

Felicitas Kleinkopf, Janina Jacke, Markus Gärtner<sup>1</sup>

## Text- und Data-Mining

### Urheberrechtliche Grenzen der Nachnutzung wissenschaftlicher Korpora und ihre Bedeutung für die Digital Humanities

*Der Beitrag geht der Frage nach, ob und inwieweit im Rahmen von Text- und Data-Mining erstellte Korpora, die geschützte Gegenstände enthalten, unter Berücksichtigung des Urheberrechts nach Abschluss der Forschungsarbeiten aufbewahrt, wissenschaftlich überprüft und Dritten zu Zwecken einer Nachnutzung zugänglich gemacht werden dürfen. Das führt zu der weiteren Frage, ob die Anforderungen der guten wissenschaftlichen Praxis angesichts urheberrechtlicher Beschränkungen eingehalten werden können. Vorgestellt wird ein Konzept, das es von rechtlicher Seite ermöglichen soll, Dritten Auszüge der Korpora zugänglich zu machen und das von technischer Seite dessen Realisierbarkeit im Bereich der Digital Humanities thematisiert. Damit werden die Nachnutzungsmöglichkeiten der Korpora aus ökonomischen wie aus wissenschaftspolitischen Gründen ausgeweitet.*

## I. Einleitung

### 1. Status quo im Urheberrecht

Im Rahmen des Text- und Data-Mining (TDM) werden große Text- und Datenmengen automatisiert analysiert, um Muster, Strukturen oder Zusammenhänge aufzudecken. Dieses Verständnis liegt sowohl der urheberrechtlichen Schranke im deutschen Recht, § 60d UrhG,<sup>2</sup> als auch der Richtlinie über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt vom 17. April 2019 (RL 2019/790 (DSM-RL)) zugrunde, Art. 2 Nr. 2 DSM-RL. Im Zuge der Umsetzung der DSM-RL erfährt das geltende Recht Änderungen, die die am 3.2.2021 als Regierungsentwurf eingebracht wurden.

In der juristischen Literatur herrscht mittlerweile Einigkeit darüber, dass die digitale Analyse als solche zumeist keine Urheber- oder verwandten Schutzrechte verletzt.<sup>3</sup> Anders verhält es sich mit den Vorbereitungshandlungen. Diese beinhalten die Erstellung sog. Korpora. Das sind Zusammenstellungen der für die Forschungsfrage relevanten und möglicherweise urheberrechtlich

---

<sup>1</sup> Felicitas Kleinkopf ist wissenschaftliche Mitarbeiterin und Doktorandin am Zentrum für Angewandte Rechtswissenschaft, Institut für Informations- und Wirtschaftsrecht, Karlsruher Institut für Technologie.

Dr. phil. Janina Jacke ist wissenschaftliche Mitarbeiterin am Institut für Sprach- und Literaturwissenschaft der TU Darmstadt und am Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart.

Dipl.-Inf. Markus Gärtner ist wissenschaftlicher Mitarbeiter und Doktorand am Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart. Die Arbeit entstand im Rahmen des interdisziplinären Forschungsprojekts "XSample - Text und Data Mining auf geschützten Werken durch Auszüge transparent erschließen". XSample ist ein Kooperationsprojekt zwischen der Universitätsbibliothek Stuttgart, dem Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart und dem Zentrum für Angewandte Rechtswissenschaft/Institut für Informations- und Wirtschaftsrecht des Karlsruher Instituts für Technologie. Es ist Teil des Förderprogramms „Wissenschaftliche Bibliotheken gestalten den digitalen Wandel (BW-BigDIWA)“ des Ministeriums für Wissenschaft, Forschung und Kunst Baden-Württemberg. Der Artikel gibt die persönliche Auffassung der Verfasserinnen und des Verfassers wieder.

<sup>2</sup> BT-Drs. 18/12329, S. 40.

<sup>3</sup> Richtlinie (EU) 2019/790 (DSM-RL), Erwägungsgrund 9; BT-Drs. 18/12329, S. 40; Stieper in: Schricker/Loewenheim, UrhG, 6. Aufl. 2020, § 60d Rn. 3; Dreier in: Dreier/Schulze, UrhG, 6. Aufl. 2018, § 60d Rn. 4; Raue, ZUM 2019, 684 (685).

geschützten Werke. Zur Korpuserstellung sind umfangreiche Vorverarbeitungen notwendig, die je nach Forschungsgebiet und Forschungsfrage variieren können. Relevant sind etwa in den Digital Humanities, die Digitalisierung analoger Forschungsmaterialien, die Texterkennung, die Segmentierung von Fließtexten in Tokens sowie die für die spätere Analyse relevante Annotation besagter Tokens.<sup>4</sup> Die segmentierten und annotierten Fassungen der Forschungsgegenstände bilden dann das Korpus, das anschließend algorithmisch analysiert wird. Im Zuge dieser Vorverarbeitung werden urheberrechtliche Schutzgegenstände vervielfältigt und öffentlich zugänglich gemacht, §§ 16, 19a UrhG. Diese Verwertungshandlungen sind der nicht-kommerziellen wissenschaftlichen Forschung<sup>5</sup> nun durch die Schranken der § 60d UrhG und Art. 3 DSM-RL erlaubt, solange rechtmäßiger Zugang zu den geschützten Forschungsobjekten besteht.<sup>6</sup>

Was das Schicksal der Korpora nach Projektabschluss betrifft, leidet die Forschung aber weiterhin unter erheblichen Rechtsunsicherheiten: Um zu vermeiden, dass "parallele Artikeldatenbanken" geschützter Güter entstehen, sieht § 60d Abs. 3 UrhG vor, dass die Korpora nach Projektabschluss zu löschen sind.<sup>7</sup> Alternativ können sie in bestimmten Institutionen archiviert werden. Diese Regelung wird in Umsetzung der DSM-RL voraussichtlich entfallen, vgl. § 60d Abs. 5 UrhG-E. Ob und wie Forschungsergebnisse anhand der Korpora wissenschaftlich überprüft werden dürfen, ist aber nicht abschließend geklärt. Auch zur Verfügbarkeit der Korpora für Anschlussforschungen wurde bislang keine ausdrückliche Regelung getroffen.

