

digital | recht

Schriften zum Immaterialgüter-, IT-,
Medien-, Daten- und Wettbewerbsrecht

Benjamin Raue & Christof Schöch (Hrsg.)

Recht und Digital Humanities

Urheberrecht und verwandte Rechtsbereiche bei der
Forschung mit digitalen Textbeständen

Band 23

Benjamin Raue/Christof Schöch
(Hrsg.)

Recht und Digital Humanities

Urheberrecht und verwandte Rechtsbereiche bei der
Forschung mit digitalen Textbeständen

digital | recht

Schriften zum Immaterialgüter-, IT-, Medien-, Daten- und Wettbewerbsrecht

Herausgegeben von Prof. Dr. Maximilian Becker, Prof. Dr. Katharina de la Durantaye, Prof. Dr. Franz Hofmann, Prof. Dr. Ruth Janal, Prof. Dr. Anne Lauber-Rönsberg, Prof. Dr. Benjamin Raue, Prof. Dr. Herbert Zech

Band 23

Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Angaben sind im Internet über <http://dnb.d-nb.de> abrufbar.

Dieses Buch ist ebenfalls als E-Book verfügbar unter: <https://www.epubli.com/shop>

Dieses Werk ist unter der Creative-Commons-Lizenz vom Typ CC BY 4.0 International (Namensnennung) lizenziert: <https://creativecommons.org/licenses/by/4.0/>. Einzelne Beiträge können unter anderer Lizenz veröffentlicht sein. Die Lizenz eines Beitrags ist im jeweiligen Beitrag kenntlich gemacht. Diese ist zu beachten.



Gefördert mit Mitteln der Forschungsinitiative Rheinland-Pfalz.



Rheinland-Pfalz

MINISTERIUM FÜR
WISSENSCHAFT
UND GESUNDHEIT

ISBN: 9783818705503

URN: hbz:385-20241021164116136-4146720-6

DOI: <https://doi.org/10.25353/ubtr-6390-a811-1faf>

© Trier 2024 Benjamin Raue und Christof Schöch

Die Schriftenreihe wird gefördert von der Universität Trier und dem Institut für Recht und Digitalisierung Trier (IRD T).

 UNIVERSITÄT
TRIER

 IRD T Institute for
Digital Law
Trier

Vorwort

Durch den Einsatz digitaler Methoden in den Geisteswissenschaften ergeben sich neue Möglichkeiten der Analyse, Visualisierung und Verknüpfung von Daten. Viele dieser Daten bestehen jedoch aus Texten, Bildern oder Datenbanken, die oft urheberrechtlich geschützt sind. Hier einen Interessenausgleich zwischen den Interessen der Forschenden und denen von Urheberinnen und Urhebern zu schaffen, ist Gegenstand eines fruchtbaren Austauschs zwischen den Digital Humanities und den Rechtswissenschaften an der Universität Trier, deren Ergebnisse wir in diesem Sammelband auch Forschenden anderer Forschungseinrichtungen zugänglich machen wollen.

Dieser Sammelband führt insbesondere die Beiträge zusammen, die aus dem Projekt „Mining and Modeling Text (MiMoText)“ (2019-2023, gefördert mit Mitteln der Forschungsinitiative Rheinland-Pfalz) hervorgegangen sind, das den Bereich der quantitativen Methoden zur Extraktion, Modellierung und Analyse geisteswissenschaftlich relevanter Informationen aus umfangreichen Textsammlungen weiterentwickelt und aus interdisziplinärer (geistes-, informatik- und rechtswissenschaftlicher) Perspektive erforscht hat.

Den Kern des Bands machen Beiträge aus, die im Arbeitsbereich 3 des Projekts in intensiver interdisziplinärer Zusammenarbeit zwischen Forschenden aus den Rechtswissenschaften (Katharina Erler-Fridgen, Karolina Benedyk und Benjamin Raue) sowie den Digital Humanities (Maria Hinzmann und Christof Schöch) entstanden sind. Sie versuchen, typische Konstellationen aus den Digital Humanities mit dem Urheber- und Datenschutzrecht zu lösen. Der Band wurde angereichert um weitere passende Beiträge, u.a. solche, die aus dem interdisziplinären Workshop „Strategien für die Nutzbarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte“ hervorgegangen sind.

Die Beiträge sind auf dem Stand ihrer ursprünglichen Veröffentlichung. Wir danken allen an dem Projekt Beteiligten, insbesondere Katharina Erler-Fridgen, Karolina Benedyk und Maria Hinzmann, sowie Leonard Vierbuchen für seinen Einsatz im Veröffentlichungsprozess.

Trier, 05.11.2024

Benjamin Raue und Christof Schöch

Inhaltsübersicht

<i>Vorwort</i>	III
<i>Inhaltsübersicht</i>	V
Teil 1: Einführung	3
<i>A. Urheberrecht und Zugang in einer digitalen Welt</i>	
Katharina de la Durantaye/Benjamin Raue	3
<i>B. Open Access in der Wissenschaft und im Urheberrecht sowie bei verwandten Schutzrechten</i>	
Karolina Benedyk	19
<i>C. Verfahrensschritte bei dem Einsatz von Text und Data Mining-Verfahren in den Geisteswissenschaften</i>	
Katharina Erler-Fridgen	29
Teil 2: Urheberrechtlicher Schutz der Ausgangsmaterialien	47
<i>A. Kriterien der urheberrechtlichen Schutzfähigkeit von Texten und Sammelwerken</i>	
Katharina Erler-Fridgen	47
<i>B. Der Schutz großer Textbestände nach dem UrbG</i>	
Florian Jotzo	59
<i>C. Die Nutzung wissenschaftlicher Ausgaben für Textanalysen</i>	
Katharina Erler-Fridgen	77
<i>D. Datenbanken als Quelle oder Ergebnis von Textanalysen</i>	
Katharina Erler-Fridgen	85
<i>E. Die internationale Anwendbarkeit des (deutschen) Urheberrechts</i>	
Karolina Benedyk	113
Teil 3: Urheberrechtsfreie Korpora durch abgeleitete Textformate	123
<i>A. Zugang zu großen Textkorpora des 20. und 21. Jahrhunderts mit Hilfe abgeleiteter Textformate</i>	
Benjamin Raue/Christof Schöch	123
<i>B. Nutzarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte</i>	
Karina Grisse	135

<i>C. Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen</i>	
Christof Schöch/Evelyn Gius/Peer Trilcke/Peter Leinen/Fotis Jannidis/Maria Hinzmann/Jörg Röpke	155
Teil 4: Gesetzlich erlaubte Handlungen	203
<i>A. Die Text und Data-Mining-Schranken und ihr Rahmen für Textanalysen in den Digital Humanities</i>	
Katharina Erler-Fridgen	203
<i>B. Text und Data Mining in Einrichtungen des Kulturerbes</i>	
Benjamin Raue	219
<i>C. Das Zitat und dessen Rahmen für Belege bei Textanalysen</i>	
Katharina Erler-Fridgen	237
<i>D. Die Wissenschaftsschranke in den Digital Humanities</i>	
Karolina Benedyk/Katharina Erler-Fridgen	265
Teil 5: Vertraglich erlaubte Handlungen	275
<i>A. Vertragsrechtliche Lösungen</i>	
Karolina Benedyk	275
<i>B. Kollektive Lizenzen mit erweiterter Wirkung</i>	
Karolina Benedyk	287
Teil 6: Weitere Aspekte	297
<i>A. Die Präsentation von Textteilen als Ergänzung von Textanalysen</i>	
Katharina Erler-Fridgen	297
<i>B. Datenschutz bei den Digital Humanities – ein Überblick</i>	
Karolina Benedyk	305
<i>Verzeichnis der Autorinnen und Autoren</i>	VI

Teil 1: Einführung

A. Urheberrecht und Zugang in einer digitalen Welt

- Urheberrechtliche Fragestellungen des Zugangs für Gedächtnisinstitutionen und die Digital Humanities*

Katharina de la Durantaye/Benjamin Raue

Version (2020), CC BY-SA 4.0.

I. Einleitung

Das Urheberrecht gewährt dem Urheber Ausschließlichkeitsrechte an Werken. Technische und werkvermittelnde Leistungen werden durch Leistungsschutzrechte geschützt. Wo diese Rechte anfangen und aufhören, wer ein Werk unter welchen Umständen und auf welche Weise nutzen darf, sind Kernfragen des Urheberrechts. Es muss einen Ausgleich schaffen zwischen Exklusivitäts- und Zugangsinteressen.¹

In der analogen Welt wurden Zugangsfragen im Urheberrecht dennoch selten problematisiert. Der Eigentümer eines Buches konnte es selbstverständlich aufschlagen und lesen, ohne den Urheber um Erlaubnis bitten zu müssen. Der analoge Werkgenuss war urheberrechtlich irrelevant. Das Urheberrecht mussten im Wesentlichen professionelle Verwerter beachten. Private und wissenschaftliche Nutzer konnten es hingegen weitgehend ignorieren.

Die Digitalisierung hat die Verhältnisse grundlegend geändert und den Anwendungsbereich des Urheberrechts erheblich ausgedehnt:² Wer ein digitales Werkstück nutzt, vervielfältigt es in aller Regel jedenfalls im Arbeitsspeicher seines

* Dieser Beitrag wurde zuerst veröffentlicht als *Durantaye/Raue* RuZ 2020, 83-94 unter einer CC BY-SA Lizenz.

¹ Dazu ausführlich: *Peukert*, GRUR-Beilage 2014, 77 ff.

² Dazu etwa: *Hofmann*, ZGE 2016, 482, 498; *Becker*, ZGE 2016, 239, 273; *Raue*, ZGE 2014, 387, 389.

Geräts. Er nimmt damit eine Handlung vor, für die er grundsätzlich der Zustimmung des Urhebers bedarf.³ Urheberrecht ist Alltagsrecht geworden. Das stellt das Urheberrecht stärker als früher vor die Herausforderung, die berechtigten Interessen der Urheber und sonstigen Rechteinhaber mit den Allgemeininteressen auf Informationsfreiheit und Kommunikation in Einklang zu bringen.⁴

Es wundert daher nicht, dass Fragen des Zugangs in den letzten Jahren nicht nur in den Fokus der Wissenschaft, sondern auch der Gesetzgebung und der Rechtsprechung geraten sind.⁵

Diese Zeitschrift wird in ihrem urheberrechtlichen Teil das Spannungsverhältnis zwischen urheberrechtlichen Ausschließlichkeitsrechten einerseits und Zugangsrechten bzw. -interessen andererseits beleuchten und die aktuellen Entwicklungen kritisch begleiten. Ihnen ist der erste Teil dieses Beitrags gewidmet. Weil wir in dieser Zeitschrift auch interdisziplinäre Fragestellungen erörtern wollen, werden im zweiten Teil urheberrechtliche Fragestellungen der Digital Humanities skizziert.

II. Gedächtnisinstitutionen

Digitale Technologien eröffnen Gedächtnisinstitutionen neue Möglichkeiten, um ihren Bestand zu erhalten, zu erweitern und zugänglich zu machen. Nicht all diese Möglichkeiten werden in der Praxis auch genutzt. Neben wirtschaftlichen hat dies vor allem rechtliche Gründe: Digitale Pendant analoger Nutzungen, die zustimmungsfrei zulässig waren, bedürfen zum Teil der Zustimmung des Rechteinhabers.

Das liegt zunächst an der oben bereits genannten Ausweitung des Urheberrechts. Grund ist aber auch der Zuschnitt der gesetzlichen Erlaubnistatbestände. Sie erklären urheberrechtlich relevante Nutzungen unter bestimmten Voraussetzungen ohne Erlaubnis des Rechteinhabers für zulässig. Die Tatbestände entstammen der analogen Welt. Nicht alle passten daher auf digitale Sachverhalte.

³ Zur Frage, ob der Erwerb eines (digitalen) Werkexemplars mit dem Erwerb eines dinglichen Genussrechts verbunden ist, vgl.: *Kuschel*, Der Erwerb digitaler Werkexemplare zur privaten Nutzung, 2019.

⁴ Vgl. nur: EuGH, Urt. v. 8.9.2016 – C-160/15, ECLI:EU:C:2016:644 = GRUR 2016, 1152 Tz. 31 – *GS Media/Sanoma*; EuGH, Urt. v. 8.9.2016 – C-160/15, ECLI:EU:C:2014:1195 = GRUR 2014, 654 Tz. 24 – *PRCA/NLA*; EuGH, Urt. v. 4.10.2011 – C-403/08 und C-429/08, ECLI:EU:C:2011:631 = GRUR 2012, 156 Tz. 164 – *FAPL u. Murphy*; *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 9; *Raue*, ZGE 2017, 514, 515.

⁵ Grundlegend dazu etwa: *Wielsch*, Zugangsregeln, 2008.

Zudem waren die Voraussetzungen und der Anwendungsbereich der Tatbestände oft unklar – alleine schon, weil sie viele unbestimmte Rechtsbegriffe enthielten. Gedächtnisinstitutionen konnten darum oftmals nur schwer einschätzen, welche Nutzungen erlaubnisfrei gestattet waren und welche nicht.⁶

In den letzten Jahren haben sowohl der nationale als auch der europäische und der internationale Gesetzgeber bestehende gesetzliche Erlaubnistatbestände bzw. deren Vorgaben verändert und neue geschaffen, um dieses Problem zu adressieren.⁷ Bislang sind viele Fragen noch nicht oder nicht umfassend erforscht. Dazu gehören Inhalt und Umfang dieser neuen bzw. geänderten Regelungen auf nationaler Ebene, das Verhältnis der europäischen Vorgaben untereinander, die bei der Umsetzung bestehenden Spielräume und die Möglichkeiten des deutschen Gesetzgebers, diese Spielräume auszunutzen.

1. Deutsches Recht

Bereits im Koalitionsvertrag von 2013 hatte die Große Koalition die Schaffung einer Bildungs- und Wissenschaftsschranke in Aussicht gestellt.⁸ Am 1. März 2018 trat schließlich das Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft (UrhWissG) in Kraft.⁹ Es regelt die gesetzlichen Erlaubnistatbestände für Unterricht, Forschung und Gedächtnisinstitutionen neu (§§ 60a-60h UrhG). Ziel war vor allem die Schaffung möglichst konkreter Regelungen, die auf unbestimmte Rechtsbegriffe verzichten und technologieneutral formuliert sind, um so die Rechtssicherheit für die Institutionen zu erhöhen.¹⁰

Inhaltlich war der Spielraum des nationalen Gesetzgebers jedoch durch europäische Vorgaben begrenzt, wie sie insbesondere in der Richtlinie zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte

⁶ Vgl. dazu nur: *de la Durantaye*, Allgemeine Bildungs- und Wissenschaftsschranke, 2014, 274 ff.

⁷ Daneben hat auch der EuGH einige Urteile gefällt, die für Gedächtnisinstitutionen von erheblicher Bedeutung sind, vgl. nur: EuGH, Urt. v. 19.12.2019 – C-263/18, ECLI:EU:C:2019:1111 = GRUR 2020, 179, 179 ff. – *Tom Kabinet* zur Zulässigkeit des Weiterverkaufs „gebrauchter“ E-Books.

⁸ Vgl. nur: Koalitionsvertrag von CDU, CSU und SPD, Deutschlands Zukunft gestalten, S. 134, abrufbar unter: www.cdu.de/artikel/der-koalitionsvertrag-von-cdu-csu-und-spd, zuletzt abgerufen am 27.1.2020.

⁹ Art. 4 UrhWissG, BGBl. 2017 I Nr. 61, 3351.

¹⁰ Vgl.: BT-Drs. 18/12329, 2.

in der Informationsgesellschaft (InfoSoc-RL)¹¹ enthalten sind. Demgemäß hat das UrhWissG die Nutzungsmöglichkeiten eher moderat erweitert.¹² Ohne Vorgänger ist aber die Text und Data Mining-Schranke des § 60d UrhG. Danach darf für die automatisierte Auswertung einer Vielzahl von Texten das Ursprungsmaterial automatisiert und systematisch vervielfältigt und das dabei entstehende Korpus öffentlich zugänglich gemacht werden, sofern gewisse Voraussetzungen erfüllt sind und der Nutzer nichtkommerzielle Zwecke verfolgt.¹³ Allein diese Neustrukturierung und vorsichtige Erweiterung der Erlaubnistatbestände führt für die Gedächtnisinstitutionen zu Erleichterungen. Zudem erweitern die klareren Konturen den realen Handlungsspielraum. Momentan wirken sich diese Vorteile in der Praxis jedoch noch nicht in vollem Maße aus, weil die Rechtssicherheit, die das UrhWissG der Praxis gebracht hat, (noch) zeitlich befristet ist: §§ 60a-60h sind zunächst nur bis Ende Februar 2023 in Kraft; 2022 ist sie zu evaluieren.¹⁴ Größere Investitionen in die neue Rechtslage lohnen sich für Gedächtnisinstitutionen mithin zunächst nicht. Allerdings erfordert die DSM-RL die Verstärkung einiger Schranken (dazu sofort unter 2.a).

2. Europarecht

Am 6. Juni 2019 trat die Richtlinie über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt (DSM-RL)¹⁵ in Kraft. Ihr Inhalt war teilweise hoch umstritten,¹⁶ obwohl die Richtlinie nicht den großen Wurf darstellt, den sich manche erhofft hatten. Für Lehrende, Forschende und Gedächtnisinstitutionen bringt sie aber wertvolle zusätzliche Möglichkeiten, urheberrechtlich geschützte Werke zustimmungsfrei zu nutzen.

¹¹ Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft, ABl. EG 2001 L 167/10.

¹² Vgl. dazu: *de la Durantaye*, GRUR 2017, 558 ff.

¹³ Vgl. dazu nur: *Raue*, CR 2017, 656 ff.; *Spindler*, ZGE 2018, 273 ff.; *Specht*, OdW 2018, 285.

¹⁴ § 142 UrhG.

¹⁵ Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG, ABl. EU 2019 L 130/92.

¹⁶ Dies gilt insbesondere für Art. 15 und 17 DSM-RL, die dem Leistungsschutzrecht für Presseverleger bzw. der Haftung von Plattformbetreibern gewidmet sind.

a) Ausnahmen und Beschränkungen

Die Ausnahmen und Beschränkungen vom Urheberrecht der DSM-RL sind verpflichtend, anders als ein Großteil der InfoSoc-RL.¹⁷ Mitgliedstaaten müssen diese Vorgaben zwingend und dauerhaft in nationales Recht umsetzen. Das betrifft gewichtige Teile des Unterabschnitts 4 „Gesetzlich erlaubte Nutzungen für Unterricht, Wissenschaft und Institutionen“. So müssen §§ 60a, d, e und f UrhG und damit jene Normen, die im Gesetzgebungsverfahren besonders kontrovers diskutiert wurden, auch nach Februar 2023 in Kraft bleiben und deren Befristung spätestens mit Ablauf der Umsetzungsfrist am 7. Juni 2021 aufgehoben werden.

Trotzdem kann die geltende Rechtslage nicht unverändert beibehalten werden. Die DSM-RL zwingt den Gesetzgeber zu einigen Anpassungen der gesetzlichen Erlaubnistatbestände. So verpflichten etwa Art. 3 und 4 DSM-RL die Mitgliedstaaten, das Text und Data Mining in größerem Umfang zu gestatten, als § 60d UrhG dies bislang tut.¹⁸ Nach Art. 3 Abs. 2 DSM-RL müssen Mitgliedstaaten vorsehen, dass die Forscher das Korpus unbefristet für Zwecke späterer Forschung aufbewahren dürfen (Art. 3 Abs. 2 DSM-RL). Überdies sollen Mitgliedstaaten das Text und Data Mining vergütungsfrei zulassen (ErwGr 17).¹⁹ In geringerem Umfang als für wissenschaftliche Zwecke müssen Mitgliedstaaten auch das Text und Data Mining für kommerzielle Zwecke erlauben (Art. 4 DSM-RL).

Nach Art. 5 DSM-RL müssen Mitgliedstaaten überdies, anders als es das geltende Recht in Deutschland (§ 60a UrhG) derzeit vorsieht, den grenzüberschreitenden *remote access* zu Materialien erlauben, die digital zur Veranschaulichung von Unterricht oder Lehre zur Verfügung gestellt wurden. Fraglich ist, ob Bereichsausnahmen für bestimmte WerkGattungen, wie sie das deutsche Recht derzeit enthält, künftig Bestand haben können.

¹⁷ ErwGr. 5; zum Verhältnis der Schranken der DSM-Richtlinie zu jenen der InfoSoc-Richtlinie, vgl. nur: *Stieper*, GRUR 2020, 1 ff.

¹⁸ Vgl. dazu nur: *Raue*, ZUM 2019, 684 ff.; *Stieper*, GRUR 2020, 1, 4.

¹⁹ So sieht es auch der Diskussionsentwurf des BMJV für ein Erstes Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarkts vom 15. Januar 2020 vor, abrufbar unter: https://www.bmjbv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/DiskE_Anpassung%20Urheberrecht_digitaler_Binnenmarkt_Versendungsschreiben.html?nn=6712350, zuletzt abgerufen am 27.1.2020.

Dem Erhalt des kulturellen Erbes ist Art. 6 DSM-RL gewidmet. Danach müssen Mitgliedstaaten Einrichtungen des Kulturerbes gestatten, Werke und sonstige Schutzgegenstände für Zwecke der Erhaltung in dem dafür erforderlichen Umfang, mithin auch redundant, zu vervielfältigen – und zwar unabhängig vom konkreten Format oder Medium. Bildungseinrichtungen gehören nach der DSM-RL nicht zu den Einrichtungen des Kulturerbes (vgl. Art. 2 Nr. 3 DSM-RL). Im deutschen Recht werden sie hingegen in § 60d Abs. 1. i.V.m. § 60e Abs. 1 UrhG privilegiert. Das darf auch künftig so bleiben: Bildungseinrichtungen werden von Art. 5 Abs. 2 lit. c InfoSoc-RL erfasst. Seinen Anwendungsbereich lässt die DSM-RL unberührt (vgl. Art. 25 DSM-RL).²⁰

Wichtig für Gedächtnisinstitutionen ist, dass die Schranken – mit Ausnahme des kommerziellen Textes und Data Mining gem. Art. 4 DSM-RL – lizenzfest sind und sich auch gegenüber technischen Schutzmaßnahmen durchsetzen (Art. 7 DSM-RL).

b) Erleichterung der Lizenzierungspraxis

Überdies enthält die Richtlinie Vorgaben, die den Abschluss von Lizenzverträgen erleichtern sollen – vor allem in Bezug auf vergriffene Werke. Nach Art. 8 Abs. 1 DSM-RL müssen die Mitgliedstaaten festlegen, dass Verwertungsgesellschaften mit Einrichtungen des Kulturerbes Lizenzvereinbarungen über die nicht-kommerzielle Nutzung vergriffener Werke und sonstiger Schutzgegenstände abschließen dürfen, deren Rechteinhaber die Verwertungsgesellschaft nicht mit der Wahrnehmung ihrer Rechte betraut haben (sog. Außenseiter).²¹ Dafür muss die betreffende Verwertungsgesellschaft in zweifacher Hinsicht repräsentativ sein: erstens für die Art von Werken und sonstigen Schutzgegenständen und zweitens für die Rechte, auf die sich die Vereinbarung bezieht. Ist die Verwertungsgesellschaft – gegebenenfalls vermittelt über Gegenseitigkeitsverträge – in mehreren Mitgliedstaaten repräsentativ, kann sie Lizenzen für all diese Länder erteilen. Für Bereiche, in denen keine repräsentative Verwertungsgesellschaft existiert, müssen die Mitgliedstaaten ihren Einrichtungen des Kulturerbes über eine Schranke ermöglichen, Werke und sonstige Schutzgegenstände aus ihren Sammlungen für nicht-kommerzielle Zwecke auf nicht-kommerziellen Internetseiten zugänglich zu machen (Art. 8 Abs. 2, 3 DSM-RL).

²⁰ So auch: *Stieper*, GRUR 2020, 1, 5.

²¹ Vgl. dazu nur: *de la Durantaye/Kuschel*, ZUM 2019, 694 ff.; *de la Durantaye*, GRUR 2020, 7 ff.

Die Nutzungen sollen in einer zentralen Datenbank dokumentiert werden (Art. 10 DSM-RL); die Interessenträger sollen einen Dialog führen (Art. 11 DSM-RL). Rechteinhaber müssen der Nutzung einfach widersprechen können (Art. 8 Abs. 4 DSM-RL).

