste Selbstverpflichtung der Verlage wird dort des Weiteren genannt:

„Commitment 1: The Sample Clause for non-commercial Text and Data Mining in scientific research

As part of authorising text and data mining, signatories commit either to implementing the STM sample clause for non-commercial text and data mining into existing agreements (on request or as part of subscription renewal), or to offering researchers and institutions to which researchers are affiliated comparable and equivalent access rights for the purpose of non-commercial text and data mining of subscribed journal content for non-commercial scientific research, at no additional cost to researchers/subscribing institutions.“¹⁴

Deutlich tritt in dieser Passage die Grundannahme zutage, dass TDM es erforderlich mache, dem Lizenznehmer im Rahmen von vertraglichen Vereinbarungen einseitig weitere Nutzungsrechte einzuräumen. Mittlerweile haben zahlreiche Verlage diesen Weg beschritten und entsprechende, ergänzende Abschnitte oder Anhänge vorgesehen, häufig sekundiert von einer TDM Policy. Als Beispiel kann an dieser Stelle die Zeitschrift *Science* genannt werden, deren ausführlicher Annex A des institutionellen Lizenzvertrags der Regelung von TDM gewidmet ist.¹⁵ Auch Springer sieht eine Ergänzung bestehender Verträge im Falle von Vertragserneuerungen oder auf Anfrage vor.¹⁶ Analoges gilt für den Verlag Wiley, der die Möglichkeit zum Abschluss eines umfangreichen Text and Data Mining Agreement anbietet.¹⁷ Ein synoptischer Vergleich der verschiedenen Regelungen würde zwar Unterschiede in der spezifischen Umsetzung verdeutlichen, insgesamt hat jedoch die Musterlizenz der International Association of STM Publishers¹⁸ zu einer gewissen Standardisierung der vertraglichen Regelungen geführt, wenngleich sich die Lage aus der Perspektive der Wissenschaftler zweifellos unübersichtlich darstellt.

Was die technische Durchführung von TDM angeht, wird von grösseren Verlagen der Weg über eine API vorgegeben, oder – als Alternative –, auch über eine Crossref API. Letztere bietet gleichzeitigen Zugriff auf die von der Einrichtung lizenzierten Inhalte verschiedener Verlage, entbindet den Wissenschaftler jedoch nicht von der Zustimmung zu einer separaten Click-through Licence (sofern durch den Verlag verlangt).¹⁹

Müsste man demnach insgesamt zum Fazit kommen, dass die anfangs beschriebenen Restriktionen eine deutliche Abschwächung erfahren haben, indem nunmehr sehr viele der genannten Verlage TDM Policies veröffentlicht haben und sich diese in detaillierten Regelungen in den Lizenzverträgen oder Zusatzvereinbarungen manifestieren, und dies sogar ohne zusätzliche

¹⁴http://www.stm-assoc.org/2015_11_10_Text_and_Data_Mining_Declaration.pdf

¹⁵Science Online Journals Institutional License Agreement. Terms and Conditions for Text and Data Mining (“TDM”): <http://www.sciencemag.org/subscribe/institutional-license-agreement>.

¹⁶<https://www.springer.com/gp/rights-permissions/springer-s-text-and-data-mining-policy/29056>

¹⁷<http://olabout.wiley.com/WileyCDA/Section/id-826542.html>

¹⁸http://www.stm-assoc.org/2012_03_15_Sample_Licence_Text_Data_Mining.pdf

¹⁹<http://tdmsupport.crossref.org/> und ausführlich vgl. Lammey (2015).

Kosten? Wäre diese dargestellte Initiative demnach das Resultat erfolgreicher Argumentation der Wissenschaftsorganisationen einerseits und Einsicht auf Seiten der Verlage andererseits, so dass zu diesem Thema kein hartes Ringen um Formulierungen in Lizenzverträgen mehr nötig wäre? Oder sind die Lizenzergänzungen vielmehr ein Danaergeschenk, vor dessen leichtfertiger Annahme Forschungseinrichtungen und Bibliotheken sich hüten sollten?