## 2. Auswirkungen der Rechtslage auf die Forschungspraxis

Empirische Studien legen den Schluss nahe, dass sich eine restriktive rechtliche Zulassung von TDM-Forschung negativ auf TDM-Aktivitäten auswirkt.<sup>8</sup> Das könnte aber wohl nicht nur das Fehlen urheberrechtlicher Erlaubnisse als solche betreffen, sondern auch einen nicht den Bedürfnissen der Wissenschaft entsprechenden Umgang mit den Korpora nach Beendigung der Forschungsarbeiten. Die Erfahrung aus den Digital Humanities bestätigt diese Annahme: Es wurde (und wird auch teilweise weiterhin trotz der Einführung des § 60d UrhG bzw. der DSM-RL) ausschließlich an gemeinfreien

---

<sup>4</sup> Genauer zu den Abläufen in Digital Humanities Reiter, Pichler, Kuhn (Hrsg.), Reflektierte algorithmische Textanalyse, 2020. Näher zu den Unterschieden *Bläß*, Korpusbildung, in: forTEXT. Literatur digital erforschen, abrufbar unter: <https://fortext.net/routinen/methoden/korpusbildung> (Zugriff: 6.8.2020).

<sup>5</sup> Nach der DSM-RL muss auch eine nicht zweckgebundene Schranke eingeführt werden (Art. 4 DSM-RL), im deutschen UrhG wird das voraussichtlich in einem neuen § 44b UrhG erfolgen.

<sup>6</sup> Verwiesen sei in Bezug auf das UrhWissG auf *Raue*, CR 2017, 656; *Schack*, ZUM 2017, 802; *Specht*, OdW 2018, 285; *Spindler*, ZGE 2018, 273; in Bezug auf die DSM-RL auf *Dreier*, GRUR 2019, 771 (772); *Dreier*, ZUM 2019, 384 (387 f.) *Geiger/Frosio/Bulayenko*, IIC 2018, 814; *Raue*, GRUR 2017, 11; *Raue*, ZUM 2019, 684; *Wandtke*, NJW 2019, 1841 (1842) sowie allgemein auf *Dreier* in: *Dreier/Schulze*, UrhG, 6. Aufl. 2018, § 60d und *Stieper* in: *Schricker/Loewenheim*, UrhG, 6. Aufl. 2020, § 60d. Zu den Details der geplanten Neuregelung s. *Raue*, ZUM 2020, 172 und die Gesetzesentwürfe, abrufbar unter:

[https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/DE/Gesetz\\_I\\_Anpassung-Urheberrecht-dig-Binnenmarkt.html?nn=6712350](https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/DE/Gesetz_I_Anpassung-Urheberrecht-dig-Binnenmarkt.html?nn=6712350) (Zugriff: 9.9.2020). Grundlegend zu der urheberrechtlichen Einordnung der einzelnen Voverarbeitungsschritte neben den o. g. auch *Spindler*, GRUR 2016, 1112.

<sup>7</sup> BT-Drs. 18/12329, S. 41.

<sup>8</sup> *Handke/Guibault/Vallbb*, in: *Schmidt/Dobrev*a (Hrsg.), New avenues for electronic publishing in the age of infinite collections and citizen science, 2015, S. 127 ff.; *Haucap/Loebert/Spindler/Thorwarth*, Ökonomische Auswirkungen einer Bildungs- und Wissenschaftsschranke im Urheberrecht, 2016, S. 45 f.

Werken geforscht.<sup>9</sup> Im Bereich der digitalen Literaturwissenschaft wird beispielsweise jüngere, zeitgenössische Literatur meist aus arbeitsökonomischen sowie aus strategischen Gründen nicht mit digitalen Methoden beforscht: Bei der Forschung an urheberrechtlich geschützten Werken ist oft die Digitalisierung der Texte notwendig, was erheblichen zusätzlichen Arbeitsaufwand bedeutet. Dadurch, dass auch die Publikationsmöglichkeiten der Forschungsergebnisse eingeschränkt sind, sinken zudem die Förderchancen für die entsprechenden Projekte. Diese Faktoren tragen dazu bei, dass sich Projektkonzeptionen in der digitalen Literaturwissenschaft oft an pragmatischen Überlegungen orientieren statt an wissenschaftlicher Relevanz und sich i. d. R. auf die Forschung an älteren, gemeinfreien Werken und die in ihnen auftretenden Phänomene beschränken. Das birgt auch den Vorteil, dass die erstellten Korpora projektübergreifend genutzt werden dürfen. Korpora, die online verfügbar sind,<sup>10</sup> bestehen deswegen bislang ausschließlich aus gemeinfreien Werken.<sup>11</sup>

### 3. Anforderungen an eine forschungsfreundliche Rechtslage

Der Beitrag wird der Frage nachgehen, ob der Umgang mit den Korpora im Urheberrecht zufriedenstellend geregelt ist und ob er insbesondere den Grundsätzen der guten wissenschaftlichen Praxis gerecht wird. Nur in diesem Fall kann das Recht das von der DSM-RL angestrebte Ziel einer rechtssicheren und forschungsfreundlichen Regelung<sup>12</sup> erreichen. Zur Veranschaulichung der Problemstellung wird eine vorstellbare Forschungsfrage im Bereich des TDM dargelegt (II.). Die urheberrechtlichen Rahmenbedingungen werden anschließend analysiert (III.). Dabei ist zu diskutieren, wie die Korpora rechtlich einzuordnen sind (III. 1.), was das geltende und künftige Recht bzgl. ihrer Aufbewahrung vorgibt, wie es nachgebessert werden könnte (III. 2.) und inwiefern wissenschaftliche Überprüfbarkeit gewährleistet wird (III. 3.). Zuletzt soll untersucht werden, welche rechtlichen Möglichkeiten es hinsichtlich der Weitergabe von Auszügen der Korpora an Dritte gibt und ob diese Vorgehensweise in den Digital Humanities inhaltlich sinnvoll und technisch umsetzbar wäre (III. 4.).