Damit werden alle Mitgliedstaaten verpflichtet, erweiterte kollektive Lizenzen für vergriffene Werke einzuführen.²² In Deutschland gibt es derartige Lizenzen *de facto* bereits (vgl. §§ 51, 52 VGG), allerdings nur für vergriffene Schriftwerke. Außerdem können bislang nur die Vervielfältigung und die öffentliche Zugänglichmachung lizenziert werden. Grenzüberschreitende Nutzungen, die die Richtlinie ausdrücklich gestatten will (Art. 9 DSM-RL), sind derzeit nicht möglich.

Elementar für den Erfolg der Regelungen wird sein, dass der nationale Gesetzgeber die Anforderungen an das Vergriffensein bei der Umsetzung nicht zu hoch ansetzt. Nach Art. 8 Abs. 5 DSM-RL soll mit „vertretbarem Aufwand“ geprüft werden, ob das Werk oder der Schutzgegenstand auf den üblichen Vertriebswegen für die Öffentlichkeit erhältlich ist. In den Erwägungsgründen wird ausgeführt, dass Stichproben in vielen Fällen ausreichen werden (ErwGr. 38 S. 6). Wer den Aufwand betreiben muss, dürfen die Mitgliedstaaten festlegen (ErwGr. 38 S. 2).

Damit hat der europäische Gesetzgeber Konsequenzen aus den Erfahrungen mit der Verwaiste Werke-RL²³ gezogen. Danach müssen die berechtigten Kultureinrichtungen eine „sorgfältige Suche“ nach dem Rechteinhaber jedes Werkes durchführen, bevor sie mit der Nutzung beginnen dürfen (Art. 2 Abs. 1, Art. 3).²⁴ Bei Massendigitalisierungsprojekten ist eine solche Suche mit ganz erheblichen Kosten verbunden. Gedächtnisinstitutionen machen von dieser Nutzungsmöglichkeit, die Deutschland in §§ 61 ff. UrhG umgesetzt hat, deswegen auch nur selten Gebrauch. Das beim Amt der Europäischen Union für geistiges Eigentum (EUIPO) geführte zentrale Register verwaister Werke weist derzeit

²² Nach Art. 12 DSM-RL dürfen sie auch in anderen Fällen, in denen die individuelle Lizenzierung so aufwändig ist, dass sie üblicherweise nicht erfolgt, erweiterte kollektive Lizenzen vorsehen; verpflichtet sind sie dazu aber nicht.

²³ Richtlinie 2012/28/EU des Europäischen Parlaments und des Rates vom 25.10.2012 über bestimmte zulässige Formen der Nutzung verwaister Werke, ABl. EU 2012 L 299/5.

²⁴ Diese Vorgaben wurden umgesetzt in §§ 61 Abs. 2, 61a UrhG.

für die gesamte EU lediglich 5.924 Werke aus.²⁵ Das beim Deutschen Patent- und Markenamt geführte Register vergriffener Werke hingegen enthält alleine für Deutschland 27.583 Einträge.

3. Völkerrecht

Im September 2016 trat der Vertrag von Marrakesch zur Erleichterung des Zugangs für blinde, sehbehinderte und anderweitig lesebehinderte Personen zu veröffentlichten Werken (Vertrag von Marrakesch)²⁶ in Kraft. Er ist der erste völkerrechtliche Vertrag, der ausschließlich Ausnahmen und Beschränkungen des Urheberrechts gewidmet ist.²⁷ Die EU hat die Vorgaben des Vertrags in der Marrakesch-RL umgesetzt.²⁸ Das deutsche Umsetzungsgesetz trat am 1. Januar 2019 in Kraft.²⁹

Die neuen §§ 45b-d UrhG modifizieren und ergänzen den Erlaubnistatbestand zugunsten von Menschen mit Behinderungen (§ 45a UrhG). Sie erlauben Blinden und sehbehinderten Menschen, veröffentlichte Texte und darin eingebettete Illustrationen in barrierefreie Formate zu konvertieren und konvertieren zu lassen, ohne den Rechteinhaber um Zustimmung bitten zu müssen (§ 45b Abs. 1 UrhG). Auch Einrichtungen für Blinde und sehbehinderte Menschen – insbesondere barrierefreie Bibliotheken – dürfen solche konvertierten Exemplare erstellen (§ 45c Abs. 1 UrhG). Sie dürfen sie untereinander tauschen und ihren Kunden sowohl analog als auch digital zur Verfügung stellen (§ 45c Abs. 2 UrhG), unabhängig davon, ob die Rechteinhaber barrierefreie Versionen anbie-

²⁵ Orphan Works Database, abrufbar unter: <https://euipo.europa.eu/orphanworks/#search/basic/all>, zuletzt abgerufen am 23.1.2020. Darin enthalten sind 6.902 eingebettete Werke.

²⁶ Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired or Otherwise Print Disabled, abrufbar unter: <https://wipolex.wipo.int/en/text/301016>, zuletzt abgerufen am 27.2.2020.

²⁷ Vgl. dazu: *Schmidt*, Maximalschutz im internationalen und europäischen Urheberrecht, 2018, Kapitel 3.

²⁸ Richtlinie (EU) 2017/1564 des Europäischen Parlaments und des Rates vom 13. September 2017 über bestimmte zulässige Formen der Nutzung bestimmter urheberrechtlich oder durch verwandte Schutzrechte geschützter Werke und sonstiger Schutzgegenstände zugunsten blinder, sehbehinderter oder anderweitig lesebehinderter Personen und zur Änderung der Richtlinie 2001/29/EG zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft, ABl. EU 2017 L 242/6.

²⁹ BGBl. 2018 Teil I Nr. 40, 2014 ff.

ten. Die Erlaubnistatbestände sind lizenzfest (§ 45d UrhG). Im Gegenzug enthält der Urheber einen Anspruch auf angemessene Vergütung (§ 45c Abs. 4 UrhG).

III. Digital Humanities

Wissenschaftliche Erkenntnis beruhte im analogen Zeitalter fast ausschließlich auf Leistungen des menschlichen Geistes und deren Vermittlung auf menschlichen Fähigkeiten. Durch das Aufkommen von Computern und der zunehmenden Digitalisierung von Erkenntnisquellen unterstützen Algorithmen immer stärker den menschlichen Erkenntnisprozess und erweitern die Möglichkeit, Informationen übersichtlich darzustellen. Dieses Potential des Digitalen nutzen zunehmend auch Geisteswissenschaftler. Die systematische Nutzung computergestützter Verfahren und digitaler Ressourcen in den Geisteswissenschaften werden schlagwortartig als „Digital Humanities“ bezeichnet.³⁰

Die Gewinnung neuer Erkenntnisse war im analogen Zeitalter weitgehend urheberrechtsneutral. Allenfalls das Kopieren von Texten oder Fotos, die in der eigenen Bibliothek nicht vorhanden waren, berührte die Interessen der Urheber. Die Digitalisierung von und die Arbeit mit digitalisierten Texten und anderen urheberrechtlich geschützten Schutzgegenständen (z.B. Bilder, Fotos, Videos, Audioaufnahmen) erfordern – wie oben unter I. bereits angesprochen – jedoch wenigstens eine vorübergehende Vervielfältigung im Arbeitsspeicher des Computers. Auch eine solch flüchtige Vervielfältigung hat der Gesetzgeber als Vervielfältigung im Sinne von § 16 UrhG eingestuft.³¹ Damit ist sie grundsätzlich dem Urheber vorbehalten. Flüchtige Verwertungshandlungen stellt typischerweise die Urheberrechtsschranke des § 44a UrhG frei. Noch nicht vollständig geklärt ist aber, ob auch die Nutzung offensichtlich rechtswidrig zur Verfügung gestellter Schutzgegenstände von § 44a UrhG gedeckt ist.³²

Forscher der Digital Humanities möchten Texte, Bilder und andere urheberrechtlich geschützte Gegenstände aber nicht nur betrachten oder durchsuchen,

³⁰ Erste Vorläufer entwickelten sich – unter anderen Namen – schon in den frühen 1960er Jahren. Ein geschichtlicher Überblick der Digital Humanities findet sich etwa bei: *Thaller*, Geschichte der Digital Humanities, in: Jannidis/Kohle/Rehbein, Digital Humanities – Eine Einführung, 2017, 2 ff.

³¹ Vgl. nur § 16 Abs. 1 UrhG: „gleichviel ob vorübergehend“ und Dreier/Schulze-Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 16 Rn. 12 m.w.N.

³² *Rau*, ZGE 2017, 514, 530 ff., 534.

sondern dauerhaft speichern, normalisieren, annotieren oder sonst anreichern und anschließend auf Trends, Korrelationen und andere Zusammenhänge (automatisiert) untersuchen. Dauerhafte Vervielfältigungen sind nicht von § 44a UrhG gedeckt, sondern müssen durch andere Schranken gerechtfertigt werden, etwa §§ 53, 60c ff. UrhG. Bei der Auslegung und Reichweite der Schranken ist insbesondere auch die Wissenschaftsfreiheit (Art. 13 GRCh; Art. 5 GG) zu berücksichtigen und mit dem Grundrecht auf Geistiges Eigentum (Art. 17 Abs. 2 GRCh; Art. 14 GG) abzuwägen.³³

Das Urheberrecht knüpft im Regelfall an menschliche Handlungen an.³⁴ Auch im digitalen Zeitalter lassen sich grob fünf Phasen der wissenschaftlichen Arbeit differenzieren: erstens die Recherche- und Materialsammelphase, zweitens die Aufbereitungsphase, in der die gesammelten Materialien geordnet, normalisiert, annotiert und sonst bearbeitet werden, drittens die Phase der Erkenntnisgewinnung, viertens die Ergebnispräsentation und fünftens die Datenaufbewahrung und -nachnutzung.

1. Recherche- und Materialsammelphase

In der Recherche- und Materialsammelphase stellen sich Zugangsfragen, die bei privaten Informationsinhabern nicht nur urheberrechtliche, sondern auch eigentums- und vertragsrechtliche Fragen aufwerfen sowie bei Informationen der öffentlichen Hand Fragen des Informationszugangs³⁵ und deren Weiterverwendung. Urheberrechtlich relevant ist im Forschungskontext, ob bereits das Betrachten bzw. Lesen digitaler Quellen, zu denen dem Wissenschaftler rechtswidrig Zugang verschafft wurde, eine Urheberrechtsverletzung darstellt.³⁶ Denn

³³ Nach dem EuGH muss die grundrechtliche Dimension der Nutzungsrechte dem Eigentumsrecht der Urheber gegenübergestellt werden; sie muss praktische Wirksamkeit entfalten, vgl. etwa: EuGH, Urt. v. 29.7.2019 – C-516/17, ECLI:EU:C:2019:625 = GRUR 2019, 940 Tz. 42 – *Spiegel Online/Volker Beck*.

³⁴ Urheberrechtlich relevant können auch unterlassene Vorsichts- und Beseitigungsmaßnahmen durch Verletzungen von Verkehrspflichten sein, im wissenschaftlichen Kontext etwa beim Betreiben von Repositorien. Vgl. allgemein zu (Verkehrs-)Pflichten von Plattformen: EuGH, Urt. v. 14.6.2017 – C-610/15, ECLI:EU:C:2017:456 = GRUR 2017, 790 Tz. 35 ff. – *Stichting Brein/Ziggo u.a. (Pirate Bay)* und Art. 17 DSM-RL; letzterer ist nur auf kommerzielle Plattformen anwendbar (Art. 2 Nr. 6 DSM-RL). Zur Urheberrechtswidrigkeit der Nutzung solcher Schattenbibliotheken: *Steinhauer*, Die Nutzung einer „Schattenbibliothek“ im Licht des Urheberrechts: Einige Überlegungen am Beispiel von Sci-Hub, 2016, abrufbar unter: <http://nbn-resolving.de/urn:nbn:de:hbz:708-dh4052>, zuletzt abgerufen am 27.1.2020.

³⁵ Zu den urheberrechtlichen Einschränkungen der Informationszugangsansprüche etwa: *Raue*, JZ 2013, 280 ff.

³⁶ Dazu: *Raue*, ZGE 2017, 514, 530 ff.

die Schutzgegenstände müssen dafür vorübergehend im Arbeitsspeicher des *digital device* gespeichert werden.

Wer digitale Materialien dauerhaft speichert und aufbereitet, muss das Urheberrecht beachten. Das gilt uneingeschränkt, wenn das Material urheberrechtlich geschützt ist. Die Schutzdauer für urheberrechtliche Verwertungshandlungen, die in Deutschland stattfinden, ist 70 Jahre nach dem Tod des Autors (§ 64 UrhG).³⁷ Aber auch an sich gemeinfreie Werke können dem Urheberrecht unterfallen. So können etwa (nicht-maschinelle) Übersetzungen urheberrechtlich geschützter Werke auch nach Ablauf der Schutzfrist des Originalwerks als Bearbeitungen nach § 3 UrhG geschützt sein. Ihr Schutz läuft erst 70 Jahre nach dem Tod des Übersetzers aus. Darüber hinaus genießen wissenschaftliche Ausgaben, also Ausgaben urheberrechtlich nicht geschützter Werke oder Texte, die das Ergebnis wissenschaftlich sichtender Tätigkeit sind und sich wesentlich von den bisher bekannten Ausgaben unterscheiden, ab Erscheinen 25 Jahre lang Schutz (§ 70 UrhG). Für die wissenschaftliche Praxis ebenfalls relevant ist, dass die Vervielfältigung von Datenbanken oder die Entnahme von Texten aus ihnen gegen das Urheberrecht verstoßen kann, wenn die Datenbanken entweder das Ergebnis schöpferischer Leistung (§ 4 Abs. 2 UrhG) oder wesentlicher Investitionen (§ 87 a UrhG) sind.

2. Aufbereitungsphase

Um digitale Korpora für anspruchsvolle wissenschaftliche Fragestellungen nutzen zu können, müssen diese im Regelfall aufbereitet werden.³⁸ Über die Frage nach dem Vervielfältigungsrecht hinaus stellt sich dabei die Frage, inwiefern dadurch in das Bearbeitungsrecht eingegriffen wird.³⁹ Der deutsche Gesetzgeber hat „ausschließlich technisch bedingte Änderungen eines Werkes nach § 60d Absatz 1, § 60e Absatz 1 sowie § 60f Absatz 2“ in § 23 S. 3 UrhG freigestellt. Dessen Verhältnis zum unionsrechtlichen Vervielfältigungsbegriff ist jedoch noch nicht vollständig geklärt.⁴⁰

³⁷ Bei mehreren Autoren richtet sich der Fristbeginn nach dem Tod des Längstlebenden, § 65 Abs. 1 UrhG.

³⁸ Vgl.: *Triaille/de Meeüs d 'Argenteuil/de Francquen*, Study on the legal framework of text and data mining, 2014, 47 f.; *Schöck*, Digitale Wissensproduktion, in: Jannidis/Kohle/Rehbein, Digital Humanities – Eine Einführung, 2017, 223, 227 ff.

³⁹ Dazu etwa: *Spindler*, GRUR 2016, 1112, 1113 f.; *Specht*, OdW 2018, 285.

⁴⁰ Vgl. etwa: *Ohly*, GRUR 2017, 964, 967 f. m.w.N.

Zugleich ist fraglich, ob die Digitalisierung und Aufbereitung von Texten Rechte nach dem Urheberrechtsgesetz oder sonstige Datenrechte begründen, wer dessen Inhaber ist und wem Nutzungsrechte an den Daten zustehen. Insbesondere in Forschungsgruppen wird die Frage nach Nutzungs- und Veröffentlichungsrechten relevant, wenn es zum Streit innerhalb der Gruppe kommt oder einzelne Forscher die Forschungseinrichtung wechseln.

3. Erkenntnisgewinnung

Zur Erkenntnisgewinnung setzen Forscher der Digital Humanities Algorithmen ein, die Zusammenhänge, Korrelationen und andere Erkenntnisse aus Texten und anderen Schutzgegenständen gewinnen können. Der europäische Gesetzgeber hat klargestellt, dass dies im Grundsatz auch dann nicht in das Urheberrecht eingreift, wenn dafür computergestützte Algorithmen verwendet werden.⁴¹ Die erforderlichen Vervielfältigungshandlungen sind für die nichtkommerzielle wissenschaftliche Forschung zurzeit nach § 60d UrhG,⁴² künftig auch für Forschungseinrichtungen und Einrichtungen des Kulturerbes, freigestellt.⁴³

4. Ergebnispräsentation

Werden die Erkenntnisse in gedruckter Form präsentiert und dabei urheberrechtlich geschützte Materialien verwendet, greift dies in das Vervielfältigungs- und Verbreitungsrecht ein (§§ 16, 17 UrhG). Für die Online-Präsentation muss das Recht der öffentlichen Zugänglichmachung (§ 19a UrhG) beachtet werden.⁴⁴ Weder die Wissenschaftsschranke des § 60c UrhG noch die Text und Data Mining-Schranke des § 60d UrhG stellen Präsentationen frei. Gegebenenfalls

⁴¹ ErwGr. 9 Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG. ABl. EU 2019 L 130/92; ferner Dreier/Schulze- Dreier, Urheberrechtsgesetz, 6. Aufl. 2018, § 60 d Rn. 1, 4; Fromm/Nordemann-A. Nordemann, Urheberrecht, 12. Aufl. 2018, § 60 d Rn. 4; Dreyer/Kotthoff/Meckel-Hentsch, Urheberrecht, 4. Aufl. 2018, § 60 d Rn. 3; Raue, ZUM 2019, 684, 685.

⁴² Dazu etwa: Oberfell, Big Data und Urheberrecht, in: Ahrens/Bornkamm/Ferzer/Koch/McGuire/Würtenberger, FS Büscher, 2018, 223, 228 ff.; Raue, CR 2017, 656 ff.; Spindler, ZGE 2018, 273, 277 ff.; Specht, OdW 2018, 285.

⁴³ Vgl. dazu den Diskussionsentwurf des Bundesministeriums der Justiz und für Verbraucherschutz, Entwurf eines Ersten Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarkts v. 15.1.2020.

⁴⁴ Specht, OdW 2018, 285, 287; Raue, CR 2017, 656, 660.

sind andere Erlaubnistatbestände, etwa die Zitatschranke (§ 51 UrhG), einschlägig.

Zu Gute kommt den Forschern aber, dass die in urheberrechtlich geschützten Materialien enthaltenen Informationen typischerweise nicht urheberrechtlich geschützt sind (mit Ausnahme von fiktionalen Werken). Lediglich wenn sie die wissenschaftlichen Erkenntnisse anhand von längeren Textpassagen oder von Bild-, Film- oder Tonaufnahmen veranschaulichen wollen, müssen sie sich auf einen gesetzlichen Erlaubnistatbestands berufen können. Auch wenn sie von gemeinfreien Datenbeständen auf urheberrechtlich geschützte Materialien verlinken, um *Linked Open Data* zu erstellen, müssen sie grundsätzlich urheberrechtliche Wertungen berücksichtigen.⁴⁵ Das Verlinken auf urheberrechtlich geschützte Inhalte ist danach urheberrechtsfrei, wenn diese mit Zustimmung des Urhebers frei verfügbar sind und keine Zugangssperren überwunden werden. Sind die Inhalte rechtswidrig veröffentlicht worden, dann haftet der Linksetzer, wenn er dies weiß oder wissen müsste. Bei der Ausgestaltung der Sorgfaltspflichten⁴⁶ sind allerdings die Informations- und die Wissenschaftsfreiheit angemessen zu berücksichtigen.

5. Nachnutzungsphase

Die gute wissenschaftliche Praxis erfordert, dass Forschungsdaten auch nach Abschluss eines Projekts gespeichert und Dritten für die Überprüfung der Ergebnisse sowie für die Anschlussforschung zur Verfügung gestellt werden.⁴⁷ Zugleich gilt: Das öffentliche Zurverfügungstellen urheberrechtlich geschützter Materialien in Forschungsrepositorien greift in das Recht der öffentlichen Zugänglichmachung (§ 19a UrhG) ein. Die Forscher müssen sich daher entweder auf eine Zustimmung des Urhebers oder auf eine gesetzliche Schranke stützen können. Dieses Problem können sie unter Umständen umgehen, indem sie Texte in einem abgeleiteten, reduzierten Textformat bereitstellen, das von der

⁴⁵ EuGH, Urt. v. 8.9.2016 – C-160/15, ECLI:EU:C:2016:644 = GRUR 2016, 1152, 1152 ff. – *GS Media/Sanoma*; überholt daher: BGH, Urt. v. 17.7.2003 – I ZR 259/00, GRUR 2003, 958, 961 f. – *Paperboy*.

⁴⁶ Dazu allgemein: Dreyer/Kotthoff/Meckel-Dreyer, Urheberrecht, 4. Aufl. 2018, § 15 d Rn. 81 ff.; Raue, ZGE 2017, 514, 528 f.

⁴⁷ Vgl. etwa: DFG, Leitlinien zur Sicherung guter wissenschaftlicher Praxis, Kodex, 2019, Leitlinie 12, 13, 17 sowie *Max-Planck-Gesellschaft*, Regeln zur Sicherung guter wissenschaftlicher Praxis, abrufbar unter: https://www.mpg.de/229457/Regeln_guter_wiss_Praxis__Volltext-Dokument_.pdf, zuletzt abgerufen am 27.1.2020.

urheberrechtlich geschützten Form entkleidet ist, aber für bestimmte Fragestellungen weiterverwendet werden kann.⁴⁸

Derzeit verpflichtet § 60d UrhG den Forscher, das Korpus und die Vervielfältigungen des Ursprungsmaterials nach Abschluss der Forschungsarbeiten zu löschen. Er darf sie zwar an Bibliotheken und andere Einrichtungen des Kulturerbes zur dauerhaften Aufbewahrung übermitteln. Unklar ist aber, inwiefern diese dem übermittelnden und anderen Forschern Zugang zu den Materialien verschaffen dürfen.⁴⁹ Künftig wird dies anders sein: In Umsetzung von Art. 3 DSM-RL wird § 60d UrhG keine Löschpflicht mehr enthalten.⁵⁰

Wie in der Bearbeitungs- und Aufbereitungsphase stellt sich auch in der Nachnutzungsphase die Frage, ob durch die Sammlung, Zusammenstellung, Ordnung und Aufbereitung der Texte sowie an den Forschungsergebnissen eigene Rechte der beteiligten Forscher entstehen, die einer Nachnutzung durch Dritte, aber auch durch andere Mitglieder der Forschungsgruppe entgegenstehen können.

IV. Fazit

Die Digitalisierung bietet Gedächtnisinstitutionen und Forschern der Digital Humanities viele Möglichkeiten. Wenn sie mit aktuellen, urheberrechtlich noch geschützten Gegenständen arbeiten wollen, gibt das Urheberrecht dafür einen Rahmen vor. Die urheberrechtlichen Schranken der §§ 60a ff. UrhG, die das UrhWissG 2018 in das UrhG eingefügt hat, haben die urheberrechtlichen Möglichkeiten für Gedächtniseinrichtungen und Forscher transparenter, übersichtlicher und anwendungsfreundlicher ausgestaltet. Allerdings haben diese zugleich neue Fragen und Probleme aufgeworfen. Darüber hinaus stellen sich ganz neue Fragen der Zuordnung von und der Nutzungsrechte an Forschungsdaten. Wir wollen mit der „Recht und Zugang“ ein Forum schaffen, in dem

⁴⁸ Vgl. dazu etwa den Workshop „Strategien für die Nutzbarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte“, abrufbar unter: <http://text-und-data-mining.de/workshops>, zuletzt abgerufen am 27.1.2020.

⁴⁹ Dazu: *Raue*, CR 2017, 656, 661; *Spindler*, ZGE 2018, 273, 285.

⁵⁰ § 60d Abs. 5 UrhG des Diskussions-Entwurfs lautet: „(5) Berechtigte nach Absatz 2 Satz 1 und Absatz 3 Nummer 1 dürfen Vervielfältigungen nach Absatz 1 mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufbewahren, solange sie für Zwecke der wissenschaftlichen Forschung oder zur Überprüfung wissenschaftlicher Erkenntnisse erforderlich sind.“

A. Urheberrecht und Zugang in einer digitalen Welt - Katharina de la Durantaye/Benjamin 17
Raue

diese Fragen aus Perspektive der Wissenschaft und der Gedächtniseinrichtungen beleuchtet werden. Sie sind herzlich eingeladen, dazu beizutragen.

B. Open Access in der Wissenschaft und im Urheberrecht sowie bei verwandten Schutzrechten

Karolina Benedyk

Version 1.0 (05.01.2024), CC BY-SA 4.0.