4 Oder doch ganz anders? „The right to read is the right to mine“

Der Grund, weshalb speziell Elseviers TDM Policy vorangehend ausführlich zitiert wurde, ist weniger ein spezifischer Unterschied oder eine ggfs. restriktivere Ausgestaltung im Vergleich zu anderen genannten Beispielen. An der Veröffentlichung der TDM Policy von Elsevier hat sich jedoch in grosser Deutlichkeit manifestiert, dass die Initiative der International Association of STM Publishers und nachfolgende Veröffentlichung von Elseviers neuer TDM Policy nicht als voraussetzungslose Eigeninitiative anzusehen war. Vielmehr war das zitierte Statement der International Association of STM Publishers ein Resultat aus dem Stakeholder-Dialog der Initiative „Licences for Europe“ der Europäischen Kommission, in deren Rahmen auch eine Working Group speziell zu TDM eingerichtet worden war.²⁰ Zu einem gemeinsamen Statement kam es jedoch nicht mehr, da zehn Verbände zur Hälfte der Beratungen ihren Austritt aus der Working Group erklärt hatten. Hintergrund war die Fokussierung auf den „Lizenzweg“ zur Schaffung eines Rahmens für TDM-Aktivitäten:

„The majority of researchers and research institutions pointed out that mining of data and texts to which a lawful access had been obtained should not be subject to an additional licence or payment. In their perspective the ‘right to read includes the right to mine’. [...] 10 organisations representing essential interests within the research community (including the only representative of open access publishers) have left the Working Group during the process. Although they welcomed the orientation debate by the Commission in December 2012 and the subsequent commitment to adapt the copyright framework to the digital age, they believe that having placed licensing as the central pillar of the discussion has not made a focused evaluation possible.“²¹

Um die Einführung der neuen TDM Policy von Elsevier und anderen Verlage zu verstehen, muss man sich diesen Kontext des grundlegenden Dissenses vergegenwärtigen. Die Einführung neuer Vertragsklauseln kann somit eher als ein Versuch der vertraglichen Fixierung von Rahmenbedingungen im Sinne der Verlegerinteressen speziell für TDM interpretiert werden, bevor sich durch potentielle Schrankenregelungen im Urheberrecht bestimmte Aspekte ggfs. nur noch schwer durchsetzen liessen (zumindest jene vertraglichen Aspekte, die durch eine Schrankenregelung nicht als hinfällig anzusehen wären).

Wenn der Versuch der Verlage durch den Kontext der Diskussion besser eingeordnet werden kann, so auch die deutliche Reaktion seitens LIBER und anderer, teilweise zuvor am Stakeholder-Dialog beteiligter Wissenschaftsorganisationen. In dem von LIBER koordinierten Offenen Brief „Realising the innovative potential of digital research methods: a call from

²⁰Vgl. <https://ec.europa.eu/licences-for-europe-dialogue/node/7.html>.

²¹<https://ec.europa.eu/licences-for-europe-dialogue/sites/licences-for-europe-dialogue/files/WG4-Text-and-Data-Mining.pdf>

the research community“²² vom 1. Juli 2014 wird Elsevier aufgefordert, seine TDM Policy zurückzuziehen. Als Kritikpunkte werden u.a. angeführt: die Vorgabe der Nutzung der API, die individuell von Wissenschaftlern abzuschliessende Click-through Licence, die jederzeit einseitig geändert werden kann, die Notwendigkeit der Erstellung von Kopien für effizientes TDM, die Konzentration auf technische Massnahmen zur Verhinderung von unerlaubten Zugriffen statt der Entwicklung von technisch stabileren Plattformen und zuletzt der Vorhalt, dass die hohen Lizenzkosten nicht nur mit dem Anbieten von Inhalten zu rechtfertigen seien, sondern jenen auch entsprechende Services gegenüberstehen müssten.

Als Kern der Argumentation kann jedoch gelten, dass die Regelungen der Verlage am Bedarf der Wissenschaft vorbeigehen, die Anwendung von TDM eher behindern denn fördern und letztlich der Weg über Lizenzen eine Sackgasse darstelle und zur Schaffung von Rechtssicherheit eine auf europäischer Ebene harmonisierte Schrankenregelung – wie sie zu diesem Zeitpunkt in Grossbritannien schon eingeführt war – geboten sei.