## II. Geisteswissenschaftliche Forschungsfragen der Digital Humanities

Exemplarisch sei die konzeptionelle Ausgestaltung eines literaturwissenschaftlichen Anwendungsfalls umrissen, wie er in den Digital Humanities vorstellbar ist.<sup>13</sup> Gegenstand ist im Fallbeispiel die Analyse unzuverlässig erzählter fiktionaler Texte, d. h. solcher Texte, deren Erzähler teilweise Unwahres über die erzählte Welt der Geschichte behauptet.<sup>14</sup> Dieses Phänomen tritt vornehmlich in urheberrechtlich geschützter Literatur aus dem 20. und 21. Jahrhundert auf<sup>15</sup> und ist ein beliebtes

---

<sup>9</sup> Z. B. in den Projekten des *Trier Center for Digital Humanities*, <https://kompetenzzentrum.uni-trier.de/de/projekte/projekte/> (Zugriff: 9.9.2020).

<sup>10</sup> Z. B. das Deutsche Romankorpus, abrufbar unter: <https://gitlab2.informatik.uni-wuerzburg.de/kallimachos/DROC-Release> (Zugriff: 9.9.2020); das Werkzeug Katharisis, abrufbar unter: <http://lauchblatt.github.io/QuantitativeDramenanalyseDH2015/index.html> (Zugriff: 9.9.2020); weitreichend auch <https://www.clarin-d.net/de/sprachressourcen-und-dienste/korpora> (Zugriff: 9.9.2020).

<sup>11</sup> Je nachdem, ob die Korpora selbst geschützt sind, kommt auch aktuell freilich nur eine Veröffentlichung in Betracht, der die Inhaber dieser Rechte zugestimmt haben, insb. Open Access.

<sup>12</sup> Richtlinie (EU) 2019/790 (DSM-RL), Erwägungsgrund 10.

<sup>13</sup> Dem liegt die Konzeption des ersten Use Cases aus "XSample" zugrunde, s. o. Fußn. 1.

<sup>14</sup> *Shen*, Unreliability, in: Hühn, Peter et al. (Hrsg.), *the living handbook of narratology*, abrufbar unter: <https://www.lhn.uni-hamburg.de/node/66.html> (Zugriff: 9.9.2020).

<sup>15</sup> *Kindt*, *Unzuverlässiges Erzählen und literarische Moderne*, Tübingen 2008.

literaturwissenschaftliches Forschungsfeld.<sup>16</sup> Das Phänomen ist als Use Case für Forschungen im Bereich des TDM besonders interessant: Für das Vorliegen unzuverlässigen Erzählens gibt es auf der einen Seite eine Reihe textueller bzw. sprachlicher/stilistischer Indikatoren, deren Vorliegen sich algorithmusbasiert feststellen lässt, etwa durch Machine-Learning-Verfahren, die auf handannotierten Daten trainiert wurden.<sup>17</sup> Auf der anderen Seite gilt die tatsächliche Entscheidung über das Vorliegen unzuverlässigen Erzählens als stark interpretativ.<sup>18</sup>

Anhand dieser Eigenschaften werden die aufwendige gedankliche Konzeption und der daraus resultierende Zeitaufwand ersichtlich, der in die Erstellung der Korpora fließt. Wenn diese für wissenschaftliche Nachnutzungen zur Verfügung stünden, müsste der Aufwand nicht doppelt betrieben werden.

### **III. Rechtlicher Rahmen einer Verfügbarkeit der Korpora nach Forschungsabschluss**

#### **1. Rechte an den Korpora**

Die Korpora können selbst Schutzgegenstand des UrhG sein. Dabei kommt einerseits ein Datenbankherstellerrecht nach §§ 87a ff. UrhG, andererseits sogar ein Datenbank-Werkschutz gem. §§ 4 Abs. 2, 2 Abs. 2 UrhG in Betracht. Letzterer liegt, jedenfalls in den Digital Humanities, aufgrund der individuellen Konzeption der Korpora nahe. I.d.R. bestehen der Schutz der in den Korpora enthaltenen Werke und derjenige der Korpora selbst parallel zueinander. Je nach Art und Umfang der Vorverarbeitungen ist allerdings an (unfreie) Bearbeitungen oder freie Benutzungen zu denken, die aktuell ebenfalls rechtlichen Änderungen unterliegen.<sup>19</sup> Zu erwähnen sind § 23 S. 3 UrhG bzw. § 23 Abs. 3 UrhG-E, die jedenfalls ausschließlich technisch bedingte Änderungen freistellen.

#### **2. Aufbewahrung der Korpora**

Eine erhebliche Einschränkung der Forschungspraxis bildet die Rechtslage bzgl. der Aufbewahrung der Korpora. Noch ist vorgesehen, dass diese nach Abschluss der Forschungsarbeiten gelöscht werden müssen, alternativ können sie auch in bestimmten Institutionen archiviert werden, §§ 60d Abs. 3, 60e, 60f UrhG. In Umsetzung der DSM-RL wird das nationale Recht voraussichtlich zugunsten nicht-kommerziell tätiger Wissenschaftsinstitutionen (denen nach § 60d Abs. 2 S. 1, Abs. 3 Nr. 1 UrhG-E) von einer Löschungspflicht Abstand nehmen. Stattdessen wird es diesen erlauben, die Korpora selbst aufzubewahren, solange sie für Zwecke der wissenschaftlichen Forschung oder zur Überprüfung wissenschaftlicher Erkenntnisse notwendig sind (§§ 60d Abs. 5 UrhG-E).