Digitaler Wandel ist auch bei der Veröffentlichung wissenschaftlicher Aufsätze allgegenwärtig. Neben der körperlichen Zeitschrift bieten digitale Datenbanken und das Internet potenziell Zugang zu allem Wissen. Allerdings gilt das Urheberrecht unabhängig vom Medium. Es ist also auch auf internetbasierte Zugänge und Verwertungen anwendbar. Dadurch schränkt das Urheberrecht die Verbreitung des Wissens ein. Das fällt insbesondere im Internet auf, das ohne Einschränkungen Zugänge in Sekundenschnelle ermöglichen kann. Open Science ist die Antwort auf den digitalen Wandel in der Wissenschaft. Diese ermöglicht weitreichenderen Zugang und vermeintlich niedrigere Publikationskosten. Allerdings können Urheber nicht auf ihr Urheberrecht verzichten.¹ Vielmehr ermöglicht ihnen das Gesetz, Lizenzen zu vergeben, die den Zugang und die Verwertung der Werke erlauben. Dabei stellen Creative Commons Musterlizenzverträge bereit, die eine einfache Veröffentlichung gewährleisten.

I. Open Science

1. Begrifflichkeiten

Im weiteren Verlauf werden einige Aspekte von Open Science dargestellt. Open Science ist der Oberbegriff für unterschiedliche Strategien und Verfahren in der Wissenschaft. Es beschreibt einen Wandel in der Forschungsmethodik, der unterschiedliche Ausprägungen hat.

Den Fokus legt die Handreichung auf Open Access. Damit ist der freie Zugang zu wissenschaftlichen Publikationen und anderen Materialien im Internet ge-

¹ Vgl. zum Werkschutz *Erler-Fridgen*, Teil 2, A. in diesem Band.

meint. Open Access ist eine Art der Lizenzierung, die den freien Zugang ermöglicht.² Das Verfahren ist aus der Kritik heraus entstanden, dass Forschung, welche häufig steuerlich finanziert ist, durch die Subskriptionsgebühren von Fachzeitschriften nur sehr teuer verfügbar ist. Dagegen halten Fachzeitschriften, dass sie die Peer-Review-Prozesse finanzieren und eine kuratierende Funktion aufweisen, also qualitative Aspekte erfüllen.³ In der Regel sind Fachzeitschriften, die im Open Access publizieren, auf andere Geschäftsmodelle angewiesen, seien es Publikationsgebühren („article processing charges“), konsortiale Finanzierungsmodelle oder auch institutionelle Förderung.

Digitale Forschung ist häufig datengetrieben. Zum Beispiel bedarf das Programmieren von Künstlicher Intelligenz einen großen Pool von Daten. Open Data bezeichnet übersetzt offene Daten, die für alle zugänglich sind und zu jedem Zweck verwendet und verarbeitet werden dürfen. Open Data ist eine Praxis, die wissenschaftliche Daten veröffentlicht und wiederverwendet werden.⁴ Wissenschaftler sehen Daten oftmals als öffentliches Gut an, wobei Verlage die Daten einschränken.⁵ Oftmals kommen nur geringe Datensätze in die Endpublikation, jedoch haben die vollen Datensätze ein größeres Gewicht.⁶

In der Informatik ist mittlerweile das Arbeiten mit Open Source verbreitet. Es hat eine lange Tradition. In der Softwareentwicklung ist der Quellcode frei zugänglich und kann mit einer Lizenz weiterverwendet werden.⁷ Die Vorteile hiervon sind unter anderem, dass Fehler schnell ausgebessert werden können.

2. Open Content-Lizenzmodell

Eine der ersten Lizenzen war die GNU General Public License (GPL) für Freie Software. Darauf aufbauend entwickelte das Creative Commons-Projekt ein allgemeines Lizenzmodell, das die Lizenzierung weiterer Werkarten ermöglichte.⁸ Die Creative Commons sind heutzutage die bedeutendsten Open Content-Lizenzmodelle.⁹ Das Ziel von Creative Commons ist es, der Allgemeinheit

² Heiser, Von Open Access zu Open Science, 238 ff.

³ Davis, 2014 Administrative Science Quarterly 59, 193 ff.

⁴ Murray-Rust, Nature Precedings 2008, 1.

⁵ Murray-Rust, Nature Precedings 2008, 1.

⁶ Murray-Rust, Nature Precedings 2008, 1.

⁷ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 69a Rn. 11.

⁸ Abrufbar sind die Lizenzen unter: <https://creativecommons.org/licenses/>, zuletzt abgerufen am 10.11.2022.

⁹ Paul, in Hoeren/Sieber/Holznel, Handbuch Multimedia-Recht, Werkstand: 58. EL 2022, Teil 7.4 Rn. 119.

freien Zugang zu kreativem Schaffen zu erleichtern.¹⁰ Wörtlich übersetzt bedeutet Creative Commons auch schöpferisches Gemeingut.

Durch die Bereitstellung von Musterlizenzverträgen ermöglichen sie eine kontrollierte Nutzungsfreigabe.¹¹ Die Werke stehen unter der „Creative Commons License“ (CCL) und machen durch standardisierte Symbole, Logos und Buttons auf sich aufmerksam.¹² Zudem bietet eine technische Vorkehrung weitere maschinenlesbare Transparenz.

CCL ermöglicht es zudem, zwischen unterschiedlichen Lizenzabstufungen zu wählen. Hierbei können sich Rechteinhaber entscheiden, ob sie ihre ganzen Rechte freigeben, oder sich einige Rechte wie z.B. das Namensrecht einbehalten.¹³ Die verschiedenen Lizenzverträge zeigen, wie und mit welchen Einschränkungen die Materialien verwendet werden dürfen. Ohne jegliche Bedingung existiert die CC0-Lizenz. Der Zusatz BY bedeutet Attribution und meint die Namensnennung des Urhebers. Bei SA (Share Alike) ist eine Weitergabe unter gleichen Bedingungen gestattet. Durch den Zusatz NC (Non-Commercial) ist eine kommerzielle Nutzung des Werkes ausgeschlossen und der Zusatz ND (No Derivatives) verbietet eine Bearbeitung. In den Digital Humanities ist eine Veröffentlichung mit einer CC-Lizenz mit unterschiedlichen Ausprägungen gängig.¹⁴

II. Digitaler Wandel in der Wissenschaft

Der digitale Wandel macht auch in der Wissenschaft keinen Halt. Daraus ergeben sich viele Vorteile.¹⁵ Wissenschaft lebt von interdisziplinärem Austausch und Erkenntnissen.¹⁶ Durch den digitalen Wandel können Forschende auf diese

¹⁰ *Paul*, in Hoeren/Sieber/Holznagel, Handbuch Multimedia-Recht, Werkstand: 58. EL 2022, Teil 7.4 Rn. 120.

¹¹ *Paul*, in Hoeren/Sieber/Holznagel, Handbuch Multimedia-Recht, Werkstand: 58. EL 2022, Teil 7.4 Rn. 121.

¹² *Paul*, in Hoeren/Sieber/Holznagel, Handbuch Multimedia-Recht, Werkstand: 58. EL 2022, Teil 7.4 Rn. 121.

¹³ *Paul*, in Hoeren/Sieber/Holznagel, Handbuch Multimedia-Recht, Werkstand: 58. EL 2022, Teil 7.4 Rn. 122.

¹⁴ Vgl. CC BY-ND: DHQ, Digital Humanities Quarterly: <http://www.digitalhumanities.org/dhqdev/about/about.html>; CC BY-SA: ZfdG, Zeitschrift für digitale Geisteswissenschaften: <http://www.zfdg.de/>; CC-BY: JCLS, Journal of Computational Literary Studies: <https://jcls.io/>.

¹⁵ *Euler*, RuZ 2020, 56 unter einer CC BY-SA Lizenz.

¹⁶ *Euler*, RuZ 2020, 56 unter einer CC BY-SA Lizenz.

Erkenntnisse jederzeit und von überall zugreifen.¹⁷ Dadurch konnte der Zugang des Wissens vom körperlichen Medium gelöst werden und ist grundsätzlich ubiquitär, also für jeden von überall zugänglich.

Auf der anderen Seite führt die Digitalisierung dazu, dass das Urheberrecht den Betroffenenkreis erweitert.¹⁸ Das UrhG regelt auch Handlungen von Endnutzenden.¹⁹ Der Zugang zu dem geschützten Wissen ist erst möglich bzw. rechtlich zulässig, soweit das Wissen keinen rechtlichen Restriktionen (mehr) unterliegt. Um das zu erreichen, veröffentlichen viele Wissenschaftlerinnen und Wissenschaftler ihre Veröffentlichungen mit lizenzfreiem Zugang.²⁰

Über den freien Zugang hinaus sind „offene Formate, Standards, Schnittstellen und freie Lizenzen wichtig für Open Access Angebote“.²¹ Open Access ermöglicht darüber hinaus nicht nur den Zugang, sondern auch die Möglichkeit, neue Erkenntnisse weiterzuverarbeiten.²² Die Nachnutzung ist allerdings urheberrechtlich als Verarbeitung einzuordnen und unterliegt rechtlichen Restriktionen.²³

In den Digital Humanities werden Korpora im Rahmen von Open Access veröffentlicht, soweit die zugrundeliegenden Texte gemeinfrei sind. Die Digitale Bibliothek von TextGrid bietet eine umfangreiche Sammlung XML/TEI-erschlossener Texte aus Belletristik und Sachliteratur vom Anfang des Buchdrucks bis zu den ersten Jahrzehnten des 20. Jahrhunderts, die in deutscher Sprache verfasst oder übersetzt wurden.²⁴ Für die germanistische und vergleichende Literaturwissenschaft ist die Sammlung von besonderem Interesse, da sie nahezu alle wichtigen kanonisierten Texte und zahlreiche weitere literaturhistorisch relevante Texte enthält, deren urheberrechtliche Schutzfrist abgelaufen ist.²⁵ Obwohl die Texte gemeinfrei waren, unterlag die Zusammenstellung Leistungsschutzrechten. Um die Digitale Bibliothek nachnutzbar zu machen, wurde sie freigekauft, um sie dann mit CC-Lizenz publizieren zu können.²⁶

¹⁷ Euler, RuZ 2020, 56 unter einer CC BY-SA Lizenz.

¹⁸ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, Einl. Rn. 25.

¹⁹ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, Einl. Rn. 25.

²⁰ Euler, RuZ 2020, 56 unter einer CC BY-SA Lizenz.

²¹ Euler, RuZ 2020, 56, 59 unter einer CC BY-SA Lizenz.

²² Euler, RuZ 2020, 56, 59 unter einer CC BY-SA Lizenz.

²³ Vgl. Erler-Fridgen, Teil 1, C. in diesem Band.

²⁴ Vgl. Die Digitale Bibliothek bei TextGrid – TextGrid: <https://textgrid.de/digitale-bibliothek>, zuletzt abgerufen am 15.03.24.

²⁵ Vgl. Die Digitale Bibliothek bei TextGrid – TextGrid: <https://textgrid.de/digitale-bibliothek>, zuletzt abgerufen am 15.03.24.

²⁶ Vgl. <https://sprache.hypotheses.org/2436>.

III. Urheberrecht

Das Urheberrecht macht Informationen zu einem handelbaren Wirtschaftsgut.²⁷ Es ordnet das Werk dem Schöpfer zu, vgl. § 7 UrhG.²⁸ Diese Person kann selbst entscheiden, wem und in welchem Maße sie die Informationen zur Verfügung stellt. Sind keine Schrankenbestimmungen einschlägig, kann Wissen zu einem exklusiven Recht vereinsamen.

Den beschränkten Zugang zu Wissen bemerken Studierende beispielsweise oft erst, wenn sie die Bildungseinrichtung verlassen. Hochschulen, Universitäten und Forschungsinstitute sind durch die übergreifenden Lizenzverträge Oasen des Wissens.²⁹ Ansonsten ist der Informationszugang im Internet stark reglementiert. Selbst wenn der Zugang oft noch ermöglicht wird, sind weitere Verwertungsmöglichkeiten nicht erlaubt. Einfluss darauf hat der Urheber. Zugang und Verwertung sind von seiner Einwilligung abhängig.

1. Nutzungsrechte

Das Urheberrecht billigt dem Urheber oder Leistungsrechtsinhaber das alleinige Nutzungsrecht zu, vgl. § 15 UrhG. Dieses entsteht von Gesetzes wegen. Die Nutzung von Werken stellt mindestens eine Vervielfältigungshandlung dar.³⁰ Für die rechtmäßige Nutzung wissenschaftlicher Werke bedarf es somit einer gesetzlichen Ausnahme für die Nutzungshandlung (Schranken)³¹ oder einer Einwilligung des Urhebers.³²

2. Verzicht auf das Urheberrecht

Soweit der Urheber auf seine Werke verzichten könnte, wären sie „frei“ und unterlägen nicht mehr gesetzlichen Restriktionen. Urheberrechte sind allerdings nicht vollständig übertragbar, vgl. § 29 I UrhG. Das ist Ausdruck der moralistischen Theorie im Urheberrecht. Diese besagt, dass zwischen den persönlichkeitsrechtlichen als auch den vermögensrechtlichen Interessen eine enge

²⁷ Peukert/Sonnenberg, in Weingart/Taubert (Hrsg.), *Wissenschaftliches Publizieren*, 2006, 211, 230.

²⁸ Vgl. hierzu Handreichung *Erler-Fridgen*, Teil 2, A. in diesem Band.

²⁹ Euler, RuZ 2020, 56, 58 unter einer CC BY-SA Lizenz.

³⁰ Vgl. *Erler-Fridgen*, Teil 1, C. in diesem Band.

³¹ Vgl. hierzu *Benedyk/Erler-Fridgen*, Teil 4, D. in diesem Band; *Erler-Fridgen*, Teil 4, C. in diesem Band; *Erler-Fridgen*, Teil 4, A. in diesem Band.

³² Euler, RuZ 2020, 56, 70 unter einer CC BY-SA Lizenz.

wechselseitige Beziehung besteht.³³ Daraus, dass das Urheberrecht nicht übertragbar ist, folgert die Literatur, dass der Urheber nicht im Ganzen auf das Urheberrecht verzichten kann.³⁴

Haben Parteien einen Vertrag auf Übertragung des Urheberrechts geschlossen, ergibt die Vertragsauslegung regelmäßig, dass sich der Vertrag auf die Einräumung von Nutzungsrechten richtet.³⁵ Der Urheber kann das Werk oder Leistungsschutzrecht nicht übertragen, jedoch hindert ihn nichts daran, es in Gänze zu lizenzieren, vgl. § 29 II UrhG. Nach der Zweckübertragungslehre räumt der Urheber nur insoweit die Rechte ein, als es der Vertragszweck erfordert, vgl. § 31 V UrhG.³⁶ Verzichtet der Urheber mithin innerhalb eines Vertrages auf seine Rechte, ergibt die Vertragsauslegung nach § 31 V UrhG, dass er dem Vertragspartner unter Umständen ein ausschließliches Recht gewährt. Die Verzichtserklärung ist eng auszulegen und nur auf das ausschließlich Gewollte zu beschränken.³⁷ Ist ein ausschließliches Recht nicht erforderlich, um den Vertragszweck zu erreichen, ergibt sich aus der Zweckübertragungslehre, dass nur die notwendigen Verwertungsrechte für den Vertragszweck übertragen werden.

3. Open Content

Eine besondere Art von Lizenzverträgen sind die Open-Content-Verträge. Im Gegensatz zu Verlagsverträgen haben sie nicht den Zweck, eine ausschließliche Rechtsposition zu verschaffen.³⁸ Vielmehr ermöglichen sie es jedem, das Werk vergütungsfrei zu nutzen.³⁹ Geben Urheber ihre Werke in Form von Open Content frei, räumen sie der Allgemeinheit vergütungsfreie, nicht-ausschließliche Nutzungsrechte gem. § 31 II UrhG zur Vervielfältigung und Verbreitung des betroffenen Inhalts ein.⁴⁰ Trotz des offenen Zugangs bleiben die Rechte beim Urheber.

³³ *Wandtke*, in *Wandtke/Bullinger*, Urheberrecht, 5. Auflage 2019, Einl. 5.

³⁴ *Hochte*, in *Wandtke/Bullinger*, Urheberrecht, 5. Auflage 2019, § 29 Rn. 15; *Obly*, in *Schricker/Löwenheim*, Urheberrecht, 6. Auflage 2020, § 29 Rn. 8.

³⁵ *Obly*, in *Schricker/Löwenheim*, Urheberrecht, 6. Auflage 2020, § 29 Rn. 8.

³⁶ BGH GRUR 2003, 234 236 – EROC III; *Schulze*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 31 Rn. 110.

³⁷ *Schulze*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 29 Rn. 10.

³⁸ *Peukert/Sonnenberger*, in *Weingart/Taubert* (Hrsg.), Wissenschaftliches Publizieren, 2006, 211, 227.

³⁹ *Hochte*, in *Wandtke/Bullinger*, Urheberrecht, 6. Auflage 2022, § 29 Rn. 22.

⁴⁰ OLG Köln CR 1996, 723, 725; *Hochte*, in *Wandtke/Bullinger*, Urheberrecht, 6. Auflage 2022, § 29 Rn. 22.

4. Urheberrechte im wissenschaftlichen Kontext

Nach § 43 UrhG gilt das vorher Beschriebene ebenso, wenn der Urheber das Werk in Erfüllung seiner Verpflichtung aus einem Arbeits- oder Dienstverhältnis geschaffen hat, soweit sich aus dem Inhalt oder dem Wesen des Arbeits- oder Dienstverhältnisses nichts anderes ergibt. Wissenschaftliches Personal unterscheidet sich in Personen, die freier wissenschaftlicher Tätigkeit nachgehen und angestellten Forschern.⁴¹ Hochschulprofessoren forschen frei und eigenverantwortlich und haben keine Pflicht zur Veröffentlichung ihrer Forschungsergebnisse.⁴² Daraus ergibt sich, dass die Urheberrechte an den Arbeitsergebnissen nur ihnen und nicht dem Dienstherrn zur Gute kommen.⁴³ Etwas anderes gilt für Arbeitsergebnisse von wissenschaftlichen Assistenten hinsichtlich Arbeiten, die im Kontext weisungsgebundener Tätigkeit entstanden sind.⁴⁴ Für ihre eigene wissenschaftliche Arbeit stehen wissenschaftlichen Assistenten jedoch nach § 43 UrhG die Urheberrechte uneingeschränkt zu. Es erfolgt keine Übertragung an die Forschungseinrichtung. Vielmehr ist der Urheber frei, die Veröffentlichung flexibel auszugestalten. Er kann einem Verlag ein zeitlich beschränktes ausschließliches Nutzungsrecht einräumen und sich vorbehalten, das Werk in Form von Open Access zu veröffentlichen.⁴⁵

Allerdings lassen sich Verlage unter Umständen nicht auf diese Möglichkeit ein, oder der Urheber verhandelt den Verlagsvertrag nicht dementsprechend. Erhält der Verlag sodann das ausschließliche Recht, würde der Urheber selbst eine Urheberrechtsverletzung begehen und gegen den Verlagsvertrag verstoßen, wenn er das Werk veröffentlichte.⁴⁶

Dieser Effekt soll durch das Zweitverwertungsrecht nach § 38 IV UrhG durchbrochen werden.⁴⁷ Hiernach haben Urheber wissenschaftlicher Beiträge, die im Rahmen einer mindestens zur Hälfte mit öffentlichen Mitteln geförderten For-

⁴¹ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 43 Rn. 12.

⁴² Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 43 Rn. 12.

⁴³ BGH GRUR 1991, 523, 525 – Grabungsmaterialien.

⁴⁴ LG Köln ZUM 2000, 579.

⁴⁵ Peukert/Sonnenberger, in Weingart/Taubert (Hrsg.), Wissenschaftliches Publizieren, 2006, 211, 235 f.

⁴⁶ Peukert/Sonnenberger, in Weingart/Taubert (Hrsg.), Wissenschaftliches Publizieren, 2006, 211, 236.

⁴⁷ Peukert/Sonnenberger, in Weingart/Taubert (Hrsg.), Wissenschaftliches Publizieren, 2006, 211, 236.

schungstätigkeiten entstanden und in einer periodisch mindestens zweimal jährlich erscheinenden Sammlung erschienen sind, das Recht, den Beitrag nach Ablauf von zwölf Monaten seit der Erstveröffentlichung in der akzeptierten Manuskriptversion öffentlich zugänglich zu machen. Dies gilt auch, wenn sie dem Verlag ein ausschließliches Nutzungsrecht eingeräumt haben.

Ein weiteres Verwertungsrecht nach 10 Jahren räumt § 40a I 1 UrhG bei pauschaler Vergütung eines Urhebers ein. Damit soll eine angemessene Vergütung während der langen Schutzdauer des Urheberrechts erreicht werden.⁴⁸ Verleiht der Urheber ein ausschließliches Nutzungsrecht zeitlich unbegrenzt für eine einmalige pauschale Vergütung, ist der Verwerter der faktische Nutznießer der langen Schutzdauer.⁴⁹ Das Recht ist allerdings nur für Verträge und sonstige Sachverhalte anwendbar, die ab dem 01.03.2017 geschlossen wurden, vgl. § 132 IIIa UrhG. Für Verträge aus der Zeit davor ist weiterhin das bisherige Recht anwendbar.⁵⁰

IV. Leistungsschutzrechte

Leistungsschutzrechte sind im Gegensatz zu Urheberrechten zum Teil übertragbar, wenn ihnen eine urheberpersönlichkeitsrechtliche Komponente fehlt.⁵¹ Der Schutz wissenschaftlicher Ausgaben nach § 70 UrhG ist nicht vollständig übertragbar.⁵² Diese schützt die persönliche Leistung und hat damit einen persönlichkeitsrechtlichen Kern.⁵³

Im Gegensatz zum Datenbankwerk ist der Datenbankherstellerschutz⁵⁴ nicht darauf ausgerichtet, die schöpferische Auswahl und Anordnung zu schützen, sondern die Investition in die Beschaffung, Sammlung, Überprüfung, Aufbereitung und Darbietung des Inhalts.⁵⁵ Geschützt ist hierbei nicht der Urheber, sondern die Person, die das Investitionsrisiko trägt.⁵⁶

⁴⁸ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 40a Rn. 1.

⁴⁹ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 40a Rn. 1.

⁵⁰ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 40a Rn. 2.

⁵¹ *Hochte*, in Wandtke/Bullinger, Urheberrecht, 5. Auflage 2019, § 29 Rn. 5, 10 ff.; *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 29 Rn. 21.

⁵² *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, Vor § 28 Rn. 3.

⁵³ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 70 Rn. 9.

⁵⁴ Vgl. hierzu Handreichung *Erlar-Fridgen*, Teil 2, D. in diesem Band.

⁵⁵ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87a Vor. Rn. 1.

⁵⁶ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87a Rn. 18.

Zudem gilt auch, dass auf Leistungsschutzrechte verzichtet werden kann, soweit sie übertragbar sind und dem keine gesetzlichen Vorschriften entgegenstehen, z.B. § 79 II 2 iVm § 32 III 1 UrhG.⁵⁷

V. Zukunft von Open Access

Das Urheberrecht bereitet der Wissenschaft Schwierigkeiten, Quellen einzubinden oder frei zu veröffentlichen. Radikale Stimmen fordern deswegen eine Aufhebung des Urheberrechts für wissenschaftliche Werke, die größtenteils durch staatliche Stellen finanziert wurden.⁵⁸ Allerdings scheitern solch drastische Änderungen bereits an den völkerrechtlichen Konventionen zum Urheberrecht.⁵⁹

Durch Open Science wird versucht, die urheberrechtlichen Defizite zu überwinden, indem auf eine offene Lizenzierung gesetzt wird.⁶⁰ Dabei soll die Digitalisierung auch dem wissenschaftlichen Prozess nützen, indem sie das Wissen im Internet offen zugänglich ausgestaltet. Neben dem freien Zugang ist auch die freie Nutzungsmöglichkeit gefordert.⁶¹ Dadurch sollen Verarbeitungsmodelle wie das Text und Data Mining den wissenschaftlichen Fortschritt ermöglichen. Open Science ist darauf ausgerichtet, transparente und zugängliche Informationen zu teilen und Wissen durch kollaborative Zusammenarbeit zu ermöglichen.⁶² Auch die Europäische Kommission hat in sich in ihrem für 2020 – 2024 eine Strategie für offene Wissenschaft in der EU gesetzt.⁶³ Unter anderem verlangt die Kommission von Empfängern von Forschungs- und Innovationsmitteln, dass sie ihre Veröffentlichung mit offenem Zugang zur Verfügung stellen.

⁵⁷ *Hochte*, in Wandtke/Bullinger, Urheberrecht, 5. Auflage 2019, § 29 Rn. 21.

⁵⁸ *Peukert/Sonnenberger*, in Weingart/Taubert (Hrsg.), Wissenschaftliches Publizieren, 2006, 211, 232; vgl. die abgeschwächte Form *Hansen*, GRUR Int. 2005, 378, 383 f.

⁵⁹ *Peukert/Sonnenberger*, in Weingart/Taubert (Hrsg.), Wissenschaftliches Publizieren, 2006, 211, 232.