Elseviers Reaktion durch Policy Director Gemma Hersh erfolgte am 10. Juli 2014 in Form eines Antwortschreibens. In diesem wies Elsevier zurück, mit der Einführung der neuen TDM Policy die Bemühungen der Wissenschaftsorganisationen unterminieren zu wollen, Wissenschaftler zu kontrollieren und mit administrativen Bürden zu versehen und technisch unzureichende und limitierende Möglichkeiten für ein effizientes TDM anzubieten. Zudem stehe die Regelung mit der britischen Schrankenregelung im Einklang.²³ In der Folge bemühte sich Elsevier zu betonen, dass die Einbindung der TDM-Regelungen in Lizenzverträgen für die lizenzierenden Einrichtungen fakultativ sei und auch abgelehnt werden könne. Die Durchführung von TDM-Aktivitäten sei jedoch an akzeptierte TDM-Regelungen gebunden.

5 Schranken für die Zukunft?

An der öffentlich geführten Diskussion um Elseviers TDM Policy tritt der Grunddissens zwischen Grossverlagen und Wissenschaftsorganisationen sehr pointiert zutage, d.h. der Lizenzweg vs. den Ansatz des „The right to read is he right to mine“ (vgl. auch Böttcher (2015) und Murray-Rust, Molloy und Cabell (2012)), der durch eine weitgefassete Schrankenregelung im Urheberrecht bzw. eine allgemeine Wissenschaftsschranke abgesichert werden soll. Ausführlich wird der Standpunkt vieler Wissenschaftsorganisationen in der von LIBER initiierten „The Hague Declaration on Knowledge Discovery in the Digital Age“ dargelegt, die im Dezember 2014 entstand²⁴; zusätzlich wird im Projekt „Future TDM“ aktuell an Empfehlungen für wissenschaftsorientierte Rahmenbedingungen von TDM gearbeitet.²⁵

Wenngleich die skizzierten Meinungsverschiedenheiten nach wie vor nicht aufgelöst sind, so war eine Folge der Auseinandersetzung zweifellos die deutlich stärkere Präsenz und Diskussion des Themas in den nachfolgenden Monaten, sei es durch die Einrichtung einer ad-hoc-Arbeitsgruppe TDM der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen²⁶, die auch eine Befragung zu den Bedarfen der

²² Vgl. <http://libereurope.eu/wp-content/uploads/2014/07/Open-Letter-To-Elsevier1.pdf>. Vgl. auch <http://sparcopen.org/our-work/developments-in-tdm-policy/>.

²³ Vgl. https://www.elsevier.com/__data/assets/pdf_file/0008/84464/TDM_openletter.pdf

²⁴ <http://thehaguedeclaration.com>

²⁵ <http://project.futuretdm.eu>

²⁶ <http://www.allianzinitiative.de/handlungsfelder/querschnittsthemen/nutzungsrechte/ad-hoc-arbeitsgruppe/>

Wissenschaft hinsichtlich TDM und einen sich anschliessendem Workshop durchgeführt hat (Allianz der Deutschen Wissenschaftsorganisationen, Schwerpunktinitiative Digitale Information 2015), sei es durch verschiedene weitere Veranstaltungen und Projekte im deutschsprachigen Raum, in deren Rahmen über Anwendungsszenarien, technische Infrastruktur und Rahmenbedingungen für TDM diskutiert wurde und wird.²⁷

Mittelfristig dürfte durch den im September 2016 veröffentlichten „Vorschlag für eine RICHTLINIE DES EUROPÄISCHEN PARLAMENTS UND DES RATES über das Urheberrecht im digitalen Binnenmarkt“ (COM/2016/0593 final – 2016/0280 (COD))²⁸ erneut Bewegung in die Diskussion kommen, der TDM als einen der wesentlichen Bereiche thematisiert, für den Rechtssicherheit herbeigeführt werden müsse, nicht zuletzt um Nachteile der Forschung in der Europäischen Union im Vergleich zu anderen Ländern zu vermeiden. Auf der Basis der Beratungen und Studien wurde unter mehreren Optionen die nachfolgende als die am stärksten zwischen den Interessen der Verlagswirtschaft und der Wissenschaftsorganisationen ausgewogene Formulierung erachtet:

TEIL II, Artikel 3: „Text- und Data-Mining

1. Die Mitgliedstaaten sehen eine Ausnahme von den in Artikel 2 der Richtlinie 2001/29/EG und in Artikel 5 Buchstabe a und Artikel 7 Absatz 1 der Richtlinie 96/9/EG sowie in Artikel 11 Absatz 1 dieser Richtlinie festgelegten Rechten für Vervielfältigungen und Entnahmen vor, die durch Forschungsorganisationen von Werken oder sonstigen Schutzgegenständen, zu denen sie für die Zwecke der wissenschaftlichen Forschung rechtmäßig Zugang haben, für das Text- und Data-Mining vorgenommen wurden.
2. Jede Vertragsbestimmung, die der in Absatz 1 festgelegten Ausnahme zuwiderläuft, ist unwirksam.
3. Rechteinhaber müssen Maßnahmen anwenden können, um die Sicherheit und Integrität der Netze und Datenbanken zu gewährleisten, in denen die Werke oder sonstigen Schutzgegenstände gespeichert sind. Diese Maßnahmen dürfen über das für die Erreichung dieses Ziels Notwendige nicht hinausgehen.
4. Die Mitgliedstaaten wirken darauf hin, dass sich Rechteinhaber und Forschungsorganisationen gemeinsam auf Verfahren einigen, die sich für die Anwendung der in Absatz 3 genannten Maßnahmen bewährt haben.“

Diese Richtlinie sieht zwar eine Schrankenregelung für TDM vor, enthält jedoch zugleich Punkte, die in der Praxis zu vielfältigen Problemen mit TDM führen können, z.B. durch die noch unbestimmten technischen Massnahmen der Rechteinhaber, die bereits bisher die Diskussion geprägt haben (TDM auf der Web-Plattform oder nur über API). Eine Andeutung der Diskussion scheint bereits im unmittelbar erfolgten Statement der International Association of

²⁷From Big Data to Smart Knowledge – Text and Data Mining in Science and Economy, Köln, 23.–24. Februar 2015, <http://textmining.congressbuero.de>; Text and Data Mining. Discovery of Knowledge in the Digital Age, Bern, 9. Juni 2015, <http://www.consortium.ch/veranstaltungen/>; Daten, nichts als Daten? Fortbildung für Fachreferentinnen und Fachreferenten der Naturwissenschaften, Berlin 28.-29.9.2015, <http://www.ub.hu-berlin.de/vdb-nawi-2015>. Ein gemeinsames Projekt Elseviers und der Humboldt-Universität zu Berlin ist das HEADT-Centre, in dessen Rahmen u.a. zu Fragen der Infrastruktur und Methoden für TDM geforscht wird: <http://headt.eu/>.

²⁸<http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016PC0593>

STM Publishers auf, die eine Unbestimmtheit des „rechtmässigen Zugangs“ beklagt: „Within Article 3 ‘Text and Data Mining’, ‘lawful access’ is not sufficiently defined. In its current framing, ‘lawful access’ could cover unforeseen uses such as copying rented or deposited content indeed accessing illegally hosted material. It is imperative that ‘lawful access’ be combined with the consent of the rightholders.“²⁹

Mit diesem Formulierungsvorschlag wird die skizzierte Diskussion somit keineswegs enden, sondern nur in eine neue Phase eintreten. Abzuwarten bleibt der Zeitplan für die weitere Behandlung und Umsetzung der Richtlinie in den Urheberrechtsgesetzen der Mitgliedsstaaten. Bereits seit 2014 verfügt Grossbritannien über eine Schrankenregelung. Mit dem unlängst verabschiedeten „LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1)“ wurde in Frankreich ebenfalls eine Schrankenregelung für TDM in den Code de la propriété intellectuelle eingeführt, die TDM für Zwecke der nichtkommerziellen wissenschaftlichen Forschung ermöglicht. Bzgl. des Umgangs mit den dabei anfallenden Kopien bestehen stärkere Auflagen als in Grossbritannien (langfristige Archivierung nur durch per Erlass bestimmte Einrichtungen, d.h. voraussichtlich die Bibliothèque nationale). Weitere Klarheit über die Ausgestaltung dieser Schrankenregelung wird mit der Veröffentlichung der Erlasse bestehen, die für Januar 2017 angekündigt sind.³⁰