Das ist als äußerst sinnvoll einzustufen: Die Grundsätze guter wissenschaftlicher Praxis gebieten gerade, dass Forschungsdaten für mindestens zehn Jahre verfügbar sein sollen, sodass die

---

<sup>16</sup> Zuletzt Kindt/Aumüller (Hrsg.), Narrative Unreliability: Scope and Limits, Journal of Literary Theory 2018, Vol. 12 Issue 1.

<sup>17</sup> Nünning, in: Nünning/Surkamp/Zerweck (Hrsg.), Unreliable Narration. Studien zur Theorie und Praxis unglaubwürdigen Erzählens, 1998, S. 3 – 39.

<sup>18</sup> Yacobi, Poetics Today 1981, Vol. 2 Issue 2, 113.

<sup>19</sup> Vgl. EuGH GRUR 2019, 929 ff - Pelham; sowie den Regierungsentwurf.

Forschungsergebnisse nachgeprüft werden können.<sup>20</sup> Die Speicherung von Forschungsdaten ist also essentiell, um die Validität der Forschungsergebnisse zu sichern. Bisher war man sich zudem in der Praxis unsicher, zu welchem Zeitpunkt die Löschungs- und die Übermittlungspflicht eingreifen.<sup>21</sup>

Ein effektives Forschungsdatenmanagement und die gute wissenschaftliche Praxis erfordern aber nicht nur die Aufbewahrung, sondern auch die Wiederauffindbarkeit der Forschungsdaten. Diese Transparenz wird unter die sog. FAIR-Prinzipien (Findable – Accessible – Interoperable – Re-Usable)<sup>22</sup> gefasst. Deren Einhaltung ist nicht gewährleistet, wenn die Korpora durch die Forschenden selbst aufbewahrt werden.<sup>23</sup> Zudem erfordert das öffentliche Interesse an Forschungsdaten, die aus mit Steuergeldern finanzierter Forschung stammen, die zentrale Speicherung.

Eine zentrale Speicherung sichert darüber hinaus einen effektiveren Schutz der Korpora vor unbefugten Zugriffen, als es einzelne Institutionen können. Die Beibehaltung der Archivierungsmöglichkeit ist auch aus Gründen der Rechtssicherheit zu bevorzugen, schließlich wurde diese Regelung erst zum 1.3.2018 eingeführt.

Um den Befürchtungen des Entstehens "paralleler Artikeldatenbanken"<sup>24</sup> entgegenzuwirken, die den Primärmarkt gefährden, sollten die eigene Aufbewahrung und die Archivierung weiterhin in einem Alternativverhältnis stehen. So existiert insgesamt eine geringere Anzahl an Vervielfältigungen.

Demzufolge ist es notwendig, dass künftig weiterhin die Option besteht, die TDM-Korpora an zentrale Stellen weiterzugeben. Eine Beibehaltung der Archivierung nach Vorbild des (noch) aktuellen § 60d Abs. 3 UrhG lässt auch das Unionsrecht zu.<sup>25</sup>

### 3. Überprüfung wissenschaftlicher Forschung

Die Überprüfbarkeit ist wesentlicher Bestandteil der Forschung und gewährleistet gerade ihre Validität. Auch die Gesetzesbegründung zum Urheberwissenschaftsgesetz (UrhWissG) v. 15.5.2017 sieht vor, dass die Korpora in Gänze verfügbar bleiben sollen, um Überprüfungen zu erlauben.<sup>26</sup> Es ist aber denkbar, dass in diesem Zuge urheberrechtliche Verwertungshandlungen erfolgen. Die entsprechenden Vorschriften, § 60d Abs. 1 Nr. 2 UrhG sowie § 60d Abs. 4 UrhG-E, enthalten zwar explizit die Überprüfbarkeit, sie beziehen sich aber nur auf Peer Review Verfahren, die entweder vor Abschluss der Forschungsarbeiten oder unmittelbar im Anschluss daran stattfinden.<sup>27</sup> Sie erfassen aber

---

<sup>20</sup> Vgl. etwa Leitlinie 17 der Leitlinien der DFG zur Sicherung der guten wissenschaftlichen Praxis, abrufbar unter: [https://www.dfg.de/download/pdf/foerderung/rechtliche\\_rahmenbedingungen/gute\\_wissenschaftliche\\_praxis/kodex\\_gwp.pdf](https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf), S. 22 (Zugriff: 9.9.2020).

<sup>21</sup> Rechtlich ist die Entscheidung, ob eine Forschungsarbeit abgeschlossen ist, unter die Forschungsfreiheit zu fassen und unterliegt wohl nur einer Missbrauchskontrolle, *Raue*, CR 2017, 656 (659), *Spindler*, ZGE 2018, 273 (284); *Dreier* in: *Dreier/Schulze*, UrhG, 6. Aufl. 2018, § 60d Rn. 12.

<sup>22</sup> Vgl. etwa Leitlinie 13 der Leitlinien der DFG zur Sicherung der guten wissenschaftlichen Praxis, abrufbar unter: [https://www.dfg.de/download/pdf/foerderung/rechtliche\\_rahmenbedingungen/gute\\_wissenschaftliche\\_praxis/kodex\\_gwp.pdf](https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf), S. 19 (Zugriff: 9.9.2020).

<sup>23</sup> AA bzgl. § 60d Abs. 5 UrhG-E dagegen erst kürzlich *Wirth*, ZUM 2020, 585 (590), der die gute wissenschaftliche Praxis schon dadurch gewährleistet sieht, dass die Korpora nach § 60d Abs. 5 UrhG-E aufbewahrt werden dürfen.