⁶⁰ *Heiser*, Von Open Access zu Open Science, 241.

⁶¹ *Euler*, RuZ 2020, 56, 57 unter einer CC BY-SA Lizenz.

⁶² *Vicente-Saez et. al.*, 2018 Journal of Business Research 88, 428.

⁶³ Vgl. https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en.

VI. Ergebnis

Open Science entstand aus der Idee heraus, wissenschaftliche Beiträge, Daten o.ä. zu teilen. Kollaboratives Schaffen von mehr Wissen ist der Hintergedanke dieser Vorgehensweise. Um das zu erreichen, müssen urheberrechtlich geschützte Werke zur freien Verfügung lizenziert werden. Urheber können nämlich nicht auf ihr Recht verzichten. Des Weiteren gibt es gesetzliche Schranken⁶⁴, die jedoch nur in bestimmten Kontexten greifen. Darüber hinaus können Rechteinhaber auf Leistungsschutzrechte zum Teil verzichten. Dabei ist zu beachten, dass die Regelungen nicht auf den ersten Teil des Urhebergesetzes hinweisen dürfen. Leistungsschutzrechte unterscheiden sich von Urheberrechten dahingehend, dass sie in erster Linie einen Investitionsschutz bieten. Der persönlichkeitsrechtliche Kern ist nur schwach ausgeprägt.

Weit verbreitete Musterlizenzverträge sind die Creative Commons-Lizenzen. Sie sind in unterschiedlichen Abstufungen vorhanden und haben zur erleichterten Einhaltung einen digitalen Fingerabdruck.

Open Science ist möglich, beruht aber weiterhin auf der freiwilligen Entscheidung der Wissenschaftlerinnen und Wissenschaftler. Hierbei können alle zusammen davon profitieren.

Leseempfehlung zur vertiefenden Lektüre: *Euler*, RuZ 2020, 56 unter einer CC BY-SA Lizenz.

⁶⁴ Vgl. hierzu die Handreichungen *Benedyk/Erler-Fridgen*, Teil 4, D. in diesem Band; *Erler-Fridgen*, Teil 4, C. in diesem Band; *Erler-Fridgen*, Teil 4, A. in diesem Band.

C. Verfahrensschritte bei dem Einsatz von Text und Data Mining-Verfahren in den Geisteswissenschaften

Katharina Erler-Fridgen

Version 1.0 (02.06.2022), CC BY-SA 4.0.

Werden urheberrechtlich geschützte¹ Texte mit Hilfe von Text und Data Mining Verfahren analysiert, unterliegen die vorgenommenen Handlungen grundsätzlich urheberrechtlichen Restriktionen. Denn werden Vervielfältigungen oder Entnahmen aus Datenbanken² vorgenommen oder der Analyse zugrundeliegende Text(teil)e präsentiert³, können die Verwertungsrechte des Urhebers beeinträchtigt werden. Ist dies der Fall, bedarf es einer urheberrechtlichen Gestattung: Die Text und Data Mining Schranken in §§ 44b und 60d UrhG können einen urheberrechtlichen Rahmen für den Einsatz von Text und Data Mining Verfahren bieten.⁴ Außerhalb dieser Gestattung finden sich in den § 44a ff. UrhG weitere Schranken, die einen gewissen Rahmen – wie beispielsweise das Zitatrecht⁵ in § 51 UrhG oder § 60c UrhG – für die Auseinandersetzung mit oder die Nutzung von Texten bieten können. Außerhalb dieser Schranken wird es nötig, die Zustimmung des Urhebers zu einer Verwertungshandlung einzuholen. Im Folgenden werden Verfahrensschritte beim Text und Data Mining in den Geisteswissenschaften geschildert und daran anknüpfend urheberrechtlich relevante Handlungen beschrieben.

I. Beschreibung des iterativen Verfahrens

Die Digital Humanities entwickeln und nutzen Verfahren zur Identifikation, Extraktion, Analyse und Vernetzung von fachwissenschaftlich relevanten Informationen aus einschlägigen Datenbeständen. Je nach Projektkontext werden unterschiedliche Text- und Datenquellen gesammelt und aufbereitet, um

¹ *Erler-Fridgen*, Teil 2, A. in diesem Band.

² *Erler-Fridgen*, Teil 2, D. in diesem Band.

³ *Erler-Fridgen*, Teil 6, A. in diesem Band.

⁴ *Erler-Fridgen*, Teil 4, A. in diesem Band.

⁵ *Erler-Fridgen*, Teil 4, C. in diesem Band.

sodann Fachinformationen zu extrahieren, zu vernetzen und schließlich zu präsentieren. Die urheberrechtliche Perspektive knüpft üblicherweise an menschliche Handlungen an.⁶ Daher sind die genannten Handlungen wesentliche Anknüpfungspunkte für die urheberrechtliche Bewertung des Verfahrens beim Text und Data Mining in den Geisteswissenschaften.⁷

Werden bei der Sammlung, Aufbereitung und Auswertung bestehender Texte Kopiervorgänge vorgenommen, so sind diese Verfahrensschritte dazu geeignet, das Vervielfältigungsrecht des Urhebers nach § 16 UrhG zu beeinträchtigen.⁸ Dabei werden auch vorübergehende Vervielfältigungen in dem Arbeitsspeicher des Computers als relevante Vervielfältigungen im Sinne des § 16 UrhG angesehen, können jedoch aufgrund der Schranke des § 44a UrhG gestattet sein.⁹ Jedoch ist etwa für die Zusammenstellung, Normalisierung und Annotation eines Datenkorpus bei fundierten Textanalysen in der Regel notwendig, die Vervielfältigungen dauerhaft zu speichern, was eine über § 44a UrhG hinausgehende urheberrechtliche Gestattung nötig macht.¹⁰

Auch die Übernahme von Textteilen kann rechtswidrige Vervielfältigungshandlungen begründen, weil auch Textteile eigenständig urheberrechtlich geschützt sein können.¹¹ Werden schließlich die Auswertungsergebnisse im Netz präsentiert sowie die Informationen zur Nachnutzung durch Dritte bereitgestellt, können – je nach Art und Umfang¹² – Eingriffe in das Recht des Urhebers auf öffentliche Zugänglichmachung nach § 19a UrhG begründet werden.¹³

⁶ *De la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 15f, siehe zu Verwertungshandlungen v. *Ungern-Sternberg*, in Schrickler/Loewenheim, Urheberrecht, 6. Auflage 2020, § 15 Rn. 10.

⁷ So ausführlich auch: *De la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 15f.

⁸ *De la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 15 f.

⁹ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 16 Rn. 7; Flüchtige Vervielfältigungshandlungen können jedoch von der urheberrechtlichen Schranke des § 44a UrhG freigestellt sein, zu den Voraussetzungen siehe: v. *Welser*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 44a Rn. 1.

¹⁰ *Erler-Fridgen*, Teil 4, A, II. in diesem Band mit Verweis auf *Raue*, ZUM 2021, 793, 795; siehe auch *Raue*, IIC 2018, 379, 381: vorbereitende Handlungen wie die Digitalisierung, die Normalisierung oder die Annotierung des Korpus verursachen typischerweise eine längere Speicherdauer als sie von Art. 5 Abs. 1 InfoSoc-RL/§ 44a UrhG freigestellt ist.

¹¹ Zum eigenständigen urheberrechtlichen Schutz von Textteilen *Erler-Fridgen*, Teil 6, A. in diesem Band.

¹² Beispielsweise könnten vollständige Primärtexte oder lediglich Analyseergebnisse veröffentlicht werden. Auch können Textteile als Ergänzung zu extrahierten Informationen präsentiert werden.

¹³ *Jotzo*, Teil 2, B. in diesem Band, 61, 63 f.

Die fünf Phasen des wissenschaftlichen Arbeitens¹⁴ können eingesetzt werden, um den Einsatz von Text und Data Mining Verfahren in den Geisteswissenschaften aus urheberrechtlicher Sicht zu beleuchten: Werden solche Textanalyseverfahren genutzt, werden Texte und Daten gesammelt (1), diese aufbereitet (2), Informationen extrahiert (3), die Analyseergebnisse und Text(teile) präsentiert (4) sowie die Forschungsdaten archiviert (5).¹⁵ Im Folgenden sollen diese einzelnen Verfahrensschritte bei der Anwendung von Text und Data Mining Verfahren in den Geisteswissenschaften dargestellt werden. Insbesondere sollen dabei mögliche urheberrechtlich relevante Handlungen identifiziert werden.

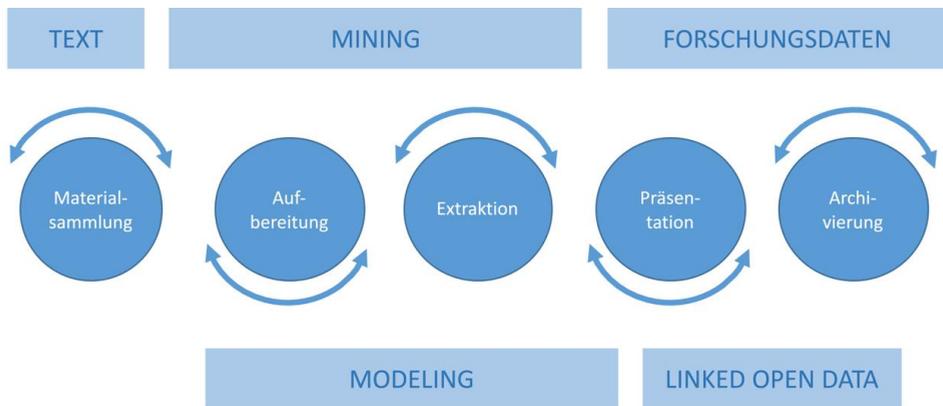


Abbildung 1: „Iterative Verfahrensschritte bei dem Einsatz von Text und Data Mining in den Geisteswissenschaften“ von Katharina Erler-Fridgen und Prof. Dr. Benjamin Raue, CC BY-SA 4.0.

Die genannten Verfahrensschritte¹⁶ erfolgen technisch wie konzeptionell in einem iterativen Verfahren, in dem die einzelnen Verfahrensschritte typischerweise ineinandergreifen. Je nach analysierter Textart und Forschungsansatz kann die Reihenfolge der Verfahrensschritte variieren. Beispielsweise kann die

¹⁴ Ausführlich zu diesen fünf Phasen siehe *de la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 15f.

¹⁵ Hierzu grundlegend *de la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 15 f.

¹⁶ Allgemein zu den fünf Phasen wissenschaftlichen Arbeitens im digitalen Zeitalter siehe: *De la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 15 f.; zu etwas gebündelten Verfahrensschritten: *Specht*, OdW 2018, 285.

(semi-) automatische Informationsextraktion als Teil der Modellierung selbst Voraussetzung zur weiteren Aufbereitung der Textquellen sein (s.u.).

II. Die Verfahrensschritte im Einzelnen¹⁷

1. Sammlung der Text- und Datenquellen

Um Fachinformationen zu gewinnen, die zur weiteren Wissensgenerierung ausgewertet werden sollen, müssen in einem ersten Schritt entsprechende Text- und Datenquellen gesammelt werden. Je nach Art der gewünschten Informationen können unterschiedliche Textarten als Quelle genutzt werden. Beispielsweise werden Primärtexte, Fachliteratur sowie bibliographische Datenbestände eingesetzt. Die gesammelten Text- und Datenquellen befinden sich meist in einem unterschiedlichen Aufbereitungszustand und können unterschiedlichen Publikationsquellen entstammen, z.B. Originalausgaben, wissenschaftliche Editionen, Sammelbänden, Bibliographien, Metadatenplattformen, Open Access Plattformen etc.

a) Sammlung verschiedener Text- und Datenarten

Für die Sammlung von Metadaten literarischer Werke werden beispielsweise bibliographische Verzeichnisse, Werkverzeichnisse eines Urhebers oder bereits vorhandene Datenbanken herangezogen. Auch Metadatendienste der Bibliotheken¹⁸ und Archive können hierfür genutzt werden.

¹⁷ Dank gilt Prof. Dr. Christof Schöch und Dr. Maria Hinzmann, Projekt MiMoText, Universität Trier, für den Austausch zum Thema.

¹⁸ Bspw. sind über die Metadatendienste der Deutschen Nationalbibliothek alle Titeldaten des Bestands aktualisiert abrufbar als Linked-Data-Service, teilweise mittels Klassifikation der Werke nach Thema siehe https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/LDS/lids_node.html, zuletzt abgerufen am 12.01.2022; darüber hinaus werden auch Entity-Facts (maschinenlesbare Faktenblätter zur Analyse von Entitäten, bspw. Namensformen) https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/Entity-Facts/entity-facts_node.html (zuletzt abgerufen am 12.01.2022) innerhalb des Bibliothekbestands zur Verfügung gestellt.

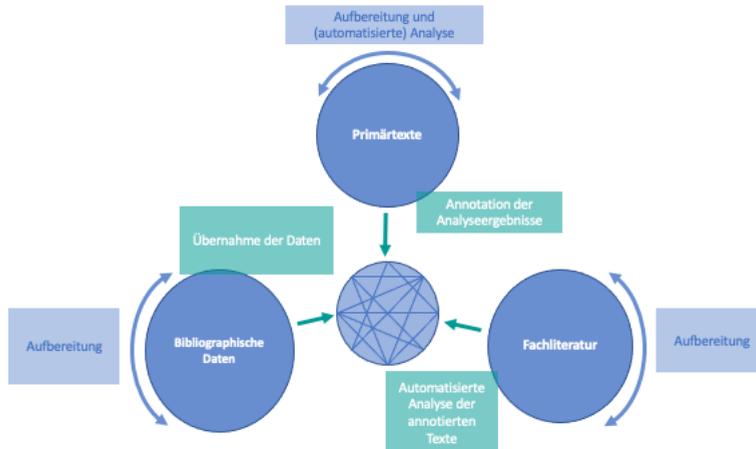


Abbildung 2: Beispiel für die Sammlung, Aufbereitung und Extraktion von unterschiedlichen Textquellen

Zum Beispiel kann eine (semi-)automatisierte Auswertung von Texteigenschaften auf Basis von Primärtexten erfolgen, die jedoch in maschinenlesbarer Form vorliegen oder entsprechend aufbereitet werden müssen (siehe Beispiel in Abbildung 2; zur Aufbereitung allgemein unten II). Bereits aufbereitete Primärtexte können in Datenbanken¹⁹ oder auf Open Source Plattformen verfügbar sein.²⁰ Primärtexte, die noch aufbereitet werden müssen, können der Originalveröffentlichung oder wissenschaftlichen Editionen entnommen werden. Des Weiteren können Analyseergebnisse aus Text und Data Mining-Analysen auch zur weiteren Auswertung annotiert werden (siehe Beispiel in Abbildung 2). Fachliteratur und fachliterarische Teile wissenschaftlicher Editionen sowie Lexika und sonstige Sekundärliteratur können ebenfalls als Quellen der Fachinfor-

¹⁹ Zur Entnahme einzelner Elemente aus vorhandenen Datenbanken siehe *Erler-Fridgen*, Teil 2, D. in diesem Band.

²⁰ Dafür kommen etwa Sammlungen gemeinfreier Primärtexte der Staatsbibliotheken in Betracht, zum Beispiel die der französischen Nationalbibliothek Gallica, auch Repositorien, die auf Open-Source Plattformen wie github (s.o.) veröffentlicht wurden, könnten dafür eingesetzt werden.

mationsgewinnung gesammelt werden. Diese können zum Beispiel nach erfolgter Annotation automatisiert ausgewertet werden (siehe Beispiel in Abbildung 2).

b) Urheberrechtliche Perspektive auf die Sammlung von Texten und Daten

Aus urheberrechtlicher Perspektive erzeugt das Sammeln und Speichern einzelner Text- und Datenquellen digitale Kopien und greift damit in das Vervielfältigungsrecht nach § 16 UrhG ein. Sind die Textquellen urheberrechtlich geschützt, so sind sie nach § 64 UrhG für einen Zeitraum von 70 Jahren nach dem Tod des Urhebers vor entsprechenden Verwertungshandlungen geschützt.²¹ Texte sind nach § 2 Abs. 2 UrhG urheberrechtlich geschützt, wenn sie eine persönliche geistige Schöpfung in wahrnehmbarer Form darstellen und ausreichende Individualität aufweisen.²² Darüber hinaus kann das Sammeln von Werken oder Daten aus Sammelbänden, etwa einer bibliographischen Sammlung, in den urheberrechtlichen Schutz des Sammelwerks nach § 4 UrhG eingreifen.²³ Frei nutzbar sind grundsätzlich gemeinfreie Werke nach § 64 UrhG und ggf. verwaiste Werke nach § 61 UrhG.²⁴ Urheberrechtlich geschützte Werke können mit Zustimmung des Urheberrechtsinhabers beispielsweise über Creative-Commons-Lizenzen genutzt werden, wobei jedoch deren Voraussetzungen beachtet werden müssen. Die Übersetzung an sich gemeinfreier Werke kann allerdings als Bearbeitung nach § 3 UrhG geschützt sein.²⁵

Außerdem kann die Sammlung o.g. verschiedener Texte und Daten aus Datenbanken einerseits den urheberrechtlichen Schutz des Datenbankwerks nach § 4 Abs. 2 UrhG verletzen.²⁶ Andererseits kann durch eine Entnahme auch in das

²¹ *Erler-Fridgen*, Teil 2, A. I. 1. in diesem Band.

²² Ausführlich hierzu: *Erler-Fridgen*, Teil 2, A, I. 1. in diesem Band.

²³ Zum urheberrechtlichen Schutz von Sammelwerken siehe *Erler-Fridgen*, Teil 2, A. in diesem Band.

²⁴ *Hagemeier*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 60d Rn. 14; siehe zur Gemeinfreiheit auch *Erler-Fridgen*, Teil 2, A, II. 1. in diesem Band.

²⁵ *De la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 16 f.

²⁶ *Erler-Fridgen*, Teil 2, D. in diesem Band.

sui generis Recht des Datenbankherstellers nach § 87a ff. UrhG eingegriffen werden, wenn wesentliche Teile einer durch wesentliche Investition geschaffenen Datenbank genutzt werden.²⁷

Die Nutzung wissenschaftlicher Editionen als Quelle für Primärtexte oder fachliterarischer Ergänzungsteile kann das Leistungsschutzrecht der wissenschaftlichen Ausgabe nach § 70 UrhG verletzen.²⁸ Solche wissenschaftliche Ausgaben sind nach § 70 UrhG für 25 Jahre nach ihrem Erscheinen leistungsschutzrechtlich geschützt.

Die Vervielfältigungen können bei rechtmäßigem Zugang zu den Werken auf Grundlage der Text und Data Mining-Schranken (§ 44b, 60d UrhG) zulässig sein.²⁹ § 44b Abs. 2 UrhG stellt es frei, bei rechtmäßigem Zugang zu den Werken, diese für die weitere Aufbereitung und Extraktion zusammenzustellen.³⁰ Ebenso umfasst § 60d UrhG bei rechtmäßigem Zugang das Recht, die Quellen zu speichern, um sie sodann aufzubereiten.³¹

2. Aufbereitung

Die verschiedenen Text- und Datenquellen müssen auf Basis einer Modellierung aufbereitet werden, um Texte mit informationstechnischen Mitteln untersuchen zu können. Als Grundlage der Aufbereitung erfolgen eine konzeptionelle Modellierung sowie eine Modellierung mit Blick auf die technische Umsetzung der eingesetzten Verfahren.³²

a) Modellierung

Ziel der Modellierung in den Digital Humanities ist es, relevante Elemente und Merkmale der Texte zu identifizieren. Diese sind dann die Basis für Mining-

²⁷ Näher zum *sui generis* Schutz des Datenbankherstellers siehe *Erler-Fridgen*, Teil 2, D. in diesem Band; siehe auch *Spindler*, GRUR 2016, 1112, 1113 f.

²⁸ *Erler-Fridgen*, Teil 2, C. in diesem Band.

²⁹ *Erler-Fridgen*, Teil 4, A. in diesem Band; *Hagemeyer*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 60d Rn. 15.

³⁰ *Raue*, ZUM 2021, 793, 795.

³¹ *Raue*, ZUM 2021, 793, 797.

³² Mit Blick auf die technische Umsetzung wird teilweise auch von logischer Modellierung gesprochen, siehe *Jannidis*, Grundlagen der Datenmodellierung, in Jannidis/Kohle/Rehbein (Hrsg.), Digital Humanities, 103.

Analyseverfahren. Hierfür werden Modelle eingesetzt, die die ausgewählten Elemente der Texte maschinenlesbar repräsentieren.³³ Die Auswahl der repräsentierten Elemente erfolgt konzeptionell und wird je nach Verwendungszweck für spätere Textanalysen getroffen.³⁴ Dabei liegt ein wesentlicher Forschungsschwerpunkt der Modellierung darin, dass sich verschiedene geisteswissenschaftliche Disziplinen auf eine für das jeweilige Forschungsvorhaben adäquate Modellierung verständigen.³⁵ Zugleich sollen Konzepte und Standards der Modellierung gewährleisten, dass Daten auch über den Projektkontext hinaus nutzbar sind. Es soll möglich sein, dass auf der entstandenen Datenbasis mit anderen Forschern zusammengearbeitet werden kann.³⁶ Außerdem ist oftmals vorgesehen, dass die geplante Präsentation der Daten in die Konzeption der Modellierung miteinfließt, insbesondere dadurch, dass Modelle und technische Verknüpfungsstrategien standardisiert werden.³⁷

Auf Basis der konzeptionellen wie technischen Modellierung erfolgt sodann die maschinenlesbare Aufbereitung mit Hilfe verschiedener Verfahren³⁸, die Auswahl der passenden Formate³⁹ sowie die Annotation der Texte.

³³ *Jannidis*, Grundlagen der Datenmodellierung, in Jannidis/Kohle/Rehbein (Hrsg.), *Digital Humanities*, 100 ff.

³⁴ Zur Auswahl und Strukturierung der Daten als „ersten hermeneutischen Akt“ siehe *Leif Scheuermann*, Die Abgrenzung der digitalen Geisteswissenschaften, in *Digital Classics Online 2* (2016), 1, 61; sowie auch: *Deck*, *Digital Humanities – Eine Herausforderung an die Informatik und an die Geisteswissenschaften*, in: Huber/Krämer (Hrsg.), *Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden*.

³⁵ *Thaller*, *Digital Humanities als Wissenschaft*, in Jannidis/Kohle/Rehbein (Hrsg.), *Digital Humanities*, 16; *McCarty*, *Modeling: A Study in Words and Meanings*, in Ray Siemens, John Unsworth, and Susan Schreibman, *A Companion to Digital Humanities*.

³⁶ *Scheuermann*, *Die Abgrenzung der digitalen Geisteswissenschaften*, in *Digital Classics Online 2* (2016), 1, 61.

³⁷ Zu den dahingehenden Funktionen der Datenmodelle: *Jannidis*, *Grundlagen der Datenmodellierung*, in Jannidis/Kohle/Rehbein (Hrsg.), *Digital Humanities*, 100.

³⁸ *Optical Character Recognition (OCR): Verfahren zur Textdigitalisierung*, in dem der Prozess der Texterfassung automatisiert vorgenommen wird, siehe <http://digitalhumanities.berkeley.edu/resources/digitization-workflows-scanning-ocr-and-audio-transcription> (zuletzt abgerufen am 12.01.2022) und *Double Keying Verfahren: Double Keying* beschreibt ein Verfahren zur Textdigitalisierung, in dem der Prozess der Texterfassung manuell vorgenommen wird, siehe <http://www.dhmuseum.uni-trier.de/node/49>, zuletzt abgerufen am 12.01.2022.