In Deutschland ist der weitere Verlauf der Reform des Urheberrechts im Sinne einer allgemeinen Bildungs- und Wissenschaftsschranke, deren Umsetzung die Regierung laut Koalitionsvertrag bis zum Ende der aktuellen Legislaturperiode plant, noch ungewiss. Mehrere Vorschläge zu einer Reform des Urheberrechts und dessen Anpassung an die Erfordernisse von Wissenschaft und Bildung unter digitalen Bedingungen – TDM eingeschlossen – liegen teilweise seit Jahren vor (vgl. Kuhlen 2016:6–7).

6 Was tun in der Zwischenzeit?

Auf dem Weg über „traditionell“ in Lizenzverträgen enthaltene Formulierungen, Erweiterungen zum Zweck von TDM und am Ende über Schrankenregelungen oder eine Allgemeine Bildungs- und Wissenschaftsschranke für Verfahren wie TDM hat sich der Überblick nur scheinbar von der Frage nach den Möglichkeiten automatisierter Verfahren entfernt, wird doch der zuletzt skizzierte Rahmen wesentlich die konkrete Praxis in ihrem Umfang, ihren organisatorischen Abläufen und technischen Möglichkeiten bestimmen.

Zum jetzigen Zeitpunkt auf den „Lizenzweg“ zu setzen, kann daher sowohl inhaltlich wie politisch als problematisch gelten. Auch die Hoffnung, dass restriktive vertragliche Vereinbarungen, wie sie von den Verlagen vertreten werden, durch zukünftige geltende Schrankenregelungen nichtig würden, und deren Abschluss daher zum jetzigen Zeitpunkt unproblematisch wäre, erscheint als riskantes Unterfangen; welche der Detailregelungen möglicherweise aufgehoben würden, weil sie einer generellen Ermöglichung von TDM zuwiderlaufen, ist keineswegs ausgemacht und müsste im Einzelfall geklärt werden.

²⁹http://www.stm-assoc.org/2016_09_14_News_Release_STM_response_to_Directive_on_Copyright_in_the_Single_Digital_Market.pdf

³⁰Vgl. <http://www.irstea.fr/toutes-les-actualites/loi-numerique-openaccess-ecrits-et-donnees-scientifiques>. Die vom Senat am 3. Mai 2016 im Rahmen des Gesetzgebungsverfahrens verabschiedete Fassung sah eine deutlich weitreichendere, wissenschaftsfreundlichere Regelung vor: Vgl. Article 18bis unter <http://www.senat.fr/leg/tas15-131.html>.

Das Konsortium der Schweizer wissenschaftlichen Bibliotheken (CSAL) hat als Reaktion auf die Anforderung von Verlagen, TDM-Amendments zu unterzeichnen, eine eigene Musterformulierung entworfen: „Authorized Users are permitted to engage in Text and Data Mining (TDM) meaning: downloading, extracting and indexing information automatically from the Licensed Materials. Where required, TDM includes mounting, loading and integrating the results on a server used for the Licensee’s and/or its Member Sites’ text-mining system and evaluate and interpret the TDM output for access and use by Authorized Users. The Licensee and its Member Sites shall ensure compliance with Licensee’s obligations of this Agreement, including but not limited to security and technical access requirements.“³¹ Auch die Musterlizenz für die sog. Allianzlizenzen oder die Grundsätze für Lizenzen der durch die DFG geförderten Fachinformationsdienste (FID) sehen wissenschaftsorientierte Regelungen von TDM vor.³²

Auch wenn man grundsätzlich darüber diskutieren kann, in welchem Masse der „Lizenzweg“ zum jetzigen Zeitpunkt beschritten werden sollte, erscheinen diese Fassungen jedoch insofern unproblematisch und gangbar, als dadurch keiner restriktiven Regelung zugestimmt wird und keine falschen Standards gesetzt werden, die zukünftigen Entwicklungen zuwiderlaufen. Muss abweichend davon in spezifischen Fällen (wie zum Beispiel bei Projekten) oder etwa zum Zwecke von Tests der TDM-Funktionalitäten auf eher restriktiver gestaltete Bedingungen eingegangen werden, so erscheint es zwingend geboten, diese ausschliesslich als separates und zeitlich befristetes Amendment abzuschliessen.