<sup>24</sup> BT-Drs. 18/12329, S. 41.

<sup>25</sup> Richtlinie (EU) 2019/790 (DSM-RL), Erwägungsgrund 15.

<sup>26</sup> BT-Drs. 18/12329, S. 41.

<sup>27</sup> *Raue*, CR 2017, 656 (659); *Spindler*, ZGE 2018, 273 (284); BT-Drs. 18/12329, S. 41.

auf Grundlage ihrer systematischen Stellung innerhalb der Normen keine zeitlich nachfolgenden Vorgänge. Im (noch) geltenden Recht können Forschende die Korpora auch nicht zugänglich machen, weil sie diese nicht (mehr) haben. Der Regierungsentwurf erwägt zudem ausdrücklich, dass die öffentliche Zugänglichmachung nach Abschluss der Forschungsarbeiten zu beenden ist.<sup>28</sup> Im Umkehrschluss ist nachfolgend keine Überprüfung mehr möglich.

Dabei wird die Überprüfung wissenschaftlicher Forschung, jedenfalls im deutschen Recht, sogar unter die Freiheit von Wissenschaft und Forschung gefasst, Art. 5 Abs. 3 S. 1 GG,<sup>29</sup> und sollte dementsprechend privilegiert werden. Es erschiene zudem widersprüchlich, wenn der Gesetzgeber einerseits eine Aufbewahrungsmöglichkeit vorsähe, die gerade den Zweck einer Überprüfbarkeit und Referenzierbarkeit hat,<sup>30</sup> die Forschungsdaten aber andererseits nicht zur Überprüfung wissenschaftlicher Forschung verwendet werden dürften.<sup>31</sup>

Für die Überprüfung wissenschaftlicher Forschung ist, anders als für Anschlussforschungen, auch keine anteilige Zugänglichmachung der Korpora ausreichend, wie es auf Grundlage des § 60c UrhG erfolgen könnte, auf den der Regierungsentwurf auch verweist.<sup>32</sup> Es muss auch nach Beendigung der Forschungsarbeiten zu Zwecken der Überprüfung Zugriff auf das gesamte Korpus gewährt werden dürfen. Infolgedessen muss § 60d Abs. 3 UrhG bereits nach aktueller Rechtslage dergestalt ausgelegt werden, dass entsprechende Verwertungshandlungen erlaubt sind.<sup>33</sup> Dies wird nach Maßgabe des Erwägungsgrundes 15 der DSM-RL und § 60d Abs. 5 UrhG-E auch künftig der Fall sein. Alternativ könnte § 60d Abs. 4 UrhG-E extensiv ausgelegt werden, die o. g. systematischen Gründe sprechen aber dagegen.

#### **4. Verfügbarkeit der Korpora für Anschlussforschungen**

Auch zur Verfügbarmachung der Korpora an Dritte, die ebenfalls Forschungszwecke verfolgen, bedarf es einer urheberrechtlichen Schranke, denn eine Übermittlung von Auszügen der Korpora beinhaltet Vervielfältigungen (§ 16 UrhG) und u. U. öffentliche Zugänglichmachungen (§ 19a UrhG) geschützter Anteile jener Werke, die im Korpus enthalten sind. Es können auch die Rechte an den Korpora selbst berührt sein.

De lege lata ist eine Nachnutzbarkeit weder auf europäischer noch auf deutscher Ebene ausdrücklich vorgesehen. Sie ist auch in der juristischen Literatur unklar,<sup>34</sup> obwohl sie mehrfach gefordert oder zumindest für wichtig befunden wurde.<sup>35</sup> Erwogen wird u. A. eine Annex-Kompetenz i. R. d.

---

<sup>28</sup> BT-Drs. 12/4022, S. 112.

<sup>29</sup> U. A. Gärditz in: Maunz/Dürig, GG, 90. Aufl. 2020, Art. 5 Abs. 3 Rn. 94 mwN. Auf Unionsebene ist Art. 13 GRCh betroffen, der bislang nicht umfassend vom EuGH thematisiert wurde.

<sup>30</sup> BT-Drs. 18/12329, S. 41.

<sup>31</sup> Ähnlich Raue, CR 2017, 656 (661).

<sup>32</sup> BT-Drs. 12/4022, S. 112.

<sup>33</sup> AA Wirth, ZUM 2020, 585 (587), der die Überprüfbarkeit schon aus der Öffentlichkeit der Einrichtungen ableitet.

<sup>34</sup> De la Durantaye/Raue, RuZ 2020, 83 (94).

<sup>35</sup> Vgl. etwa Raue, GRUR 2017, 11 (15); Raue, CR 2017, 656 (660 f.); Spindler, ZGE 2018, 273 (285),

§ 60d Abs. 3 UrhG bzw. § 60d Abs. 5 UrhG-E.<sup>36</sup> Auch wird eine Heranziehung der Kompetenzen nach §§ 60e, f UrhG diskutiert.<sup>37</sup> Es werden aber auch technische Lösungswege gesucht, um sich des Schutzes der untersuchten Gegenstände zu entledigen.<sup>38</sup>

Der betreffende Erwägungsgrund 15 der DSM-RL verweist in Bezug auf Nutzungen außerhalb des TDM auf Art. 5 Abs. 3 lit. a der Richtlinie zum Urheberrecht in der Informationsgesellschaft (RL 2001/29/EG (InfoSoc-RL)). Im deutschen Recht ist diese Norm in § 60c UrhG umgesetzt. § 60c UrhG erlaubt der nicht-kommerziellen wissenschaftlichen Forschung, einen Anteil von 15 Prozent eines Werkes oder einer Datenbank, §§ 87 a ff. UrhG, § 87c Abs. 1 Nr. 2 UrhG, zu vervielfältigen oder öffentlich zugänglich zu machen. Auf Grundlage dessen sollte § 60c UrhG als Möglichkeit herangezogen werden, Auszüge der Korpora für Anschlussnutzungen zu erstellen und herauszugeben.<sup>39</sup>