³⁹ Beispielsweise werden Texte im Format XML (Extensible Markup Language, eine Auszeichnungssprache, die zur Annotation eingesetzt werden kann und als Datenaustauschformat eine einheitliche Repräsentation in unterschiedlichen Applikationskontexten ermöglicht, siehe *Rhem*, 2.5 Texttechnologische Grundlagen, in K.-U. Carstensen, Ch. Ebert, C. Ebert, S. Jekat, R. Klabunde, H. Langer (Hrsg.), *Computerlinguistik und Sprachtechnologie*, 3. Aufl., 159.)

b) Annotationen

Annotation meint das Setzen von ergänzenden Hinweisen, also die Kommentierung einzelner Token im Text als sogenannter Beitzext.⁴⁰ Je nach Einsatz stellen Annotationen einerseits einen wesentlichen wissenschaftlichen Zwischenschritt zwischen Aufbereitung und Informationsextraktion dar. Andererseits können mit ihrer Hilfe auch Ergebnisse der Informationsextraktion aufgenommen werden. Zum Teil werden Annotationen manuell vorgenommen, etwa um technische Merkmale des Digitalisats sowie linguistische Merkmale des Textes hervorzuheben. Beispielsweise können dabei die Haltung des Aussagegebenden zum Satz, Quellenangaben sowie Aussagen zur Beziehung aus Kookkurrenz⁴¹ manuell annotiert werden. Weiterhin werden automatisiert extrahierte Informationen durch Annotationen als Kommentierung am Text ergänzt. Hier zeigt sich der iterative Prozess der Verfahrensschritte, Aufbereitung und Informationsextraktion, an deren Schnittstellen Annotationen genutzt werden.

c) Urheberrechtliche Perspektive auf die Aufbereitung

Bei der Aufbereitung der gesammelten Text- und Datenbestände können urheberrechtlich relevante Handlungen vorgenommen werden. Werden nämlich in diesem Verfahrensschritt Kopien erzeugt oder die Dateien im Arbeitsspeicher zwischengespeichert, so kann dies in das Vervielfältigungsrecht nach § 16 UrhG eingreifen. Beispielsweise können das Scannen der Texte,⁴² das Digitalisieren⁴³

oder im Format RDF (Resource Description Framework, ist ein Standard zur Annotation von Metadaten in der Form von Triplets (Subjekt, Prädikat, Objekt), *Rhem*, 2.5 Texttechnologische Grundlagen, in K.-U. Carstensen, Ch. Ebert, C. Ebert, S. Jekat, R. Klabunde, H. Langer (Hrsg.), *Computerlinguistik und Sprachtechnologie*, 3. Aufl., 166) erschlossen.

⁴⁰ *Rapp*, Manuelle und automatische Annotation, in Jannidis/Kohle/Rehbein (Hrsg.), *Digital Humanities*, 253 f.

⁴¹ Als Kookkurrenz wird das gemeinsame Vorkommen zweier oder mehrerer Wörter in einem Kontext von fest definierter Größe bezeichnet, *Kunze/Lemnitzer*, *Computerlexikographie. Eine Einführung*, 391f.; Näheres zur Kookkurrenzanalyse: *Engelberg*, *Kookkurrenzanalyse, Linguistische Methodenlehre*, FS 2009, Uni Mannheim, https://www1.ids-mannheim.de/fileadmin/lexik/lehre/engelberg/Webseite_LingMeth/Skript_05.pdf, zuletzt abgerufen am 12.01.2022.

⁴² BGH GRUR 2002, 246, 247 – Scanner.

⁴³ Zur Erzeugung eines Datensatzes mittels eines Digitalisierungsgeräts: KG ZUM 2002, 828, 830 – Pressespiegelversand.

der Texte, etwa mittels Optical Character Recognition (OCR)⁴⁴, aber auch das Speichern⁴⁵ nach erfolgter Annotation der Texte oder die Umwandlung der Dateiformate relevante Vervielfältigungshandlungen nach § 16 UrhG sein. Solche Vervielfältigungen zur Aufbereitung von Texten sind jedoch freigestellt, wenn die Voraussetzungen der Text und Data Mining-Schranken (§ 44b/§ 60d UrhG) vorliegen.⁴⁶ Daneben stellt § 44a UrhG lediglich vorübergehende Vervielfältigungen frei, die der technischen Übermittlung dienen und keine eigene wirtschaftliche Bedeutung aufweisen.⁴⁷ Insbesondere bedeutet vorübergehend in diesem Kontext, dass die Daten nach einer nicht ins Gewicht fallenden Zeit automatisch wieder gelöscht werden müssen.⁴⁸

aa) Maschinenlesbare Aufbereitung

Wie geschildert, reicht jedoch die reine Vervielfältigung in Form von Kopien oftmals nicht aus, sondern es kann notwendig sein, die Texte etwa in eine maschinenlesbare Form umzuwandeln.⁴⁹ Würde das eine Bearbeitung darstellen, dann wäre diese zwar ohne Zustimmung der rechteinhabenden Person möglich, jedoch wäre nach § 23 UrhG die Verwertung, also auch die Vervielfältigung, nicht ohne eine solche Zustimmung zulässig.⁵⁰ Das Gesetz besagt jedoch in § 23 Abs. 3 UrhG, dass „ausschließlich technisch bedingte Änderungen eines Werkes nach § 44b Abs. 2, 60d Abs. 1, § 60e Abs. 1 sowie § 60f Abs. 2 UrhG“ als Bearbeitung nach § 23 UrhG nicht erfasst werden.⁵¹ Damit stellt § 23 Abs. 3 UrhG klar, dass die Digitalisierung⁵², die Umwandlung in ein maschinenlesbares Format sowie die rein technisch bedingte Anreicherung mit Metadaten wohl keine Bearbeitung im Sinne des § 23 UrhG darstellen.⁵³ Hierfür spricht auch, dass insbesondere bei der Digitalisierung das geistige Wesen des Werkes unverändert

⁴⁴ Verfahren zur Textdigitalisierung, in dem der Prozess der Texterfassung automatisiert vorgenommen wird, siehe <http://digitalhumanities.berkeley.edu/resources/digitization-workflows-scanning-ocr-and-audio-transcription>, zuletzt abgerufen am 10.03.2020.

⁴⁵ Landgericht München I ZUM-RD 2003, 607, 610 – Vervielfältigung von Musiktiteln.

⁴⁶ *Erler-Fridgen*, Teil 4, A. in diesem Band.

⁴⁷ *Dreier*, in *Dreier/Schulze, Urheberrechtsgesetz*, 7. Aufl. 2022, § 44a Rn. 4 ff.

⁴⁸ EuGH C-5/08, ECLI:EU:C:2009:465, GRUR Int 2010, 35 Rn. 64 – Infopaq; *Dreier*, in *Dreier/Schulze, Urheberrechtsgesetz*, 7. Aufl. 2022, § 44a Rn. 4.

⁴⁹ *Spindler*, GRUR 2016, 1112, 1113.

⁵⁰ *Spindler*, GRUR 2016, 1112, 1113.

⁵¹ *De la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 16 f.

⁵² *Loewenheim*, in *Schricker/Loewenheim, Urheberrecht*, 6. Auflage 2020, § 23 Rn. 7; *Ernst*, in *Hoeren/Sieber/Holznel, Multimedia-Recht*, 7.1 Rn. 51.

⁵³ *Specht*, OdW 2018, 285.

bleibt.⁵⁴ Wie diese Norm im Verhältnis zum unionsrechtlichen Vervielfältigungsbegriff steht, wird in der rechtswissenschaftlichen Literatur diskutiert und ist nicht abschließend geklärt.⁵⁵

bb) Rechte der Forschenden durch Aufbereitung

Ob durch die Aufbereitung von Texten eigene Rechte für die Forschenden entstehen, ist fraglich und wird in vielen Forschungsgruppen eine wichtige Rolle spielen.⁵⁶ Gerade Anreicherungen mit Metainformationen wie Annotationen werden regelmäßig in Forschungsprojekten unter hohem Aufwand erstellt. Für den urheberrechtlichen Schutz von Annotationen als Bearbeitung (§ 3 UrhG) wäre erforderlich, dass die Aufbereitung die Schwelle hinreichender Individualität überschreitet.⁵⁷ Für Ergänzungen an einem vorhandenen Text gilt: Je mehr sich der Gestaltungsspielraum durch Ergänzungen am Text erweitert, desto eher könnte Individualität vorliegen.⁵⁸ Ob Annotationen als Ergänzungen oder Abänderungen des Ausgangstextes und bei hinreichender Individualität als schutzfähige Bearbeitungen gewertet werden können, ist zweifelhaft. Teilweise wird vertreten, dass linguistische Annotationen bei erforderlicher Schöpfungshöhe Bearbeitungen im Sinne von § 3 UrhG darstellen können.⁵⁹ Hierbei müsse jedoch zwischen vorbereitenden Annotationsschemata, entsprechender Standardsetzung und der Durchführung der Annotation am Text unterschieden werden: es sei denkbar, dass Annotationsschemata Schöpfungshöhe erreichen, die Durchführung am Text als solche jedoch nicht.⁶⁰ An-

⁵⁴ Siehe zur Digitalisierung und Bearbeitung allgemein *Loewenheim*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 23 Rn. 7.

⁵⁵ *De la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 18; ausführlich *Raue*, AfP 2022, 1 ff.; zur Diskussion *Obly*, GRUR 2017, 964, 967 mwN: Der Ansicht nach, der die ständige Rechtsprechung des BGH entspreche, wäre die Bearbeitung ein Unterfall der unionsrechtlichen Vervielfältigung.

⁵⁶ *De la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 18.

⁵⁷ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 3 Rn. 11.

⁵⁸ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 3 Rn. 17.

⁵⁹ *Lehmberg, Chiarcos, Rbem, Witt*, Rechtsfragen bei der Nutzung und Weitergabe linguistischer Daten, in Rehm/Witt/Lemmitzer, Datenstrukturen für linguistische Ressourcen und ihre Anwendungen, 93, 98.

⁶⁰ *Lehmberg, Chiarcos, Rbem, Witt*, Rechtsfragen bei der Nutzung und Weitergabe linguistischer Daten, in Rehm/Witt/Lemmitzer, Datenstrukturen für linguistische Ressourcen und ihre Anwendungen, 93, 98.

derer Ansicht nach liegt bei der Anreicherung mit Metainformationen keine Bearbeitung vor.⁶¹ Denn die angefügten Metainformationen würden nicht das eigentliche Werk berühren, sondern es nur anreichern und schematisieren.⁶² Hierin liege gerade keine Veränderung des ursprünglichen Werkes, die Informationen charakterisieren und ordnen das Werk lediglich ein.⁶³ Es liegt keine höchstichterliche Rechtsprechung zur Beurteilung von linguistischen Annotationen als Bearbeitungen bzw. deren Schöpfungshöhe vor.⁶⁴ Festzuhalten ist: eine Bearbeitung erfordert eine Veränderung der schutzfähigen Merkmale des ursprünglichen Werks, während die Übernahme ohne Veränderung eine Vervielfältigung darstellt.⁶⁵ In Abgrenzung dazu legt § 23 Abs. 3 UrhG jedenfalls die Annahme nahe, dass eine ausschließlich technisch bedingte Anreicherung mit Metadaten ebenso wie die Digitalisierung bereits keine Bearbeitung im Sinne des § 23 Abs. 1 UrhG darstellt.⁶⁶

3. Informationsextraktion

Die Informationsextraktion kann mittels verschiedener Analysemethoden durchgeführt werden. Zunächst geht es beim Text und Data Mining darum, verborgene Zusammenhänge und Strukturen im Text aufzufinden. Quantitative Verfahren der Informationsextraktion sind beispielsweise statistische Analysen, das maschinelle Lernen, Netzwerkanalysen sowie das Topic Modeling⁶⁷ und die Sentiment Analyse⁶⁸.⁶⁹ Je nach Art der Textquelle können im konkreten Projektkontext unterschiedliche Analyseverfahren eingesetzt werden. Auf Grundlage der Ergebnisse – etwa nach der Topic Modeling Analyse – können

⁶¹ *Spindler*, GRUR 2016, 1112, 1114.

⁶² *Spindler*, GRUR 2016, 1112, 1114: Dies entspreche in der analogen Welt am ehesten einer Exzerption von Teilen eines Werkes mit Hilfe bestimmter Schlagwörter oder Indizes.

⁶³ *Spindler*, Text und Data Mining – urheber- und datenschutzrechtliche Fragen, GRUR 2016, 1112, 1114.

⁶⁴ *Lehmberg, Chiarcos, Rhem, Witt*, Rechtsfragen bei der Nutzung und Weitergabe linguistischer Daten, in Rehm/Witt/Lemmitzer, Datenstrukturen für linguistische Ressourcen und ihre Anwendungen, 93, 98.

⁶⁵ *Loewenheim*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 23 Rn. 6.

⁶⁶ *Specht*, OdW 2018, 285.

⁶⁷ *Blei*, Probabilistic Topic Models, <http://www.cs.columbia.edu/~blei/papers/Blei2012.pdf>, zuletzt abgerufen am 12.01.2022.

⁶⁸ *Ignatow/Mibalcea*, Text Mining, 148 ff.

⁶⁹ Ausführlich zu den unterschiedlichen Verfahrensmethoden siehe *Schöb*, Quantitative Analyse, in Jannidis/Kohle/Rehbein (Hrsg.), Digital Humanities, 279 ff.

die Texte dann wiederum annotiert werden. Annotierte Texte können umgekehrt aber auch Trainingsdaten für weiterführende Analysen und damit die Grundlage für zusätzliche Informationsextraktion bieten. Ergänzend kann die Auffindbarkeit bereits enthaltener Informationen mittels Information Retrieval sichergestellt werden.⁷⁰

Grundsätzlich ist die Extraktion von Informationen aus urheberrechtlich geschützten Quellen als solche nicht vom Zuweisungsgehalt des Urheberrechts erfasst.⁷¹ Grundsätzlich gilt dies auch dann, wenn die Extraktion mit technischen Mitteln wie etwa Algorithmen erfolgt.⁷² Erfordert die Extraktion von Informationen es jedoch, dass Kopien erzeugt oder die Dateien in den Arbeitsspeicher geladen werden, was in der Regel der Fall sein wird, so kann aus urheberrechtlicher Perspektive auch dieser Verfahrensschritt zu Vervielfältigungshandlungen nach § 16 UrhG führen.⁷³ Für das Text und Data Mining vorgenommenen Vervielfältigungshandlungen fallen unter die Text und Data Mining-Schranken nach § 44b UrhG oder § 60d UrhG und sind unter deren Voraussetzungen freigestellt.⁷⁴ Voraussetzung ist jedoch insbesondere der rechtmäßige Zugang zu den genutzten Werken.⁷⁵ Hiernach zulässig sind Vervielfältigungen zum Zweck des Text und Data Minings, sodass neben der automatisierten Analyse, die Quellen auch zuvor gespeichert, normalisiert, annotiert oder auf sonstige Weise bearbeitet (§ 23 Abs. 3 UrhG) werden können (zur Aufbereitung siehe III.).⁷⁶ Die Schranke der vorübergehenden Vervielfältigungen nach § 44a UrhG setzt hingegen voraus, dass die Kopien zeitlich beschränkt sind, der technischen Übermittlung dienen und keine eigene wirtschaftliche Bedeutung aufweisen.⁷⁷

⁷⁰ *Klinke*, Information Retrieval, in Jannidis/Kohle/Rehbein (Hrsg.), Digital Humanities, 268 f.

⁷¹ *Raue*, ZUM 2019, 684, 686.

⁷² *Raue*, GRUR 2017, 11, 13; *De la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 18.

⁷³ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 16 Rn. 7.

⁷⁴ Zur Zulässigkeit von Vervielfältigungen bei Text und Data Mining siehe *Erler-Fridgen*, Teil 4, A. in diesem Band; siehe zur Freistellung *de la Durantaye/Raue*, Teil 1, A. in diesem Band, 9, 18.

⁷⁵ *Erler-Fridgen*, Teil 4, A. in diesem Band.

⁷⁶ *Raue*, ZUM 2021, 793, 797 und 798.

⁷⁷ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 44a Rn. 4 ff.: in zeitlicher Hinsicht müssen die Daten nach einer nicht ins Gewicht fallenden Zeit automatisch gelöscht werden.

4. Präsentation

Die Präsentation der extrahierten Informationen umfasst mehrere Grundentscheidungen. In welchem Umfang sollen die Daten veröffentlicht werden? Wie frei sollen die Daten verfügbar sein? Sind personenbezogene Daten mitenthalten?⁷⁸ Wie soll die Benutzeroberfläche gestaltet sein? In welchem Format sollen die Daten veröffentlicht werden?

Dabei spielt die Nachnutzbarkeit vorhandener Forschungsergebnisse bzw. die Kombinationsmöglichkeit mit vorhandenen Datenbeständen eine wesentliche Rolle. Insbesondere Anbindungsmöglichkeiten der gewonnenen Informationen an vorhandene externe Netzwerke können die Nachnutzbarkeit der Informationen erhöhen. Beispielsweise können extrahierte Daten als Linked Open Data verfügbar gemacht werden, wodurch die Vernetzung mit weiteren Informationen möglich ist.⁷⁹ Nachnutzungsoptionen frei verfügbarer Datenbestände durch Dritte müssen dabei mitgedacht werden: Sollen die Datenbestände durch andere Nutzer ergänzbar sein?⁸⁰ Werden im Projekt beispielsweise Tools auf den Datensätzen trainiert,⁸¹ dann steht auch die Präsentation eines solchen Tools und die Möglichkeit in Frage, dieses in anderen Kontexten einzusetzen.

Werden die Ergebnisse der Analyse präsentiert und veröffentlicht, so kann aus urheberrechtlicher Perspektive das Recht auf öffentliche Zugänglichmachung nach § 19a UrhG beeinträchtigt werden. Insbesondere wenn die Volltexte oder längere Textteile mitveröffentlicht werden, steht eine solche Verletzung in Rede. Eine Präsentation der zugrundeliegenden Texte kann für die Forschenden interessant sein, da diese als Nachweise⁸² der gewonnenen Erkenntnisse sowie zur Nachnutzung dienen können.⁸³ Der hierfür vorausgesetzte urheberrechtliche Schutz entfällt jedoch bei gemeinfreien Texten, wie beispielsweise Romanen

⁷⁸ Auf mögliche datenschutzrechtliche Implikationen wird in der vorliegenden Handreichung nicht eingegangen.

⁷⁹ Linked Open Data bedeutet, dass auf einer Ontologie basierte Datenbanken veröffentlicht werden und zur Vernetzung die Technologie des Semantic Web genutzt wird, *Rehbein*, *Ontologien in den Digital Humanities*, in Jannidis/Kohle/Rehbein (Hrsg.), *Digital Humanities*, 173.

⁸⁰ Beispielsweise könnte die Einbindung in ein bestehendes Wissensnetzwerk wie das Wikidata-Netzwerk erwogen werden, <https://www.wikidata.org/wiki/Wikidata:Introduction>, zuletzt abgerufen am 12.01.2022.

⁸¹ Beispielsweise kann für die Verbesserung der maschinenlesbaren Aufbereitung mittels Optical Recognition Verfahren ein Tool auf den Texten trainiert werden.

⁸² Zum Zitatrecht und Nachweisen von Textteilen bei Textanalysen siehe *Erler-Fridgen*, Teil 4, C. in diesem Band.

⁸³ *Spindler*, GRUR 2016, 1112, 1113.

aus der zweiten Hälfte des 18. Jahrhunderts, wegen Zeitablaufs (§ 64 UrhG).⁸⁴ Werden Bestandteile wissenschaftlicher Ausgaben mitveröffentlicht, kann hierdurch das Leistungsschutzrecht wissenschaftlicher Ausgaben nach § 70 UrhG verletzt werden.⁸⁵ Werden Textteile, wie einzelne Kapitel, Seiten oder auch Sätze, veröffentlicht, kann das Urheberrechte verletzen, sofern diese Werkteile urheberrechtlich nach § 2 UrhG geschützt werden.⁸⁶ Die Text und Data Mining Schranke des § 60d Abs. 4 UrhG ermöglicht die öffentliche Zugänglichmachung⁸⁷ von Vervielfältigungen für einen abgegrenzten Kreis von Personen zum Zweck deren gemeinsame wissenschaftliche Forschung.⁸⁸ Außerdem stellt sie dies auch einzelnen Dritten zur Überprüfung der Qualität der wissenschaftlichen Forschung frei.⁸⁹ Hier ist jedoch keine allgemein zugängliche Präsentation, sondern nur eine an einen abgegrenzten Personenkreis bzw. an einzelne Dritte vorgenommene Präsentation möglich.

Außerdem ist es möglich, dass an einer Sammlung von Informationen das *sui generis*-Leistungsschutzrecht des Datenbankherstellers nach § 87a ff. UrhG – u.a. bei wesentlicher Investition – entsteht.⁹⁰ Liegt eine schöpferische Leistung in Auswahl oder Anordnung der Datenbankelemente vor, so kommt auch ein Schutz der präsentierten Sammlung als Datenbankwerk nach § 4 Abs. 2 UrhG in Betracht.⁹¹

5. Archivierung

Die Archivierung der extrahierten, analysierten und vernetzten Forschungsergebnisse und Daten ist Voraussetzung für die Überprüfbarkeit und Nach-

⁸⁴ Zum urheberrechtlichen Schutz siehe *Erler-Fridgen*, Teil 2, A. in diesem Band.

⁸⁵ Zum Leistungsschutzrecht wissenschaftlicher Ausgaben siehe *Erler-Fridgen*, Teil 2, C. in diesem Band.

⁸⁶ Zum urheberrechtlichen Schutz von Textteilen siehe *Erler-Fridgen*, Teil 6, A. in diesem Band.

⁸⁷ In der Regel ist bei der Zugänglichmachung für einen abgegrenzten Personenkreis ohnehin kein öffentliches Zugänglichmachen iSv. § 19a UrhG gegeben: *Raue*, ZUM 2021, 793, 799; zum Begriff der „öffentlichen Zugänglichmachung“ auch *Erler-Fridgen*, Teil 4, A, IV. in diesem Band.

⁸⁸ Zu § 60d Abs. 4 UrhG und dem Begriff der „öffentlichen Zugänglichmachung“ siehe *Erler-Fridgen*, Teil 4, A, IV. in diesem Band.

⁸⁹ Siehe *Erler-Fridgen*, Teil 4, A. in diesem Band.

⁹⁰ Siehe hierzu *Erler-Fridgen*, Teil 2, D. in diesem Band.

⁹¹ Siehe *Erler-Fridgen*, Teil 2, D. in diesem Band.

nutzbarkeit der erzielten Forschungsdaten. Mit dem Ziel der Langzeitarchivierung soll die langfristige Nachnutzbarkeit von Daten trotz technischen Wandels ermöglicht werden.⁹² Auch hier sind viele grundsätzliche Entscheidungen zu treffen: In welchem Umfang und auf Basis welchen Projektstandes wird archiviert? Werden personenbezogene Forschungsdaten gespeichert?⁹³ Welche Metadaten, beispielsweise zu angewandten Forschungsmethoden, Soft- und Hardware etc., werden mitarchiviert? Hierbei müssen auch die passenden Formate der Metadatensätze und Archivdatensätze ausgewählt werden.⁹⁴

Sofern bei der Archivierung der erzielten Forschungsdaten auch Kopien erzeugt werden und (Teile der) Ausgangstexte davon betroffen sind, kann auch an dieser Stelle aus urheberrechtlicher Perspektive in das Vervielfältigungsrecht nach § 16 UrhG eingegriffen werden. Die Archivierung kann beispielsweise auf einem projekt-eigenen Server oder mittels Cloud-Repositories (geschlossen oder frei verfügbar)⁹⁵ vorgesehen sein. Werden die Forschungsdaten wie auch Volltexte oder Textteile frei verfügbar archiviert, so kann auch hierdurch in das Recht auf öffentliche Zugänglichmachung nach § 19a UrhG eingegriffen werden. Auch die Zugänglichmachung der Vervielfältigungen für Text und Data Mining an einen abgeschlossenen Kreis – wie sie die Text und Data Mining-Schranke nach § 60d Abs. 4 UrhG zulässt⁹⁶ – ist nach Abschluss der gemeinsamen wissenschaftlichen Forschung bzw. Überprüfung der Qualität der wissenschaftlichen Forschung zu beenden.⁹⁷ Nach der alten Rechtslage in § 60d UrhG alte Fassung (a.F.) musste Ursprungsmaterial und Korpus nach Abschluss der Forschungsarbeiten gelöscht werden.⁹⁸ Nunmehr ist jedoch eine Archivierung der Vervielfältigungen bei der Forschungseinrichtung selbst nach § 60d Abs. 5 UrhG mög-

⁹² <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/langzeitarchivierung/>, zuletzt abgerufen am 12.01.2022.

⁹³ Auf mögliche datenschutzrechtliche Implikationen wird in der vorliegenden Handreichung nicht eingegangen.

⁹⁴ *Schumann*, Einführung in die digitale Langzeitarchivierung, 44 https://www.ssoar.info/ssoar/bitstream/handle/document/45740/ssoar-2012-schumann-Einfuehrung_in_die_digitale_Langzeitarchivierung.pdf, zuletzt abgerufen am 12.01.2022.

⁹⁵ Cloud-Repositories sind beispielsweise die OpenAccess Cloud Zenodo (<https://about.zenodo.org/>, zuletzt abgerufen am 12.01.2022) oder die geschlossen oder frei zugänglich nutzbare Infrastruktur der github-Plattform.

⁹⁶ Im Regelfall ist im Zugänglichmachen an einen abgegrenzten Personenkreis jedoch wohl kein öffentliches Zugänglichmachen nach § 19a UrhG zu sehen: *Raue*, ZUM 2021, 793, 799.