Bis die Möglichkeiten von TDM und anderen Verfahren allgemein in einer Form genutzt werden können, die den Bedarfen der wissenschaftlichen Forschung entspricht, werden zweifellos noch intensive Diskussionen nötig sein. Auch an diesem Punkt ist letztlich nur Open Access ein Teil der Antwort auf die beschriebenen Herausforderungen, nicht zuletzt für das Grundproblem, dass sich TDM in der Regel lediglich auf die jeweils von einer Einrichtung lizenzierten Inhalte bezieht. Das Publizieren im Open-Access-Modell unter Nutzung von Lizenzen, die TDM und andere Verfahren nicht erschweren, kann somit Schlüssel und Treiber sein, um die skizzierten Schwierigkeiten für eine Übergangszeit zu minimieren und langfristig zu lösen (vgl. Eskevich 2016; Herb 2015).

Literatur

Allianz der Deutschen Wissenschaftsorganisationen, Schwerpunktinitiative Digitale Information (2015). *Bedarf und Anforderungen an Ressourcen für Text und Data Mining: Zusammenfassung der Ergebnisse einer Umfrage aus dem Zeitraum April bis Mai 2015*. DOI: [10.2314/GBV:836171284](https://doi.org/10.2314/GBV:836171284).

Böttcher, C. (2015). Text and data mining freedom at stakes in copyright reform. In: *EuroScientist* 3. Juni. URL: <http://www.euroscientist.com/text-and-data-mining-freedom-at-stakes-in-copyright-reform/>.

³¹http://www.consortium.ch/wp-content/uploads/2015/05/Text_and_Data_Mining_Pascal_Boutsiouci.pdf. Diese Formulierung ist gar als Standardformulierung in den Mustervertrag der ASM (American Society for Microbiology) eingegangen, vgl. <http://www.asmscience.org/files/multisubscription1.pdf>.

³²Vgl. <https://www.nationallizenzen.de/tools/al-musterlizenz> bzw. http://www.fid-lizenzen.de/ueber-fid-lizenzen/dateien/Grundsaeetze_FIDLizenzen.pdf.

- Eskevich, M. (2016). Stakeholders in Academic Publishing. Text and Data Mining Perspective and Potential. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas; Proceedings of the 20th International Conference on Electronic Publishing*. Hrsg. von F. Loizides und B. Schmidt. Amsterdam: IOS Press. URL: <http://resolver.sub.uni-goettingen.de/purl?gs-1/13343>.
- Herb, U. (2015). Text Mining. Der Streit um digitales Gold in der Wissenschaft. In: *Telepolis*. 17.06.2015. URL: <http://www.heise.de/tp/artikel/45/45203/>.
- Kuhlen, R. (2016). Die Allgemeine Bildungs- und Wissenschaftsschranke im Urheberrecht nimmt Fahrt auf – Kreativität und Innovation werden die Gewinner sein. In: *Information. Wissenschaft & Praxis* 67.1, S. 1–9. DOI: [10.1515/iwp-2016-0004](https://doi.org/10.1515/iwp-2016-0004).
- Lammey, R. (2015). CrossRef text and data mining services. In: *Science Editing* 2.1, S. 22–27. DOI: [10.6087/kcse.32](https://doi.org/10.6087/kcse.32).
- Murray-Rust, P., Molloy, J. C. und Cabell, D. (2012). *Open Content Mining*. Conference for the Fellows of OpenForum Academy – 24th September 2012 Brussels. URL: <http://www.dspace.cam.ac.uk/handle/1810/243749>.