Die Möglichkeit, Schranken miteinander zu kombinieren, wurde schon in der Vergangenheit durch den *EuGH* thematisiert und ist nach dessen Ansicht möglich, wenn die Anforderung der jeweiligen Bestimmung erfüllt sind.<sup>40</sup> Die Kumulation zweier Schranken ist von der DSM-RL in Erwägungsgrund 15 auch explizit bezweckt. Diesem Ansatz steht auch keine Subsidiarität der InfoSoc-RL gegenüber der DSM-RL entgegen. Letztere sieht mit Art. 25 gerade vor, dass weiterreichende Regelungen auf Grundlage der InfoSoc-RL erlassen werden dürfen. Zudem ergäbe gerade der Verweis in Erwägungsgrund 15 der DSM-RL andernfalls keinen Sinn.

Dem dargestellten Ansatz dürfte auch nicht der in Art. 5 Abs. 5 InfoSoc-RL kodifizierte Drei-Stufen-Test entgegenstehen, der nicht nur Grundlage für die Normierung urheberrechtlicher Schranken, sondern im Rahmen der Unionsrechtskonformität auch relevant für die richterliche Rechtsauslegung ist.<sup>41</sup> Er kann der an dieser Stelle nur in seinen Grundzügen geprüft werden.

Der Drei-Stufen-Test besagt, dass urheberrechtliche Schranken auf bestimmte Sonderfälle begrenzt sein müssen (Stufe eins), sie die normale Verwertung des Werks nicht beeinträchtigen dürfen (Stufe zwei) und sie die berechtigten Interessen des Urhebers nicht unzumutbar verletzen dürfen (Stufe drei). Ungeachtet der Frage, welche Bedeutung der ersten Stufe im Kontext der InfoSoc-RL zukommt,<sup>42</sup> muss jedenfalls die wissenschaftliche Nachnutzung von TDM-Korpora als ein derartiger Sonderfall angesehen werden. Sofern nur Anteile der Korpora nachgenutzt werden, ist auch nicht der Primärmarkt betroffen.

---

<sup>36</sup> So bzgl. des UrhWissG *Spindler*, ZGE 2018, 273 (285); befürwortend bzgl. § 60d Abs. 3 UrhG auch *Raue*, GRUR CR 2017, 656 (661); skeptisch die Allianz der deutschen Wissenschaftsorganisationen, abrufbar unter: [https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Stellungnahmen/2019/Downloads/090519\\_Stellungnahme\\_Allianz\\_EU-Richtlinien\\_Urheberrecht.pdf?\\_\\_blob=publicationFile&v=2](https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Stellungnahmen/2019/Downloads/090519_Stellungnahme_Allianz_EU-Richtlinien_Urheberrecht.pdf?__blob=publicationFile&v=2), S. 3 (Zugriff: 9.9.2020).

<sup>37</sup> *Spindler*, ZGE 2018, 273 (285).

<sup>38</sup> Verwiesen sei damit u. A. auf das *Trier Center for Digital Humanities*, dessen Forschung vorwiegend an gemeinfreien Werken stattfindet oder den Urheberrechtsschutz mittels technischer Methoden beseitigen möchte, vgl. <https://kompetenzzentrum.uni-trier.de/de/projekte/projekte/> (Zugriff: 9.9.2020); s. auch den Bericht über einen DFG-Workshop zu diesem Thema, *Erlor*, RuZ 2020, 108.

<sup>39</sup> Befürwortend in Bezug auf Art. 5 Abs. 3 lit. a InfoSoc-RL auch *Raue*, ZUM 2019, 684 (688).

<sup>40</sup> *EuGH* GRUR 2014, 1078 (1081) - Eugen Ulmer.

<sup>41</sup> *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, Tübingen 2009, S. 73 m. w. N.; *BGH* GRUR 2002, 963 (967 ff.) - Elektronischer Pressespiegel.

<sup>42</sup> *Dreier* in: *Dreier/Schulze*, UrhG, 6. Aufl. 2018, Vor § 44a Rn. 21; *Stieper* in: *Schricker/Loewenheim*, UrhG, 6. Aufl. 2020, Vor § 44a Rn. 30; näher *Senftleben*, GRUR Int 2004, 200 (206).

Die Anforderungen an eine Unzumutbarkeit für die Rechteinhaber sind tendenziell als hoch anzusehen.<sup>43</sup> Berechtigte Interessen an den TDM-Korpora bestehen von Seiten der wissenschaftlichen Forschung, diese basieren auf der Wissenschaftsfreiheit, Art. 13 GRCh. Demgegenüber sind die Urheber der in den Korpora enthaltenen geschützten Werke zu berücksichtigen, die durch ihre Normierung in Art. 17 Abs. 2 GRCh ebenfalls erhebliche Bedeutung haben. Dazu zählen auch die Inhaber der Rechte an den Korpora als solche.

Zugunsten der Nachnutzung kann die Vergütungspflicht des § 60c UrhG, § 60h UrhG, berücksichtigt werden:<sup>44</sup> Zwar ist das TDM selbst (jedenfalls künftig)<sup>45</sup> entschädigungslos hinzunehmen.<sup>46</sup> Doch erfordert die Schranke des § 60d UrhG in diesem Kontext den rechtmäßigen Zugang zu geschützten Ausgangsmaterialien, wodurch den Rechteinhabern mittelbare Einkünfte zufließen. Bei einer Nachnutzung soll dieser Zugang gerade erst gewährt werden. Über diese Divergenz kann eine Vergütung hinweghelfen.

Auf dieser Grundlage ist eine Nachnutzbarkeit der TDM-Korpora jedenfalls nicht als gänzlich unzumutbar für die Rechteinhaber einzustufen.