⁹⁷ Zur Text und Data Mining Schranke siehe *Erler-Fridgen*, Teil 4, A. in diesem Band.

⁹⁸ *Raue*, ZUM 2021, 793, 799.

lich. Es ist hiernach zulässig, entsprechende Vervielfältigungen mit angemessenen Sicherheitsvorkehrungen aufzubewahren, solange sie für die Zwecke der wissenschaftlichen Forschung oder zur Überprüfung wissenschaftlicher Erkenntnisse erforderlich sind.⁹⁹ Das bedeutet, dass Datenkorpora für Anschlussforschungen genutzt werden können, jedoch als Voraussetzung angemessene Sicherheitsvorkehrungen gegen Missbrauch getroffen werden müssen.¹⁰⁰ Aus diesem Grund wird in der rechtswissenschaftlichen Literatur empfohlen, die Daten in Forschungsdatenrepositorien zu speichern, die die Sicherheitsstandards zur Verfügung stellen können.¹⁰¹ Entsprechend den Grundsätzen guter wissenschaftlicher Praxis ist die Aufbewahrung zur Überprüfung wissenschaftlicher Erkenntnisse jedenfalls zehn Jahre lang erforderlich.¹⁰² Dass eine längere Aufbewahrungsdauer erforderlich ist, müssen die Forschenden beispielsweise unter Verweis auf Anschlussforschung ausreichend plausibel machen.¹⁰³

6. Ergebnis

Die iterativen Verfahrensschritte – Sammlung, Aufbereitung, Extraktion, Präsentation und Archivierung – greifen beim Text und Data Mining ineinander. Insbesondere die Modellierung setzt konzeptionell das Mitdenken der übrigen Verfahrensschritte voraus. Auch die Annotationen können die Schritte der Aufbereitung und Extraktion miteinander verschränken. In jedem Verfahrensschritt sind urheberrechtlich relevante Handlungen – Vervielfältigungen nach § 16 UrhG oder das öffentliche Zugänglichmachen nach § 19a UrhG – zu identifizieren. Vervielfältigungen, die für das Text und Data Mining vorgenommen werden, sind nach §§ 44b/60d UrhG unter deren Voraussetzungen zulässig. Die Text und Data Mining-Schranken ermöglichen jedoch

⁹⁹ Hierzu näher siehe *Erler-Fridgen*, Teil 4, A. in diesem Band.

¹⁰⁰ *Raue*, ZUM 2021, 793, 799; die genaue Ausgestaltung angemessener Sicherheitsvorkehrungen ist zum jetzigen Zeitpunkt nicht geklärt.

¹⁰¹ *Raue*, ZUM 2021, 793, 799; zu solchen vertrauenswürdigen Stellen siehe Erwägungsgrund 15 S. 3 der Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG (DSM-RL).

¹⁰² *Raue*, ZUM 2021, 793, 799 mit Blick auf: DFG, Leitlinien zur Sicherung guter wissenschaftlicher Praxis, Kodex, 2019, Leitlinie 12, 13, insbesondere 17; Argumente für eine längere Aufbewahrungsdauer: *Heesen/Jüngels*, RuZ 2021, 45, 50 unter einer CC BY-SA Lizenz.

¹⁰³ *Raue*, ZUM 2021, 793, 799: Forschende haben hier eine Einschätzungsprärogative, die gerichtlich lediglich eingeschränkt auf Missbrauch überprüft werden kann; dazu auch *Erler-Fridgen*, Teil 4, A. in diesem Band.

nicht, Ausgangstexte allgemein zu präsentieren. Nach § 60d Abs. 4 UrhG ist lediglich die öffentliche Zugänglichmachung an einen abgegrenzten Personenkreis für spezifische Zwecke zulässig.¹⁰⁴ Die Aufbewahrung von Daten bei der Forschungseinrichtung selbst ist nunmehr nach § 60d Abs. 5 UrhG möglich, solange sie für den Zweck der wissenschaftlichen Forschung oder für die Überprüfung wissenschaftlicher Erkenntnisse erforderlich ist.

Vertiefende Literaturhinweise: *De la Durantaye/Raue*, Urheberrecht und Zugang in einer digitalen Welt, Teil 1, A. in diesem Band; *Specht*, Die neue Schrankenregelung für Text und Data Mining und ihre Bedeutung für die Wissenschaft, *Ordnung der Wissenschaft (OdW)* 2018, 285; *Raue*, Die Freistellung von Datenanalysen durch die neuen Text und Data Mining-Schranken (§§ 44b, 60d UrhG), *Zeitschrift für Urheber- und Medienrecht (ZUM)* 2021, 793.

¹⁰⁴ Zum Begriff der „öffentlichen Zugänglichmachung“ in § 60d Abs. 4 UrhG siehe *Erler-Fridgen*, Teil 4, A. IV. in diesem Band.

Teil 2: Urheberrechtlicher Schutz der Ausgangsmaterialien

A. Kriterien der urheberrechtlichen Schutzfähigkeit von Texten und Sammelwerken

Katharina Erler-Fridgen

Version 1.0 (23.06.2021), CC BY-SA 4.0.

Werden Texte und Sammlungen von Texten oder Daten gesammelt, aufbereitet und hieraus Informationen extrahiert, sind urheberrechtliche Rahmenbedingungen zu beachten. Sind die genutzten Texte und Sammlungen urheberrechtlich geschützt, können Kopiervorgänge Vervielfältigungshandlungen nach § 16 UrhG begründen. Auch kann in das Recht auf öffentliche Zugänglichmachung nach § 19a UrhG eingegriffen werden, wenn die Ausgangstexte gemeinsam mit den Auswertungsergebnissen im Netz oder für weitere Forschende präsentiert werden. Im Folgenden sollen daher die Kriterien der urheberrechtlichen Schutzfähigkeit von Texten dargestellt, die Unterschiede verschiedener Textarten erläutert und auf den Schutz von Sammelwerken eingegangen werden.

I. Kriterien der urheberrechtlichen Schutzfähigkeit von Texten

1. Allgemeine Kriterien des Werkschutzes

Zunächst ist der urheberrechtliche Werkschutz nach § 2 UrhG zeitlich begrenzt und erlischt 70 Jahre nach dem Tod des Urhebers nach § 64 UrhG.¹ Beispielsweise ist diese Schutzfrist bei französischen Romanen aus der zweiten Hälfte des 18. Jahrhunderts überschritten, diese sind demzufolge *gemeinfrei*. Das bedeutet, dass Werke frei verwertet werden können, ohne dass die Zustimmung des Urhebers oder seines Rechtsnachfolgers eingeholt werden muss.² Innerhalb dieses Schutzzeitraums sind Werke nach § 2 UrhG dann geschützt, wenn die folgenden Voraussetzungen erfüllt sind.

¹ Steht das Urheberrecht mehreren Miturhebern (§ 8 UrhG) zu, beträgt die Schutzfrist nach § 65 Abs. 1 UrhG 70 Jahre nach dem Tod des längstlebenden Miturhebers.

² *Katzenberger/Metzger*, in *Schricker/Loewenheim, Urheberrecht*, 6. Aufl. 2020, § 64 Rn. 52.

a) Persönliche Schöpfung

Voraussetzung des Schutzes urheberrechtlicher Werke nach § 2 Abs. 2 UrhG³ ist das Vorliegen einer persönlichen geistigen Schöpfung.⁴ Hierzu muss eine persönliche und individuelle Schöpfung vorgenommen werden, der ein geistiger Gehalt in einer wahrnehmbaren Form innewohnt.⁵ Eine *persönliche Schöpfung* liegt dann vor, wenn ein Werk von einem Menschen geschaffen worden ist.⁶ Wird Technik als vom Menschen gesteuertes Hilfsmittel eingesetzt, um ein Werk zu generieren, steht dies dem Urheberrechtsschutz zunächst nicht entgegen.⁷ Rein durch Computer erstellte Gegenstände ohne Einwirkung des Menschen genießen jedoch keinen Urheberrechtsschutz. Daher können zum Beispiel computergenerierte Übersetzungen vom Urheberrechtsschutz ausgeschlossen sein.⁸ Beispielsweise entsteht demzufolge auch kein eigener, neuer Werkschutz eines digitalisierten Textes, der durch ein Tool automatisch erstellt wurde.⁹

³ Auch für Sammelwerke und Datenbanken als eigenständige Werke setzt § 2 Abs. 2 UrhG für den Schutz eine Manifestation als persönliche geistige Schöpfung, hier speziell mit Blick auf die Auswahl und Anordnung der Werke oder Daten, voraus, siehe *Dreier*, in *Dreier/Schulze, Urheberrechtsgesetz*, 6. Aufl. 2018, § 4 Rn. 11.

⁴ Zum europäischen Werkbegriff, der „eigenen geistigen Schöpfung“ siehe bereits EuGH C-5/08, ECLI:EU:C:2009:465, GRUR 2009, 1041 Rn. 51 – Infopaq/DDF, siehe aktuell EuGH C-310/17, ECLI:EU:C:2018:899, GRUR 2019, 73 Rn. 35 ff. – Levola Hengelo; ausführlich zum unionsrechtlichen Rahmen des Werkbegriffs: *Leistner*, in *Schricker/Loewenheim, Urheberrecht*, 6. Aufl. 2020, § 2 Rn. 3 ff.

⁵ Ausführlich zu den Kriterien der Schutzfähigkeit *Jotzo*, Teil 2, B. in diesem Band, 61, 64 ff und *Schack*, *Urheber- und Urhebervertragsrecht*, 9. Aufl. 2019, Rn. 181 ff., siehe auch *A. Nordemann*, in *Fromm/Nordemann, Urheberrecht*, 12. Aufl. 2018, § 2 Rn. 20 ff.

⁶ *Loewenheim/Leistner*, in *Schricker/Loewenheim, Urheberrecht*, 6. Aufl. 2020, § 2 Rn. 39.

⁷ *A. Nordemann*, in *Fromm/Nordemann, Urheberrecht*, 12. Aufl. 2018, § 2 Rn. 21 mwN, werden klare Anweisungen an die Maschine eindeutig geplant und festgelegt, so bestehe kein Zweifel am menschlichen Schaffen, so *Loewenheim/Leistner*, in *Schricker/Loewenheim, Urheberrecht*, 6. Aufl. 2020, § 2 Rn. 40.

⁸ *Schulze*, in *Dreier/Schulze, Urheberrechtsgesetz*, 6. Aufl. 2018, § 2 Rn. 8.

⁹ Zum ähnlichen Beispiel einer computergenerierten Übersetzung siehe *Loewenheim/Leistner*, in *Schricker/Loewenheim, Urheberrecht*, 6. Aufl. 2020, § 2 Rn. 39 mwN.

b) Wahrnehmbare Form

§ 2 Abs. 2 UrhG setzt voraus, dass die Ausdrucksform des Werkes der *menschlichen Wahrnehmung* zugänglich und dabei genau und objektiv identifizierbar ist.¹⁰ Es genügt irgendeine Ausdrucksform des Werkes, eine dauerhafte Form ist hingegen nicht erforderlich.¹¹ Selbst mündliche Spontanäußerungen in einem Interview können als Werk schutzfähig sein, wenn die übrigen Voraussetzungen des § 2 Abs. 2 UrhG vorliegen.¹² Auch reicht es aus, dass das Werk mit technischen Hilfsmitteln wahrnehmbar, etwa auf einem Datenträger, gespeichert ist.¹³ Werden Text- und Datenquellen zur darauffolgenden Auswertung in digitale Form überführt oder existieren bereits als genuin digitale Quelle, so genügen diese Ausdrucksformen den Anforderungen des § 2 Abs. 2 UrhG.

c) Geistiger Gehalt

Aus dem Werk muss außerdem ein geistiger Gehalt, eine *menschliche Gedanken- oder Gefühlsäußerung* hervorgehen.¹⁴ Bei Sprachwerken nach § 2 Abs. 1 Nr. 1 UrhG, die bei literarischen Texten in der Regel vorliegen, vermittelt Sprache diese Gedanken- oder Gefühlsäußerung.¹⁵ Der geistige Gehalt von Romanen findet seinen Niederschlag beispielsweise in der Gedankenführung ihrer Inhalte und ihrer geistvollen Form.¹⁶ Generell stellt der geistige Gehalt bei Sprachwerken eine niederschwellige Hürde dar.¹⁷ Entsteht ein Werk jedoch rein aus mechanischer Tätigkeit ohne Ausdruck der Gedanken oder Gefühle des Werk-schaffenden, so ist dieses Zufallswerk mangels geistigen Gehalts nicht schutzfähig.¹⁸ Dies führt dazu, dass von Robotern erzeugte Gegenstände keinen geistigen

¹⁰ *Loewenheim/Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 47 unter Verweis auf EuGH C-310/17, ECLI:EU:C:2018:899, GRUR 2019, 73 Rn. 40 f. – Levola Hengelo/Smile Foods und BGH GRUR 1985, 1041, 1046 – Inkasso-Programm.

¹¹ *A. Nordemann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 2 Rn. 23.

¹² OLG Köln GRUR-RR 2010, 143, 144 – Wie ein Tier in einem Zoo.

¹³ Für das Fernsehbild: BGH GRUR 1962, 470, 472 – AKI.

¹⁴ *A. Nordemann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 2 Rn. 25.

¹⁵ *Bullinger*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 2 Rn. 58.

¹⁶ Allgemein zu den Voraussetzungen des geistigen Gehalts bei Sprachwerken: *Loewenheim/Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 2 Rn. 45 mwN.

¹⁷ *Jotzo*, Teil 2, B. in diesem Band, 61, 65 unter Verweis auf *Bullinger*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 2 Rn. 64.

¹⁸ *A. Nordemann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 2 Rn. 25.

Gehalt haben, wenn ihre Erstellung durch Menschen lediglich ausgelöst wird und die Programmierung keine Gestaltungselemente vorgibt.¹⁹

d) Individualität

Zentrales Kriterium des Werkbegriffs ist die Individualität.²⁰ Sie setzt voraus, dass das Werk vom individuellen Geist des Urhebers geprägt sein muss und das Ergebnis eines *individuellen Schaffens* darstellt.²¹ Die rein mechanische Aneinanderreihung und das Zusammenfügen von Informationsmaterial würden beispielsweise keine Schutzfähigkeit eines Registers auslösen.²² Auch der rein routinemäßigen, alltäglichen Leistung, wie sie etwa in einem Anwaltsschriftsatz Ausdruck findet, wird die ausreichende Individualität abgesprochen.²³ Außerdem fehlt es an einer eigenschöpferischen Leistung, wenn sich die Gestaltung allein aus der Natur der Sache ergibt oder durch Zweckmäßigkeitserwägungen vorgegeben ist.²⁴ So handelt es sich beispielsweise bei rein beschreibenden Ergänzungen in einer Bibliographie, wie die Aufzählung von Handlungsort, Handlungszeit sowie der Akteure, typischerweise um objektive Elemente, deren Darstellung reinen Zweckmäßigkeitserwägungen folgt.

Der individuelle Charakter des Werkes kann sich in Inhalt, Form oder Konzeption der Schöpfung ausdrücken.²⁵ Auch die *Auswahl oder Anordnung* des Stoffes kann eine individuelle geistige Schöpfung begründen.²⁶ So kann etwa das Re-

¹⁹ Zur Abgrenzung und zum möglichen Schutz zugunsten des Softwareschöpfers siehe *A. Nordemann*, in *Fromm/Nordemann, Urheberrecht*, 12. Aufl. 2018, § 2 Rn. 25 und *Czychowski*, in *Fromm/Nordemann, Urheberrecht*, 12. Aufl. 2018, § 69a Rn. 18 ff.

²⁰ Ausführlich zur Voraussetzung der Individualität im Kontext der IT-gestützten Forschung mit großen Textbeständen siehe *Jotzo*, Teil 2, B. in diesem Band, 61, 65 f, zum Begriff der Individualität in der Terminologie des BGH ausführlich siehe *Loewenheim/Leistner*, in *Schricker/Loewenheim, Urheberrecht*, 6. Aufl. 2020, § 2 Rn. 50.

²¹ *Loewenheim/Leistner*, in *Schricker/Loewenheim, Urheberrecht*, 6. Aufl. 2020, § 2 Rn. 50 f.

²² BGH GRUR 1987, 704, 706 Warenzeichenlexika.

²³ BGH GRUR 1986, 739, 741 – Anwaltsschriftsatz.

²⁴ So für die Beschreibungen von Produkten in einem Versandkatalog nach Merkmalen wie Schnittform, Eleganz, Verarbeitung usw. LG Stuttgart ZUM-RD 2011, 649, 652.

²⁵ *A. Nordemann*, in *Fromm/Nordemann, Urheberrecht*, 12. Aufl. 2018, § 2 Rn. 24 mwN.

²⁶ *Loewenheim/Leistner*, in *Schricker/Loewenheim, Urheberrecht*, 6. Aufl. 2020, § 2 Rn. 50 f. unter Verweis auf BGH GRUR 2011, 134 Rn. 36 – Perlentaucher; zur individuellen Auswahl und Anordnung als Voraussetzung des Schutzes von Sammelwerken siehe unten II. 1.

gister einer Sammlung mittelalterlicher Texte schutzfähig sein, weil die Konzeption von Auswahl und Anordnung seiner Einträge eine wissenschaftliche Bearbeitung der enthaltenen Texte ermöglicht.²⁷

Entscheidend für die Beurteilung der Individualität ist der *Gesamteindruck* der konkreten Gestaltung im Gesamtvergleich zu vorbestehenden Gestaltungen.²⁸ So kann etwa der Text einer Bedienungsanleitung eigenschöpferisch gestaltet sein, wenn man den Gesamteindruck von Auswahl und Anordnung der ihn ergänzenden Zeichnungen und Bilder sowie die wechselseitige Aufgabenzuweisung der Bild- und Textinformationen in den Blick nimmt.²⁹

Das Maß der individuellen Prägung des Werkes kann variieren, dabei wird von *Gestaltungshöhe*³⁰ gesprochen.³¹ Ein Mindestmaß an Individualität, wie bei der sogenannten „kleinen Münze“ des Urheberrechts, ist jedoch erforderlich.³² Die „kleine Münze“ beschreibt dabei geistige Schöpfungen, die an der untersten Schutzuntergrenze liegen und gerade noch als urheberschutzfähig einzuordnen sind.³³ Die Voraussetzung der Individualität und deren Ausmaß kann einerseits je nach Werkart³⁴ einzelfallbezogen unterschiedlich ausgelegt werden³⁵ und spielt andererseits eine wesentliche Rolle mit Blick auf den Schutzzumfang bei Verletzungshandlungen.³⁶

²⁷ BGH GRUR 1980, 227, 231 – Monumenta Germaniae Historica.

²⁸ *Loewenheim/Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 55.

²⁹ BGH GRUR 1993, 34, 36 – Bedienungsanweisung.

³⁰ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 2 Rn. 20 beschreibt mit Blick auf die Gestaltungshöhe unterschiedliche Terminologien: beispielsweise wird auch von einem hinreichenden schöpferischen Eigentümlichkeitsgrad (BGH GRUR 1988, 533, 535 – Vorentwurf II) oder von einer Schöpfungshöhe gesprochen.

³¹ *Loewenheim/Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 51.

³² *Loewenheim/Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 51.

³³ Siehe beispielsweise BGH GRUR 1995, 581, 582 – Silberdistel; zum Begriff der „kleinen Münze“ und dessen Herkunft siehe die ausführliche Erläuterung von *Schulze*, Die kleine Münze und ihre Abgrenzungsproblematik bei den Werkarten des Urheberrechts, 1 ff. und *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 293 ff.

³⁴ Zur Relevanz der Unterscheidung zwischen Textarten siehe unten 2.

³⁵ Beschreibung der Schutzuntergrenze und relevanter Rechtsprechung siehe *Loewenheim/Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 2 Rn. 59.

³⁶ *Loewenheim/Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 52.

Je stärker die individuelle Prägung des Werkes ausfällt, desto größer ist auch der *Schutzumfang*.³⁷ Umgekehrt folgt einem geringen Maß an Eigentümlichkeit, beispielsweise bei Zeichnungen in einer Bedienungsanleitung, ein entsprechend enger *Schutzumfang*.³⁸

Erforderlich ist das Vorhandensein eines wenn auch geringen *Gestaltungsspielraums*³⁹, der dem Werkschaffenden zur Entfaltung seiner individuellen Schöpfungstätigkeit verbleibt.⁴⁰ So führt für den Fall von Satzteilen aus zusammengefassten Artikeln der Tagespresse erst die Auswahl, Anordnung und Kombination der Wörter zu einer geistigen Schöpfung mit individueller Prägung.⁴¹ Für Sprachwerke erhöht die Länge eines Textes an sich typischerweise den vorhandenen Gestaltungsspielraum für Individualität.⁴² Je länger ein Text also ist, desto eher kann Spielraum für Individualität vorliegen. Es kann aber auch einzelnen Satzteilen⁴³ schutzfähige Individualität zugesprochen werden.⁴⁴

2. Unterscheidung zwischen Textarten

Die Anforderungen an die Schutzfähigkeit können in der Rechtsprechung *je nach Werkart* unterschiedlich ausfallen.⁴⁵ Die in den Geisteswissenschaften analysierten Texte werden vornehmlich Sprachwerke nach § 2 Abs. 1 Nr. 1 UrhG darstellen, bei denen unterschiedliche Maßstäbe angelegt werden.⁴⁶

Wesentlich ist dabei die Unterscheidung zwischen *literarischen Texten*, *wissenschaftlichen* und *reinen Gebrauchstexten*, an die die Rechtsprechung höhere Anforderungen stellt.⁴⁷ *Literarische Texte* wie Romane erfahren in der Regel den

³⁷ Loewenheim/Leistner, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 93.

³⁸ BGH GRUR 1993, 34, 35 – Bedienungsanweisung.

³⁹ Schack, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 192 verweist hier auf die Rechtsprechung des BGH GRUR 1980, 227, 230 – Monumenta Germaniae Historica, in der der BGH eine schöpferische Leistung in der Formgebung, im bloßen Sammeln, Einteilen und Anordnen des Stoffes als ausreichenden Gestaltungsspielraum gelten lässt.

⁴⁰ Schack, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 192.

⁴¹ EuGH C-5/08, ECLI:EU:C:2009:456 Rn. 45 – Infopaq/DDF.

⁴² OLG Köln GRUR-RR 2016, 59 Rn. 26 unter Verweis auf OLG Köln ZUM-RD 2012, 35 (für Werbetexte und darin enthaltene Produktbeschreibungen).

⁴³ EuGH C-5/08, ECLI:EU:C:2009:456 Rn. 46 – Infopaq/DDF.

⁴⁴ Ausführlich zur Länge von Texten und deren individuellen Prägung siehe Jotzo, Teil 2, B. in diesem Band, 61, 66 f.

⁴⁵ Schulze, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 2 Rn. 24.

⁴⁶ Jotzo, Teil 2, B. in diesem Band, 61, 64.

⁴⁷ Schulze, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 2 Rn. 27 unter Verweis auf BGH GRUR 1993, 34, 36 Bedienungsanweisung.

Schutz als Sprachwerk.⁴⁸ Die Anforderungen an die Schutzfähigkeit literarischer Texte werden in der Rechtsprechung tendenziell niedriger angesetzt als bei wissenschaftlichen Werken oder Gebrauchstexten.⁴⁹ Neben der Form der Darstellung eines Romans sind auch die Inhalte, die Phantasiegeschichte oder die Kunstfiguren eines Textes, bei unverwechselbarer Charakterisierung,⁵⁰ schutzfähig.⁵¹

Bei *wissenschaftlichen Texten* und Registern⁵² können die Darstellungsform, die individuelle Gedankenführung sowie Auswahl und Anordnung des Materials geschützt sein.⁵³ Jedoch sind die Grenzen der Schutzfähigkeit bei wissenschaftlichen Texten enger, da wissenschaftliche Erkenntnisse und Ideen grundsätzlich frei bleiben sollen.⁵⁴ Außerdem weisen eine im wissenschaftlichen Fachbereich übliche Sprache, eine aus wissenschaftlichen Gründen gebotene Darstellungsform sowie der übliche Textaufbau keine hinreichende Individualität auf.⁵⁵ Gleichzeitig kann beispielsweise auch ein wissenschaftliches Register, das neben der reinen Auffindbarkeit von Schlagworten auch wissenschaftliche Erläuterungen enthält, trotz geringer Individualität urheberrechtsschutzfähig sein.⁵⁶ An *Gebrauchstexte* wie Gebrauchsanleitungen, Merkblätter, Lexika oder Werbetexte stellt die Rechtsprechung höhere Anforderungen an die Schutzfähigkeit, also ein deutliches Übertreten des Durchschnitts.⁵⁷ Sind Aufbau, Ausdrucksweise und Darstellungsart einer Bedienungsanweisung oder eines Warenzeichenlexikon durch ihre Zweckbestimmung weitestgehend vorgegeben und üblich, so wird die individuelle Eigenart diesem Gebrauchstext abgesprochen.⁵⁸

⁴⁸ Siehe oben I. c.