Auch aus urheberrechtsdogmatisch ist eine derartige Auslegung denkbar: In den letzten Jahren macht sich erfreulicherweise eine Entwicklung dahingehend erkennbar, dass man sich von einer grenzenlosen Ausweitung von Schutzrechten wegbewegt und immer häufiger Nutzungsinteressen als legitim erachtet. Auf ein maximales Schutzniveau zielte noch die InfoSoc-RL ab, die nunmehr beinahe 20 Jahre alt ist. Diese sieht in ihrem umfangreichen Kanon nur eine einzige verpflichtende Ausnahme oder Beschränkung vor. Darüber hinaus wurde der erforderliche Grad an Individualität, der für die Anerkennung eines Werks notwendig ist, auf Werkteile von elf Wörtern abgesenkt.<sup>47</sup>

Demgegenüber gestaltet Art. 3 DSM-RL das TDM sogar verpflichtend aus. Auch ist eine Tendenz erkennbar, die sich von der traditionell engen Schrankenauslegung entfernt:<sup>48</sup> Trotz des abschließenden Charakters des unionsrechtlichen Schrankenkanons sowie einer Beschränkung der Grundrechtsprüfung auf den Dreistufentest gehen mittlerweile nicht nur Stimmen der Literatur,<sup>49</sup> sondern jüngst auch der *EuGH* von einer Auslegung der Schrankenbestimmungen nach ihrem Zweck aus.<sup>50</sup> Insofern sind insgesamt die Nutzungsinteressen stärker in den Fokus gerückt.

---

<sup>43</sup> *Senftleben*, GRUR Int 2004, 200 (210 f.).

<sup>44</sup> *Stieper* in: Schrickler/Loewenheim, UrhG, 6. Aufl. 2020, Vor § 44a Rn. 31; *EuGH* GRUR 2014, 1078 ff. (1080) Rn. 48 - Eugen Ulmer; vertiefend *Senftleben*, GRUR Int. 2004, 200 (211).

<sup>45</sup> Anders verhält es sich im noch geltenden Recht, § 60h Abs. 1 UrhG.

<sup>46</sup> Richtlinie (EU) 2019/790 (DSM-RL), Erwägungsgrund 17.

<sup>47</sup> *EuGH* GRUR 2009, 104 – Infopaq.

<sup>48</sup> Zur früheren engen Schrankenauslegung u. A. *Stieper* in: Schrickler/Loewenheim, UrhG, 6. Aufl. 2020, Vor §§ 44a ff. UrhG Rn. 36, *BGH* GRUR 1968, 607 (608 f) - Kandinsky; *BGH* GRUR 1972, 614 (616) - Landesversicherungsanstalt; *BGH* GRUR 1985, 874 (875) - Schulfunksendung; *BGH* GRUR 1991, 903 (905) - Liedersammlung; *BGH* GRUR 1983, 562 (563) - Zoll- und Finanzschulen; *BGH* GRUR 2012, 819 (822) Rn. 28 - Blühende Landschaften.

<sup>49</sup> U. A. *Stieper* in: Schrickler/Loewenheim, UrhG, 6. Aufl. 2020, Vor §§ 44a UrhG Rn. 37 mwN; tendenziell *Dreier* in: *Dreier/Schulze*, UrhG, 6. Aufl. 2018, Vor § 44a Rn. 7; *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, 2009, S. 65 ff.

<sup>50</sup> *EuGH* GRUR 2019, 934 (939) Rn. 69 ff - Funke Medien/Bundesrepublik Deutschland [Afghanistan Papiere]; *EuGH* GRUR 2019, 940 (944) Rn. 53 ff - Spiegel Online/Volker Beck [Reformistischer Aufbruch].



Auch rechtspolitisch ist die Nachnutzbarkeit der Korpora erstrebenswert: Die Verfügbarkeit von Forschungsdaten für Anschlussnutzungen ist abseits von Individualinteressen ein anerkannter, nicht wegzudenkender Teil eines Forschungsprozesses, der ebenfalls zur guten wissenschaftlichen Praxis zählt.<sup>51</sup> Die Nachnutzbarkeit ist aber auch aus ökonomischen Gründen geboten: Die Datenaufbereitung nimmt einen großen Teil der Forschungsarbeiten ein und ist demnach besonders kostenintensiv. Diese Kosten treffen i. d. R. die öffentliche Hand. Um einerseits einen möglichst weiten Nutzen aus investierten Geldern zu ziehen und andererseits auch inhaltlich von bereits betriebener Forschung zu profitieren, ist es daher nur effizient, Korpora jedenfalls anteilig zugänglich zu machen. Eine derartige Regelung kann auch Anreize für bislang nicht stattgefundene Forschung setzen und damit entsprechenden Fortschritt begünstigen. Dieser ist im internationalen Vergleich dringend notwendig, um die EU als attraktiven Standort der Digitaltechnik zu positionieren.<sup>52</sup>

Diese Argumentation spricht insgesamt für eine Anwendbarkeit des § 60c UrhG. Im Gegensatz zu den sonstigen Vorschlägen zugunsten einer Nachnutzung (s. o.) ist diese durch den Verweis des Erwägungsgrund 15 der DSM-RL sogar vom Unionsgesetzgeber intendiert. Darüber hinaus verweist auch der Regierungsentwurf darauf, dass Nutzungen zu wissenschaftlichen Zwecken außerhalb des TDM weiterhin unter § 60c UrhG fallen, sofern dessen Voraussetzungen erfüllt sind.<sup>53</sup>

Eine Auszugslösung ist jedenfalls für eine bestimmte Bandbreite an Forschungsfragen der Digital Humanities vorstellbar, etwa für die beschriebene Erforschung „Unzuverlässigen Erzählens“: Das Phänomen wird zwar einerseits gemeinhin als „textglobal“ eingeschätzt, d. h. es kann nur festgestellt werden, wenn der Text in seiner Gesamtheit betrachtet wird. Andererseits gibt es aber typischerweise Schlüsselstellen, in denen die Unzuverlässigkeit aufgedeckt werden kann. Grundsätzlich dürfte also die Weitergabe von Auszügen geeignet sein, relevante Anhaltspunkte für die literaturwissenschaftlichen Forschung zu liefern.