⁴⁹ Bullinger, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 2 Rn. 58.

⁵⁰ BGH GRUR 1999, 984, 987 – Laras Tochter, BGH GRUR 2014, 258 Rn. 44 – Pippi Langstrumpf Kostüm.

⁵¹ Bullinger, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 2 Rn. 48 und 58.

⁵² BGH GRUR 1980, 227, 229 – Monumenta Germaniae Historica.

⁵³ Loewenheim/Leistner, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 83 ff., Bullinger, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 2 Rn. 50, 56.

⁵⁴ Schulze, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 2 Rn. 26.

⁵⁵ Schulze, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 2 Rn. 93 unter Verweis auf BGH GRUR 1981, 352, 355 – Staatsexamensarbeit.

⁵⁶ KG GRUR 1991, 596, 598 – Schopenhauer-Ausgabe.

⁵⁷ BGH GRUR 1993, 34, 36 – Bedienungsanweisung, BGH GRUR 1987, 704, 706 – Warenzeichenlexika, OLG Frankfurt WRP 2015, 1004, 1005 – Bedienungsanleitung, LG Köln ZUM-RD 2012, 45 – Werbetext, siehe dazu A. Nordemann, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 2 Rn. 61 mwN.

⁵⁸ BGH GRUR 1993, 34, 36 – Bedienungsanweisung, BGH GRUR 1987, 704, 706 – Warenzeichenlexika.

Deutliches Übertreten des Durchschnitts und damit Schutzfähigkeit einer Bedienungsanweisung kann hingegen beispielweise zugesprochen werden, wenn Auswahl und Anordnung enthaltener Zeichnungen und ihr Zusammenspiel mit dem Text einen eigenschöpferischen Charakter besitzen.⁵⁹

Neben den unterschiedlich hohen Anforderungen an die individuelle Prägung stehen *Auswahl und Anordnung* des Materials als weitere Anknüpfungspunkte für den urheberrechtlichen Schutz im Fokus. So kann bei Sammelwerken nach § 4 UrhG wie beispielweise Anthologien, Lexika oder Handbüchern eine eigenschöpferische Leistung in der Auswahl und Anordnung der einzelnen Elemente zu einem individuell gestalteten Ganzen liegen.⁶⁰

Abzugrenzen ist das *Leistungsschutzrecht* des Datenbankherstellers nach § 87a UrhG, für das – im Gegensatz zur Frage nach der Individualität in § 2 Abs. 2 UrhG – der Aufwand sowie die Mühe und Kosten der Herstellung einer Datenbank im Vordergrund stehen.⁶¹ Für das Leistungsschutzrecht der wissenschaftlichen Ausgabe nach § 70 UrhG steht dagegen die wissenschaftlich sichtende Tätigkeit bei der Erstellung einer solchen Ausgabe im Fokus.⁶²

II. Urheberrechtlicher Schutz von Sammelwerken

Sammlungen von Werken, Daten oder sonstigen Elementen können unabhängig vom schöpferischen Wert dieser Elemente als urheberrechtsschutzfähige Sammelwerke nach § 4 Abs. 1 UrhG geschützt sein.⁶³ Regelmäßig als Sammelwerke nach § 4 Abs. 1 UrhG sind beispielsweise Anthologien, Lexika, Handbücher oder Bildbände geschützt.⁶⁴ Auch Sammlungen bibliographischer Daten können bei Vorliegen der Voraussetzungen unter dem Schutz eines Sammelwerkes nach § 4 Abs. 1 UrhG stehen. Darüber hinaus können auch Zeitschriften

⁵⁹ BGH GRUR 1993, 34, 36 – Bedienungsanweisung.

⁶⁰ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 289 mwN.

⁶¹ *Jotzo*, Teil 2, B. in diesem Band, 61, 66.

⁶² Dazu *Erler-Fridgen*, Teil 2, A. in diesem Band.

⁶³ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 6. Aufl. 2018, § 4 Rn. 11 f., *Bullinger*, in *Wandtke/Bullinger*, Urheberrecht, 5. Aufl. 2019, § 2 Rn. 50.

⁶⁴ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 289 unter Verweis auf BGH GRUR 2013, 1213, Rn. 57 f. – SUMO.

oder Tageszeitungen den Schutz als Sammelwerk erfahren.⁶⁵ In der Regel ist jedoch die Zusammenstellung der Elemente und die thematische Zuordnung zu Sachgebieten in einer Zeitschrift keine schöpferische Tätigkeit.⁶⁶

1. Schutzvoraussetzungen

Schutzvoraussetzung solcher Sammelwerke unter § 4 Abs. 1 UrhG ist, dass die *Auswahl oder Anordnung* der einzelnen Elemente eine persönliche geistige Schöpfung ist.⁶⁷ Dafür muss der geistige Gehalt des Werkes nach dessen Gesamteindruck über die bloße Summe der Inhalte der Einzelwerke, Daten und sonstigen Elemente hinausgehen.⁶⁸ Auswahl meint hier das Sammeln und Aufnehmen, Anordnung beschreibt die Einteilung und Präsentation von Elementen.⁶⁹ Die schöpferische Leistung muss nur hinsichtlich einer von beiden Gestaltungen, Auswahl oder Anordnung vorliegen.⁷⁰ Beispielsweise kann sich die schöpferische Leistung daraus ergeben, dass die Informationsauswahl für ein Warenzeichenlexikon besonders konzipiert ist.⁷¹ Sie fehlt jedoch etwa bei einer sich zwangsläufig ergebenden Liste der Popmusik-Charts.⁷² Die schöpferische Leistung kann aber auch sowohl in der Auswahl als auch in der Anordnung der Elemente liegen.⁷³

Findet also keine individuelle Auswahlentscheidung hinsichtlich der aufgenommenen Werke oder Daten statt, so kann die Schutzfähigkeit eines Sammelwerkes nicht allein aus diesem Grund verneint werden.⁷⁴ Daher schließt die Vollständigkeit der enthaltenen Beiträge – trotz Fehlen einer Auswahlentscheidung –

⁶⁵ LG Bielefeld GRUR-RR 2010, 324, 325 – Math. Fachzeitschrift; OLG Hamm ZUM 2008, 598, 601.

⁶⁶ OLG München MMR 2007, 525, 526.

⁶⁷ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 4 Rn. 11.

⁶⁸ BGH GRUR 1990, 669, 673 f. – Bibelreproduktion; Marquardt, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 4 Rn. 5; zu den Voraussetzungen der Schutzfähigkeit nach § 2 Abs. 2 UrhG siehe oben I.

⁶⁹ Leistner, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 4 Rn. 23.

⁷⁰ Leistner, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 4 Rn. 23.

⁷¹ BGH GRUR 1987, 704, 705 – Warenzeichenlexika.

⁷² BGH GRUR 2005, 857, 858 – HIT Bilanz.

⁷³ BGH GRUR 2007, 685 Rn. 18 f. – Gedichttitelliste.

⁷⁴ Für die Schutzfähigkeit einer Datenbank: BGH GRUR 2011, 79 Rn. 38 – Markenheftchen; allgemein zur Alternativität der Schutzvoraussetzungen Auswahl und Anordnung: Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 4 Rn. 11.

zum Beispiel bei einer auf Vollständigkeit angelegten Bibliographie französischer Romane den Schutz des Sammelwerkes nicht zwingenderweise aus.⁷⁵ Wenn nämlich eine Sammlung *Vollständigkeit* anstrebt, bleibt trotzdem Raum für schöpferische Gestaltung, etwa bezüglich der Auswahl der zu sammelnden Gesamtmenge.⁷⁶

Eine rein handwerkliche, schematische oder routinemäßige Auswahl oder Anordnung reicht nicht aus.⁷⁷ Insbesondere liegt keine individuelle Schöpfung vor, wenn sich die Anordnung aus der Natur der Sache ergibt, etwa rein chronologisch, alphabetisch, numerisch oder nach Fachgebieten vorgenommen wird.⁷⁸ Anderes gilt jedoch, wenn die Anordnung ausreichenden *Spielraum für Individualität* lässt: Werden die Kriterien zur Anordnungsentscheidung beispielsweise individuell ausgewählt und gewichtet, so kann dies dem Maßstab der „kleinen Münze des Urheberrechts“ genügen und die Schutzfähigkeit als Sammelwerk auslösen.⁷⁹ Zweifelhaft ist aber, ob auch die Auswahl nach statistischen Kriterien, etwa nach der Anzahl von Nennungen in einer Sammlung, ausreichende Individualität begründen kann.⁸⁰

Bibliographische Daten als solche sind frei und ihre bloße Sammlung unterliegt zunächst keinem urheberrechtlichen Schutz.⁸¹ Anderes gilt nach § 4 UrhG dann, wenn Anordnung und Darbietung der Daten eine eigene schöpferische Leistung darstellen.⁸² Wird zum Beispiel ein Register über die reine Zusammenstellung einzelner Fakten hinaus wissenschaftlich konzipiert, so dass Einträge zur wissenschaftlichen Bearbeitung erleichtert auffindbar sind, kann in dieser Anordnung eine schöpferische Leistung liegen.⁸³

⁷⁵ Für eine Datensammlung als Datenbankwerk: BGH GRUR 2011, 79 Rn. 38 – Markenheftchen.

⁷⁶ *Czychowski*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 4 Rn. 13.

⁷⁷ *Czychowski*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 4 Rn. 13; *Marquardt*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 4 Rn. 5.

⁷⁸ *Czychowski*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 4 Rn. 13 unter Verweis auf OLG Nürnberg GRUR 2002, 607 – Stufenaufklärung nach Weißnauer; OLG Hamburg GRUR 2000, 319, 320 – Börsendaten.

⁷⁹ OLG Frankfurt MMR 2003, 45, 46 – IMS-Health.

⁸⁰ LG Mannheim ZUM-RD 2004, 547, 549 – Gedichtsammlung.

⁸¹ OLG Hamburg ZUM 1997, 145 – Hubert Fichte Biographie.

⁸² BGH GRUR 1980, 227, 231 – Monumenta Germaniae Historica.

⁸³ BGH GRUR 1980, 227, 231 – Monumenta Germaniae Historica.

2. Gegenstand

Gegenstand des Urheberrechts am Sammelwerk ist die Sammlung als solche, also die in *Auswahl oder Anordnung* der Elemente vorgenommene Leistung.⁸⁴ Der Schutz eines Sammelwerkes wird daher nur dann beeinträchtigt, wenn die Auswahl oder Anordnung ganz oder in wesentlichen Teilen übernommen wird.⁸⁵ Werden die Gedichte aus der Auswahl einer Gedichttitelliste fast vollständig übernommen, so ist daher beispielsweise von einer Verletzung der individuellen Auswahlkonzeption auszugehen.⁸⁶ Umgekehrt reichte die Übernahme von acht finanzgerichtlichen Entscheidungen aus einer Sammlung von 63 Entscheidungen nicht für einen Eingriff in den Schutz des Sammelwerkes aus.⁸⁷ Der übernommene Teil muss also die geschützte Auswahl oder Anordnung des Sammelwerkes widerspiegeln.⁸⁸ Die Übernahme einzelner Elemente reicht aus diesem Grund für eine Verletzung des Schutzes nicht aus. Es ist hierfür die Übernahme des Schutzgegenstandes – Auswahl und Anordnung der Elemente – nötig.⁸⁹

III. Ergebnis

Texte, die eine persönliche geistige Schöpfung in wahrnehmbarer Form darstellen und ausreichende Individualität aufweisen, erfahren nach § 2 Abs. 2 UrhG urheberrechtlichen Schutz. Zentrales Kriterium ist die Individualität, die voraussetzt, dass das Werk vom individuellen Geist des Urhebers geprägt sein muss, also das Ergebnis eines individuellen Schaffens ist. Literarische Texte sind von wissenschaftlichen Werken und reinen Gebrauchstexten zu unterscheiden, an deren Schutzfähigkeit die Rechtsprechung höhere Anforderungen stellt. Sammlungen von Werken oder Daten können nach § 4 Abs. 1 UrhG als Sammelwerke geschützt sein, wenn in der Auswahl oder Anordnung ihrer Elemente eine persönliche geistige Schöpfung liegt.

⁸⁴ *Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 4 Rn. 34.

⁸⁵ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 4 Rn. 15.

⁸⁶ BGH GRUR 2007, 685 Rn. 25 f. – Gedichttitelliste II.

⁸⁷ BGH GRUR 1992, 382, 384 – Leitsätze.

⁸⁸ BGH GRUR 2013, 1213 Rn. 57 – SUMO, *Czychowski*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 4 Rn. 40 mwN.

⁸⁹ *Czychowski*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 4 Rn. 40.

Leseempfehlungen zur vertiefenden Lektüre: *Jotzo*, Der Schutz großer Textbestände nach dem UrhG, Teil 2, B. in diesem Band, *Loewenheim/Leistner*, in *Schricker/Loewenheim*, Urheberrecht, 6. Aufl. 2020, § 2 und *Czychowski*, in *Fromm/Nordemann*, Urheberrecht, 12. Aufl. 2018, § 4.

B. Der Schutz großer Textbestände nach dem UrhG

Die Nutzbarmachung fremder Textbestände für die Forschung^{*}, ^{**}

Florian Jotzo

CC BY-SA 4.0.

I. Einleitung

Text und Data Mining Anwendungen (TDM) eröffnen Forschenden faszinierende Möglichkeiten.¹ Die Digital Humanities schaffen spannende Schnittstellen zwischen den Geisteswissenschaften und der Informatik. Sie tragen Textbestände zusammen, werten diese z.B. unter sprachwissenschaftlichen Aspekten oder für die Geschichtsforschung aus. Andere nutzen große Textmengen und trainieren mit ihnen künstliche Intelligenzen. Neben den klassischen Beständen in Bibliotheken bieten digitale Datenbanken und das offene Internet reiche Ressourcen für solche Projekte. Doch wollen Forschende nicht nur alte gemeinfreie Texte verwenden, dann stellen sich eine Reihe urheberrechtlicher Fragen. Der folgende Beitrag ist Teil eines interdisziplinären Projekts, das hier rechtssichere Antworten liefern soll. Ziel ist es, Strategien zu entwickeln, mit denen Forschende im Rahmen von TDM-Prozessen große Textbestände für ihre Zwecke erschließen können, ohne Urheber- und verwandte Schutzrechte Dritter zu verletzen, und die dennoch eine große Bandbreite von Forschungsvorhaben ermöglichen.

^{*} Der Beitrag ist im Rahmen des Workshops „Strategien für die Nutzbarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte“ am 28.11.2019 und 17.01.2020 in Trier entstanden. Der Autor dankt allen Teilnehmenden für die spannenden Diskussionen. Besonderer Dank gilt Dr. *Karina Grisse*, Prof. Dr. *Benjamin Raue* und Prof. Dr. *Christof Schöch* für den produktiven Austausch.

^{**} Dieser Beitrag wurde zuerst veröffentlicht als *Jotzo* RuZ 2020, 128-142 unter einer CC BY-SA Lizenz.

¹ Siehe nur *Datenethikkommission der Bundesregierung* Gutachten Oktober 2019, 124, abrufbar unter: <https://t1p.de/dkk-gutachten>, zuletzt abgerufen am 25.3.2020.

Der Gesetzgeber hat das enorme Potenzial TDM-basierter Forschung erkannt und die urheberrechtlichen Schranken ausgebaut. Im Zuge des UrhWissG² wurde eine eigene Schranke für Text- und Data Mining zu Forschungszwecken in § 60d Urheberrechtsgesetz (UrhG) geschaffen. Die neue UrhR-RL 2019/790/EU³ sieht in Art. 3 und 4 ebenfalls explizite TDM-Schranken vor.⁴ Sie erlauben insbesondere Vervielfältigungen zur Erstellung von Korpora, das Annotieren von Texten und deren automatisierte Auswertung. Damit steht das Urheberrecht dem Zusammentragen und Auswerten von Texten zu Forschungszwecken grundsätzlich nicht mehr im Weg. Die Reformen haben den Möglichkeitsraum für die Forschung spürbar erweitert.⁵ Probleme verbleiben jedoch aus Forschersicht insbesondere bei der Weitergabe von Korpora an Dritte, wie sie zB für ressourcensparende Anschlussnutzungen sinnvoll und auch außerhalb formeller wissenschaftlicher Qualitätssicherung (*peer review*) für eine nachvollziehbare transparente Forschungsarbeit förderlich wäre.⁶ Diese Defizite und der Wunsch nach Rechtssicherheit führen zu der Überlegung, dass Forschende oft gar nicht die Primärtexte in ihrer geschützten Ursprungsform brauchen. Worthäufigkeitsstatistiken, lemmatisierte Wortlisten, N-Gramme, Term-Dokument Matrices und Wort-Vektoren auf Grundlage von Word Embedding Modellen sind für viele Einsatzfelder der Digital Humanities wertvolle Ressourcen, die der Beitrag von *Christof Schöch* ausführlich vorstellt.⁷ Gemeinsam haben diese Formate, dass sie die urheberrechtlich geschützten Elemente der Primärtexte aussparen können und dann die Interessen der Rechteinhaber nicht berühren. Dazu müssten die geschützten Ausgangstexte im Rahmen der TDM-Prozesse in eine urheberrechtsfreie Form konvertiert

² Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft, BGBl 2017 Teil I, 3346.

³ Richtlinie (EU) 2019/790 vom 17.4.2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG.

⁴ Siehe dazu ausführlich *Raue* ZUM 2019, 684 ff; *Spindler* CR 2019, 277, 278 f. Zur geplanten Umsetzung in §§ 44b und 60d UrhG-E siehe *Raue*, ZUM 2020, 172 ff.

⁵ Überblick durch *Hauk/Pflüger* ZUM 2020, 383 ff.

⁶ *Raue/Schöch*, Teil 3, A. in diesem Band, 129, 131 f. Zu den Defiziten der TDM-Schranken siehe ausführlich *Grisse*, Nutzbarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte, RuZ 2020, 143, 144 f. unter einer CC BY-SA Lizenz. Mehr Raum für Anschlussforschungen und Review-Verfahren sehen dagegen.

⁷ *Schöch et al.*, Text und Data Mining mit urheberrechtlich geschützten Textbeständen: Aktuelle Überlegungen aus Perspektive der Informatik, Digital Humanities und Gedächtnisinstitutionen, RuZ 2020, 160, 161 ff. unter einer CC BY-SA Lizenz.

werden, die keinen klassischen Werkgenuss mehr ermöglicht. Solche abgeleiteten Speicherformate wären für Forschende dann eine chancenreiche Alternative, um Textkorpora langfristig aufbewahren und mit Dritten teilen zu dürfen, ohne das enge Korsett der urheberrechtlichen Schranken anlegen zu müssen. Der folgende Beitrag bildet den Ausgangspunkt dieser Strategie. Er gibt einen Überblick über die verschiedenen Schutzrechte, die Forschende bei TDM-Vorhaben typischerweise berühren. Im Detail analysiert der Beitrag, welche Bestandteile von Texten und großen Textmengen das UrhG schützt. Darauf baut *Karina Grisse* auf. Ihr Beitrag nimmt die urheberrechtlichen Schranken in den Blick und zeigt, wie mit deren Hilfe Verfahren gestaltet werden können, um Texte in solche Formate zu übertragen, die keine urheberrechtlich geschützten Bestandteile mehr enthalten.⁸ Beide Beiträge beleuchten also jeweils die unterschiedlichen Seiten derselben Medaille und ergänzen einander.

II. Urheberrechtliche Befugnisse und TDM-Prozesse

Das UrhG schützt verschiedene Gegenstände. Neben urheberrechtlich geschützten Werken (§ 2 UrhG) sind für Forschende bei TDM-Vorhaben insbesondere die Leistungsschutzrechte der Datenbankhersteller (§§ 87a-e UrhG), der Presseverleger (§§ 87f-h UrhG), für wissenschaftliche Ausgaben (§ 70 UrhG) und für nachgelassene Werke (§ 71 UrhG) relevant. Das UrhG weist den Urhebern eine Reihe von Befugnissen zu, die deren ideelle und vermögensrechtliche Interessen schützen (§ 11 UrhG). Zu den Verwertungsrechten gehört insbesondere das Vervielfältigungsrecht (§ 16 UrhG). Wer beim Text und Data Mining vorbestehende Texte sammelt, digital aufbereitet und anschließend auswertet, erzeugt – technisch bedingt – eine Reihe von Kopien und greift damit in das Vervielfältigungsrecht des Urhebers ein. Auch Werkteile können Schutz genießen, wenn der übernommene Teil die schutzfähigen Elemente des Ausgangswerkes enthält.⁹ Auch verändernde Bearbeitungen und andere Umgestaltungen greifen in das Vervielfältigungsrecht ein (vgl. § 23 UrhG). Bislang wurde dazu gefordert, dass die eigenschöpferischen Züge des entlehnten Werkes in der nachgeschaffenen Gestaltung nicht völlig verblasen.¹⁰ Der EuGH hat jedoch zuletzt

⁸ *Grisse*, Teil 3, B. in diesem Band, 141 ff.

⁹ *Loewenheim/Leistner*, in: Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 87. Zu den schutzfähigen Bestandteilen siehe unten C.

¹⁰ BGH, Urt. v. 1.12.2010 – I ZR 12/08, GRUR 2011, 134 Tz. 33 – *Perlentaucher. Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 417, 268.

diese äußerste Grenze des Vervielfältigungsrechts nicht mit Hilfe des „Verblasens“ gezogen, sondern auf die Wiedererkennbarkeit des kopierten Fragmentes abgestellt.¹¹ Da die Vorgaben aus Art. 2 RL 2001/29/EG (InfoSoc-RL) für die EU-Mitgliedstaaten zwingend sind,¹² wird der BGH seine Linie anpassen.¹³ Offen ist dann jedoch, wie das unbestimmte Kriterium der Wiedererkennbarkeit auszufüllen sein wird.¹⁴

Darüber hinaus greifen Forschende in das Recht der öffentlichen Zugänglichmachung (§ 19a UrhG) ein, wenn sie die Ausgangstexte etwa in einem Korpus ihren Kollegen oder zum Download im freien Netz bereitstellen, um die Überprüfung ihrer Forschungsergebnisse oder Anschlussnutzungen zu ermöglichen. Doch welche Elemente führen dazu, dass einzelne Texte oder Textbestände durch die genannten Ausschließlichkeitsrechte nach dem UrhG monopolisiert werden?

III. Schutz einzelner Texte

1. Sprachwerke, § 2 I Nr. 1 UrhG

Für Texte kommt insbesondere der Schutz als Sprachwerk (§ 2 I Nr. 1 UrhG) in Betracht. Der urheberrechtliche Werkschutz setzt voraus, dass die jeweilige Gestaltung die Hürde der „persönlichen geistigen Schöpfung“ iSv. § 2 II UrhG nimmt.¹⁵ Werkschaffende müssen dazu durch ihre persönliche und individuelle Tätigkeit einen geistigen Gehalt in einer wahrnehmbaren Form zum Ausdruck bringen.

¹¹ EuGH, Urt. v. 29.7.2019 – C-476/17, ECLI:EU:C:2019:624 = GRUR 2019, 929 Rn. 37 – *Pelham* (Leistungsschutzrecht des Tonträgerherstellers). Dazu *Leistner*, GRUR 2019, 1008, 1009 ff. und *Stieper*, ZUM 2019, 713, 718 f. Mit Blick auf TDM-Prozesse *Jakl*, MMR 2019, 711, 714.

¹² Zur verändernden Bearbeitung als Teil des Vervielfältigungsrechts der InfoSoc-RL *Jotzo*, ZGE 2017, 447, 467 f. und ausführlich *Einfeldt*, Open Content Lizenzen und das Bearbeitungsrecht, Göttingen 2020, 48 ff.

¹³ Gegen das Kriterium der Wiedererkennbarkeit noch BGH Urt. v. 16.4.2015 – I ZR 225/12, GRUR 2015, 1189 Tz. 109 – Goldraper.

¹⁴ *Grisse*, Teil 3, B. in diesem Band, 141, 148 mwN.