Für die technische Umsetzung müssten Verfahren entwickelt werden, die eine für die jeweiligen Formate der Primärdaten passende Segmentierung enthalten, anhand derer die Auszüge prozentual berechnet und mit der die interessanten Schlüsselstellen automatisch aufgefunden werden können: Da die Zielgruppe der Auszüge Forschende mit einem großen Spektrum unterschiedlichster Interessen umfasst, wäre es nicht zielführend, einen fixen Anteil der Daten als Auszug zu liefern.

Die Größe der Auszüge ist anhand der Seitenzahl des Werks inkl. Einleitung, Vorwort, Inhalts- und Sachverzeichnis zu bestimmen.<sup>54</sup> Die nicht-kommerziellen wissenschaftlichen Zwecke, an die § 60c UrhG gebunden ist, können mittels E-Mail-Adressen oder Formularen verifiziert werden.

---

<sup>51</sup> Vgl. etwa Leitlinie 13 der Leitlinien zur Sicherung guter wissenschaftlicher Praxis der DFG, abrufbar unter: [https://www.dfg.de/download/pdf/foerderung/rechtliche\\_rahmenbedingungen/gute\\_wissenschaftliche\\_praxis/kodex\\_gwp.pdf](https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf), S.18 f. (Zugriff: 9.9.2020); so auch *de la Durantaye/Raue*, RuZ 2020, 83 (90).

<sup>52</sup> Die Sicherung ihrer Wettbewerbsposition war auch eine Intention, die die EU im Rahmen der DSM-RL verfolgte, s. dazu Richtlinie (EU) 2019/790 (DSM-RL), Erwägungsgrund 10. Dazu hat sie auch die EU-Digital-Strategie veröffentlicht, abrufbar unter: [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age\\_de](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_de) (Zugriff: 16.9.2020). Ob die EU sich in der Digitaltechnik im internationalen Vergleich noch zu positionieren vermag, bleibt abzuwarten.

<sup>53</sup> BT-Drs. 12/4022, S. 112.

<sup>54</sup> *Dreier* in: *Dreier/Schulze*, UrhG, 6. Aufl. 2018, § 60c Rn. 7.; *BGH GRUR* 2014, 549 (551) Rn. 24 - Meilensteine der Psychologie. Eine Ausnahme bilden Werke geringen Umfangs oder einzelne Beiträge aus Fachzeitschriften oder wissenschaftlichen Zeitschriften, § 60c Abs. 3 UrhG.

Ausgenommen offensichtlicher Missbrauchsfälle dürfen den Einrichtungen dabei keine näheren Prüfpflichten zukommen, andernfalls würden ihnen unzumutbare Haftungsrisiken auferlegt.<sup>55</sup>

#### **IV. Schlussfolgerungen**

Trotz einer positiv zu bewertenden europäischen Entwicklung von einem urheberzentrierten Recht zu einem Recht, das jedenfalls auch Nutzungsinteressen als förderungswürdig erachtet, weist das derzeitige deutsche Recht in Bezug auf das TDM noch Lücken auf, die auf Grundlage des Europarechts ohne Weiteres gefüllt werden könnten.

Obwohl nach der in diesem Beitrag vertretenen Auffassung wissenschaftliche Überprüfungen auch nach Abschluss der Forschungsarbeiten möglich sind und TDM-Korpora auf Grundlage des § 60c UrhG nachgenutzt werden können, sind gesetzliche Klarstellungen anzuempfehlen. Immerhin bezweckte die EU mit ihrer unionsrechtliche Erlaubnisnorm zugunsten des TDM unter anderem, zu mehr Rechtssicherheit beizutragen.<sup>56</sup> Zudem sollte die Archivierungsmöglichkeit nach Vorbild des § 60d Abs. 3 UrhG im neuen § 60d Abs. 5 UrhG-E beibehalten werden. Auf diese Weise kann ein forschungsfreundlicher und interessengerechter Rechtsrahmen geschaffen werden.

#### **Schnell gelesen...**

- Das aktuelle und geplante Urheberrecht ermöglicht es nach seinem wörtlichen Verständnis nicht, im Bereich von Text- und Data-Mining die Anforderungen der guten wissenschaftlichen Praxis zu befolgen und schränkt etwa die Forschung in den Digital Humanities stark ein.
- Die Möglichkeit, Korpora zentral zu archivieren, ist zur Gewährleistung wissenschaftlicher Transparenz und Validität der Forschung erforderlich und sollte nach Vorbild des aktuellen § 60d Abs. 3 UrhG beibehalten werden.
- Zu Zwecken der Überprüfung wissenschaftlicher Forschung muss auch nach Forschungsabschluss Zugriff auf die gesamten Korpora gewährt werden dürfen.
- Eine Anwendung des § 60c UrhG zur Übermittlung von Auszügen der Korpora zu Zwecken von Anschlussforschungen böte eine adäquate Möglichkeit, langfristigen Nutzen aus Forschungen zu ziehen.
- Die Einführung der urheberrechtlichen Schranken zugunsten des Text- und Data-Mining stellt gleichwohl einen wichtigen Schritt in Richtung eines innovationsfreundlichen Urheberrechts dar.

---

<sup>55</sup> Ähnlich bzgl. § 60e UrhG Dreier in: Dreier/Schulze, UrhG, 6. Aufl. 2018, § 60e Rn. 27.

<sup>56</sup> Richtlinie (EU) 2019/790 (DSM-RL), Erwägungsgründe 10, 11.