¹⁵ Vgl. EuGH, Urt. v. 13.11.2018 – C-310/17, ECLI:EU:C:2018:899 = GRUR 2019, 73 Tz. 35 ff. mwN. – *Levola Hengelo* („eigene geistige Schöpfung“), Anm. *Schack*, GRUR 2019, 75. Die EuGH-Rechtsprechung vergleicht mit dem deutschen Recht *Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 19 ff. Die Anforderungen an den Werkbegriff gelten nicht speziell für Sprachwerke, sondern allgemein für alle Werkarten, *Schack*, (Fn. 10) Rn. 180 ff.

a) Voraussetzungen

1. Der Einsatz von IT als Hilfsmittel („computer-aided works“) steht der *persönlichen* Schöpfung nicht entgegen, so lange wie eine natürliche Person den kreativen Schaffensprozess steuert.¹⁶ Wenn dagegen künstliche Intelligenzen eigenständig Wetterberichte, Börsennachrichten oder andere Texte erzeugen, erlangen diese keinen Urheberrechtsschutz.¹⁷

2. Sprachwerke vermitteln durch das Mittel der Sprache Gedanken und/oder Gefühle (*geistiger Gehalt*).¹⁸ Auf eine literarische Qualität oder das ästhetische Empfinden kommt es insoweit nicht an.¹⁹ Dass der geistige Gehalt bei Sprachwerken keine sonderliche Hürde darstellt, zeigt der (unionsrechtlich gebotene) Schutz von Computerprogrammen (§ 2 II Nr. 1 UrhG)²⁰ und Wörterbüchern.²¹

3. Diesen geistigen Gehalt muss der Urheber in eine *wahrnehmbare Form* bringen.²² Einen „Geistesblitz“ kann niemand anderes erfassen, sodass er noch kein Urheberrecht auslöst.²³ Nötig ist eine Ausdrucksform, die gewährleistet, dass andere Personen mit hinreichender „Genauigkeit und Objektivität“ das geschützte Werk identifizieren können.²⁴ Wenn Forschende im Rahmen von TDM-Prozessen mit fremden Texten arbeiten, liegen diese jedoch zumindest digital vor und sind damit hinreichend wahrnehmbar.

4. Dreh- und Angelpunkt des Werkbegriffs ist die *Individualität*. Sie verlangt, dass im Werk die Persönlichkeit des Urhebers zum Ausdruck kommt.²⁵

¹⁶ Peifer, in: FS Walter, Wien 2018, 222, 226 f.; A. Nordemann, in: Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 2 Rn. 21; Lauber-Rönsberg, GRUR 2019, 244, 247; Schack, (Fn. 10) Rn. 184.

¹⁷ Lauber-Rönsberg, GRUR 2019, 244, 247; Ory/Sorge, NJW 2019, 710, 711 f.; A. Nordemann, in: F/N (Fn. 16), UrhG, § 2 Rn. 25.

¹⁸ Loewenheim/Leistner, in: S/L (Fn. 10), UrhG, § 2 Rn. 45 f.; G. Schulze, in: Dreier/Schulze, UrhG, 6. Aufl. 2018, § 2 Rn. 12; Schack, (Fn. 10) Rn. 186.

¹⁹ Ahlberg, in: Möhring/Nicolini, Urheberrecht, 4. Aufl. 2018, § 2 Rn. 76.

²⁰ Zur Absenkung der Schutzanforderungen Ahlberg, in: M/N (Fn. 19), UrhG, § 2 Rn. 93

²¹ Bullinger, in: Wandtke/Bullinger, UrhG, 5. Aufl. 2019, § 2 Rn. 64 mwN.

²² A. Nordemann, in: F/N (Fn. 16), UrhG, § 2 Rn. 23.

²³ Prägnant Schack, (Fn. 10) Rn. 187.

²⁴ EuGH, Urt. v. 13.11.2018 – C-310/17, ECLI:EU:C:2018:899 = GRUR 2019, 73, Tz. 41 – *Levola Hengelo*. Gedanken müssen dazu nicht zwingend verkörpert oder gar verschriftlicht werden; auch improvisierte Reden, Vorträge und Sketche sind schutzfähig, G. Schulze, in: D/S (Fn. 18), UrhG, § 2 Rn. 82; Loewenheim/Leistner, in: S/L (Fn. 10), UrhG, § 2 Rn. 47, 101.

²⁵ EuGH, Urt. v. 1.12.2011 – C-145/10, ECLI:EU:C:2013:138 Tz. 99, 43 – *Painer*; GA *Wathelet*, Schlussanträge v. 25.8.2019 – C-310/17, ECLI:EU:C:2018:618, Rn. 43 – *Levola Hengelo*.

Wie stark eine Person ihr Werk geprägt hat, ist nicht nur die zentrale Schutzvoraussetzung,²⁶ sondern bestimmt zugleich den Schutzzumfang des Werkes bei Übernahmen durch Dritte.²⁷ So ist bereits die Kopie eines Textausschnitts eine Vervielfältigung (§ 16 UrhG), wenn der übernommene Ausschnitt seinerseits eigenschöpferische Bestandteile des Ausgangswerks enthält.²⁸

Individuelles Schaffen muss nicht unbedingt objektiv neu sein.²⁹ Anders als beim Leistungsschutzrecht des Datenbankherstellers (§ 87a UrhG) haben auch der Aufwand, die Mühe und die Kosten, die jemand beim Erstellen eines Textes aufgebracht hat, keinen Einfluss auf die Individualität.³⁰ Keine Rolle spielt ferner, ob ein Text sitten- oder gesetzeswidrige Inhalte enthält.³¹ Anstößige Texte können genauso urheberrechtlichen Schutz beanspruchen, wie solche, die rechtswidrig entstanden sind oder die (Persönlichkeits-) Rechte Dritter verletzen.

Urheber können ihr Werk nur dann prägen, wenn es einen Gestaltungsspielraum gibt, innerhalb dessen sie ihre Individualität entfalten können.³² Dieser Spielraum fehlt, soweit fachliche Notwendigkeiten, Methoden oder schlichte Zweckmäßigkeit die Gestaltung vorgeben.³³ Längere Texte bieten mehr Raum für Individualität als kurze.³⁴ Einzelne Worte und kurze Satzteile sind daher in aller Regel nicht individuell;³⁵ insoweit besteht ein Freihaltebedürfnis zu Gunsten der Kommunikationsfreiheiten. Deshalb können Forschende im Rahmen

²⁶ *Loewenbeim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 50; *Schack*, (Fn. 10) Rn. 189.

²⁷ *Loewenbeim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 52.

²⁸ Siehe oben Fn. 6.

²⁹ Anders als die gewerblichen Schutzrechte Patent (§ 1 I PatG) und eingetragenes Design (§ 2 I DesignG), schützt das Urheberrecht auch unabhängige Doppelschöpfungen, wenn Urheber vorbestehende Gestaltungen bei ihrer Arbeit nicht gekannt haben, *Schack* Kunst und Recht, 3. Aufl. 2017, Rn. 18; *Loewenbeim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 64 f.; *A. Nordemann*, in: F/N (Fn. 16), UrhG, § 2 Rn. 26.

³⁰ *G. Schulze*, in: D/S (Fn. 18), UrhG, § 2 Rn. 53 mwN. Vgl. auch: EuGH, Urt. v. 29.7.2019 – ECLI:EU:C:2019:623 = GRUR 2019, 934 Tz. 23 – Funke Medien NRW GmbH.

³¹ BGH, Urt. v. 23.2.1995 – I ZR 68/93, Z 129, 66 = GRUR 1995, 673, 675 – *Mauer Bilder*; *Schack*, KuR (Fn. 29) Rn. 230.

³² BGH, Urt. v. 13.11.2013 – I ZR 143/12, Z 199, 52 = GRUR 2014, 175 Tz. 41 – *Ge-burtstagszug*; EuGH, Urt. v. 16.7.2009 – C-5/08, ECLI:EU:C:2009:465 = GRUR 2009, 1041 Tz. 45 – *Infopaq I*; *Schack*, (Fn. 10) Rn. 192; *Lauber-Rönsberg*, GRUR 2019, 244, 246.

³³ *Loewenbeim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 102; *A. Nordemann*, in: F/N (Fn. 16), UrhG, § 2 Rn. 40.

³⁴ OLG Köln, Urt. v. 12.6.2015 – 6 U 5/15 = GRUR-RR 2016, 59 Tz. 28 mwN. – *Afghanistan Papiere*; OLG Hamburg, Urt. v. 26.4. 2010 – 5 U 160/08, ZUM-RD 2010, 467 Tz. 3 – *Solange du wild bist* (Schutzfähigkeit einer Liedtextzeile); *Schack*, (Fn. 10) Rn. 202.

³⁵ *Schack*, (Fn. 10) Rn. 194 ff.

von TDM-Prozessen einzelne Worte oder kurze N-Gramme nutzen, ohne vorbestehende Urheberrechte an den Ausgangstexten zu berühren. Auf den ersten Blick ist es daher verlockend, die Wahrscheinlichkeit des Urheberschutzes an Hand der Textlänge zu prognostizieren, da diese vergleichsweise einfach softwaregestützt ermittelt werden kann. Dieses Vorgehen hat jedoch Grenzen. Mit dem Umfang wächst die Wahrscheinlichkeit, dass ein Text oder ein Ausschnitt daraus das nötige Maß an schöpferischer Leistung aufweisen. So hat der EuGH etwa schon beim Extrahieren von elf aufeinanderfolgenden Worten aus Zeitungsartikeln den Werkschutz und damit einen Eingriff in das Vervielfältigungsrecht für möglich gehalten.³⁶ Auch jenseits dieser Grenze liefert das schlichte Zählen von Wörtern nur Orientierung, aber keine verlässlichen Aussagen.³⁷ Selbst ein Anagramm – wie „Adolf Hitler – Folterhilda“³⁸ –, einzelne Sätze³⁹ oder kurze Gedichte⁴⁰ können in Ausnahmefällen Individualität aufweisen, während umfangreiche aber deskriptive technische Ausschreibungsunterlagen⁴¹ oder routinemäßige Anwaltsschriftsätze⁴² oft als nicht hinreichend individuell angesehen werden.

Die Individualität erfordert keine statistische, sondern eine normative Prüfung. Anknüpfungspunkte sind der schöpferische *Inhalt* von Texten und die schöpferische *Formgebung* der Inhalte.⁴³ Die schöpferische Leistung kann sich in „der von der Gedankenführung geprägten Gestaltung der Sprache“, aber auch in der „Sammlung, Auswahl, Einteilung und Anordnung des Stoffs“ ausdrücken.⁴⁴ Keine eigene Leistung ist es, vorbestehende Inhalte zu übernehmen.⁴⁵ Auch wer

³⁶ EuGH, Urt. v. 16.7.2009 – C-5/08, ECLI:EU:C:2009:465 = GRUR 2009, 1041 Tz. 48 – *Infopaq I*.

³⁷ Vgl. *Loewenheim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 68 mit weiteren Beispielen und *Schack*, (Fn. 10) Rn. 202.

³⁸ KG, Urt. v. 22.1.1971 – 5 U 2412/70, GRUR 1971, 368, 370 – *Buchstabenschütteln*.

³⁹ EuGH, Urt. v. 16.7.2009 – C-5/08, ECLI:EU:C:2009:465 = GRUR 2009, 1041 Tz. 46 ff. – *Infopaq I*.

⁴⁰ OLG München, Urt. v. 17.9.2009 – 29 U 3271/09, ZUM 2009, 970 – *Typisch München!*.

⁴¹ BGH, Urt. v. 29.3.1984 – I ZR 32/82, GRUR 1984, 659, 661 – *Ausschreibungsunterlagen*.

⁴² BGH, Urt. v. 17.4.1986 – I ZR 213/83, GRUR 1986, 739, 741 – *Anwaltsschriftsatz; Schack*, (Fn. 10) Rn. 205; großzügiger *A. Nordemann*, in: F/N (Fn. 16), UrhG, § 2 Rn. 71.

⁴³ OLG Hamburg, Urt. v. 14.4.2016 – 5 U 117/12, ZUM-RD 2016, 576, 588 – *Tagebuch der Anne Frank*. Überblick bei *Loewenheim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 76 ff.

⁴⁴ BGH, Urt. v. 1.12.2010 – I ZR 12/08, GRUR 2011, 134 Tz.36 mwN. – *Perlentaucher*.

⁴⁵ *Loewenheim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 64.

Tatsachen, Informationen oder Lehren bloß wiedergibt, leistet keinen schöpferischen Beitrag,⁴⁶ da solche Inhalte urheberrechtlich frei bleiben müssen. Ebenso wie einzelne Wörter⁴⁷ und die Sprache als solche⁴⁸ sind sie die Grundbausteine der gesellschaftlichen, wissenschaftlichen und kulturellen Kommunikation, die allen zur freien Verfügung stehen müssen.⁴⁹

Dieses Freihaltebedürfnis berücksichtigt die Rechtsprechung auch bei dem für den Urberschutz notwendigen Mindestmaß an Individualität, der so genannten „Gestaltungshöhe“.⁵⁰ Für literarische Texte – wie Gedichte, Romane und Lieder – akzeptieren die Gerichte den Schutz von Leistungen, die nur einen „geringen Grad individuellen Schaffens und damit eine geringe Gestaltungshöhe“ aufweisen („kleine Münze“).⁵¹ Bei „Gebrauchstexten“ ist die Rechtsprechung dagegen strenger und verlangt „ein deutliches Übertagen“ des Alltäglichen, des Handwerksmäßigen oder der mechanisch-technischen Aneinanderreihung von Material.⁵² Ob jemand den *Inhalt* und / oder die *Form* eines Textes hinreichend individuell gestaltet hat, entzieht sich jeder schematischen Bewertung. Nötig ist vielmehr eine normative Gesamtbetrachtung, in die alle Gestaltungselemente einfließen, die im Einzelfall genutzt worden sind.⁵³ Diese Kriterien sollen im Folgenden an Hand von Beispielen näher vorgestellt werden.

⁴⁶ EuGH, Urt. v. 29.7.2019 – C-469/17, ECLI:EU:C:2019:623 = GRUR 2019, 934 Tz. 24 – *Funke Medien NRW GmbH*; GA Szpunar, Schlussanträge v. 25.10.2018 – C-469/17, ECLI:EU:C:2018:870 Tz. 15 f. – *Funke Medien NRW GmbH*; A. Nordemann, in: F/N (Fn. 16), UrhG, § 2 Rn. 43 ff.

⁴⁷ EuGH, Urt. v. 16.7.2009 – C-5/08, ECLI:EU:C:2009:465 = GRUR 2009, 1041 Tz. 46 – *Infopaq I*.

⁴⁸ G. Schulze, in: D/S (Fn. 18), UrhG, § 2 Rn. 14. Für die Programmiersprache von Software siehe ErwGr. 11 RL 2009/24/EG.

⁴⁹ Schack, (Fn. 10) Rn. 194 ff.

⁵⁰ Vgl. BGH, Urt. v. 21.11.1980 – I ZR 106/78, GRUR 1981, 352, 353 – *Staatsexamensarbeit*.

⁵¹ BGH, Urt. v. 15.9.1999 – I ZR 57/97, GRUR 2000, 144, 145 – *Comic-Übersetzungen II*.

⁵² BGH, Urt. v. 10.10.1991 – I ZR 147/89, GRUR 1993, 34, 36 – *Bedienungsanweisung*; A. Nordemann, in: F/N (Fn. 16), UrhG, § 2 Rn. 61 mwN.

⁵³ G. Schulze, in: D/S (Fn. 18), UrhG, § 2 Rn. 57 ff.; Loewenheim/Leistner, in: S/L (Fn. 10), UrhG, § 2 Rn. 54 f.; Ahlberg, in: M/N (Fn. 19), UrhG, § 2 Rn. 75.

b) Beispiele anhand verschiedener Textgattungen

aa) Literarische Texte

Umfangreiche Romane, wie die von *Frank Schätzing*, bieten viel Raum für Kreativität und sind deshalb urheberrechtlich geschützt. Aber auch in weniger kreativen literarischen Texten (Groschenromanen) lassen sich in aller Regel eigenschöpferische Inhalte und Formulierungen nachweisen. Übernehmen Forschende Auszüge aus literarischen Texten, die mehrere zusammenhängende Sätze umfassen, greifen sie daher mit hoher Wahrscheinlichkeit ins Vervielfältigungsrecht der Urheber ein. Das ohnehin hohe Risiko wächst, wenn solche Kopien automatisch ohne Prüfung der einzelnen Texte erfolgen, so wie bei TDM-Prozessen.

Zu den geschützten Inhalten von Romanen gehören grundsätzlich die erdachten Erzählstränge, Charakteristika der Figuren und deren Beziehungsgefüge.⁵⁴ Selbst wenn Forscher die konkreten Fassungen der Ausgangstexte mit Hilfe von Software auflösen,⁵⁵ besteht in Einzelfällen die Gefahr, dass die schöpferischen Züge von Romanszenen oder -figuren in abgeleiteten Textformaten erkennbar bleiben. Ein solcher Figurenschutz ist zwar selten und auf Erzählungen mit hoher individueller Prägung beschränkt, wie etwa *Astrid Lindgrens* Pippi Langstrumpf, *Boris Pasternaks* Dr. Schiwago⁵⁶ und auch *J.K. Rowlings* Harry Potter.⁵⁷ Ungeachtet dessen sind die Motive und Ideen aus Romanen weiterhin Gemeingut,⁵⁸ sodass auch andere Geschichten über selbstbewusste junge Mädchen aus Schweden und die Abenteuer von Zauberlehrlingen erzählen und Rezensenten über die Inhalte berichten dürfen.⁵⁹ Es bleibt jedoch eine risikobehaftete Wertung, wie stark die Individualität des Ausgangswerkes ist und deshalb in einem abgeleiteten Textformat erkennbar bleibt.

⁵⁴ BGH, Urt. v. 17.7.2013 – I ZR 52/12, GRUR 2014, 258 Tz. 25 f. – *Pippi-Langstrumpf-Kostüm I; Loewenheim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 103; *G. Schulze*, in: D/S (Fn. 18), UrhG, § 2 Rn. 86.

⁵⁵ Hierzu *Grisse*, Teil 3, B. in diesem Band, 141, 152 f.

⁵⁶ BGH, Urt. v. 29.04.1999 – I ZR 65/96, Z 141, 267, 279 = GRUR 1999, 984 – *Laras Tochter*.

⁵⁷ Vgl. *Loewenheim/Leistner*, in: S/L, 5. Aufl. 2017, UrhG, § 2 Rn. 103. Anderes noch (nach damals erst vier erschienenen Bänden) LG Köln, Urt. v. 7.3.2001 – 28 O 14/01, GRUR-RR 2002, 3, 4 – *Harry Potter*. Zum urheberrechtlichen Figurenschutz *Stieper*, GRUR 2017, 649, 651 ff.

⁵⁸ *Knopp*, GRUR 2010, 28, 29.

⁵⁹ Vgl. BGH, Urt. v. 1.12.2010 – I ZR 12/08, GRUR 2011, 134 Tz. 36 f. – *Perlentaucher*.

Weniger individuell sind Schilderungen, die Fiktion und Wirklichkeit miteinander verschwimmen lassen. Historienromane und Biografien bauen auf Tatsachen auf und verfremden diese mehr oder weniger stark. Die geschichtlichen Vorbilder und Abläufe sind gemeinfrei und leisten deshalb keinen Beitrag zur Individualität des Textes. An den Ereignissen selbst, die *Anne Frank* in ihren Tagebüchern schildert, besteht daher kein Urheberrecht. Ihre konkreten sprachlichen und gestalterischen Darstellungen ermöglichen jedoch „einzigartige Einblicke in ihre Gedankenwelt“ und sind ihrerseits individuell und geschützt.⁶⁰ Gleiches gilt für die Geschichten, die *Jonas Jonassons* Romanheld Allan Karlsson erlebt. Nicht die Erzählungen der vielen historischen Vorgänge sind es, die hier die eigenschöpferischen Züge ausmachen. Mit dem Urheberrecht belohnt wird aber die einzigartige, ironische und fantasievolle Verknüpfung der historischen Vorlagen mit dem fiktiven Leben des Protagonisten, der den Mächtigen und der Gesellschaft bis hinein in die Gegenwart den Spiegel vorhält.⁶¹

Besondere Herausforderungen bestehen im Übrigen bei der IT-gestützten Auswertung von Lyrik. Solche Texte sind zwar in der Tendenz kürzer. Lyrische Texte zeichnen sich jedoch durch eine hohe schöpferische Gestaltung aus. Dass in München lebende Menschen eher zu privilegierten Kreisen gehören, ist daher keine Erkenntnis des Lyrikers *Eugen Roth*. Die originelle Formulierung

*„Vom Ernst des Lebens halb verschont
Ist der schon, der in München wohnt.“*

hielt das OLG München dennoch für urheberschutzfähig und verlangte von einem Verlag deshalb, eine Lizenz des Autors einzuholen, um diesen kurzen Reim zu verwerten.⁶²

⁶⁰ OLG Hamburg, Urt. v. 14.4.2016 – 5 U 117/12, ZUM-RD 2016, 576, 588 – *Tagebuch der Anne Frank*.

⁶¹ Vgl. *Binal*, Ein ironischer Ritt durch die Gegenwart, Deutschlandfunk Kultur, Beitrag vom 8.9.2018, abrufbar unter: <https://t1p.de/dlf-rez-jonasson>, zuletzt abgerufen am: 25.3.2020.

⁶² OLG München, Urt. v. 17.9.2009 – 29 U 3271/09, ZUM 2009, 970 f. – *Typisch München!*.

bb) Gebrauchstexte

Größere Freiräume haben Forschende, die für ihre TDM-Projekte Gebrauchstexte nutzen. Nachrichten, Werbeaussagen, Beipackzettel oder Bedienungsanleitungen weisen im Vergleich zu anderen Textarten meist eine hohe Dichte an Tatsachen, Informationen oder wissenschaftlichen Lehren und damit gemeinfreien Inhalten auf.⁶³ Wie gezeigt, verlangen die Gerichte bei Gebrauchstexten ein höheres Maß an individueller Leistung.⁶⁴ So verneinte etwa das LG Bielefeld einen Urheberrechtsschutz für den Tweet „Wann genau ist aus „Sex, Drugs & Rock n Roll“ eigentlich „Laktoseintoleranz, Veganismus und & Helene Fischer“ geworden?“⁶⁵. Trotz der prägnanten und ironischen Beschreibung der Ernährungsgewohnheiten junger Großstädter meinte das Gericht, dass der Tweet das Maß des Alltäglichen nicht deutlich übersteige.⁶⁵ Jedenfalls im Ergebnis ist die Zurückhaltung des Landgerichts richtig. Einzelne Wörter, Satzteile oder schlichte Beschreibungen gesellschaftlicher Prozesse dürfen niemandem exklusiv durch das Urheberrecht zugeschrieben werden. Gleiches gilt für kurze Tickermeldungen und Presseberichte, die gemeinfreie Nachrichten wiedergeben. Der Vergleich mit dem Reim von *Eugen Roth* zeigt jedoch, warum die Unterscheidung zwischen verschiedenen Textarten Forschenden in der Praxis kaum Nutzen bringt. Denn die Grenze zwischen zweckfreier Literatur und Gebrauchstexten ist durchaus fließend und damit kaum rechtssicher softwaregestützt im Einzelfall zu bestimmen.

Hinzu kommt, dass auch Gebrauchstexte die erforderliche Gestaltungshöhe aufweisen können. Urheberrechtlichen Schutz genießen oft z.B. ausführlichere journalistische Beiträge, Interviews oder umfangreiche Hintergrundberichte. In diesen Fällen überragen die Auswahl, Anordnung oder Darstellung der Tatsachen oft das Maß alltäglicher journalistischer Leistungen deutlich.⁶⁶ Daher besteht vor allem bei journalistischen Texten das Risiko, dass Übernehmende in Urheberrechte von Journalisten eingreifen.⁶⁷ Das zeigt sich exemplarisch im Fall der „Afghanistan Papiere“. Weder der Inhalt noch die Gestaltung dieser militärischen Lageberichte sind wirklich kreativ. Sie fassen die Situation der Bundeswehr im Einsatzgebiet zusammen und geben damit hauptsächlich gemeinfreie

⁶³ S.o. Fn. 42.

⁶⁴ S.o. Fn. 48.

⁶⁵ LG Bielefeld, Urt. v. 3.1.2017 – 4 O 144/16, MMR 2017, 641 Rn. 10.

⁶⁶ *Loewenheim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 142.

⁶⁷ *Jotzo*, AfP 2019, 481 Rn. 15 mwN.

Fakten wieder. Form und Sprache sind außerdem im klassischen Behördendeutsch auf Kürze und Verständlichkeit getrimmt. Dennoch bejahte das LG Köln den Urheberschutz, indem es in der Auswahl und prägnanten Zusammenfassung der Fakten eine hinreichend individuelle Leistung der Referenten sah.⁶⁸ Ähnlich diffus ist die Grenze zwischen urheberrechtlich geschützten Leistungen und gemeinfreien Elementen bei wissenschaftlichen Texten. Einigkeit besteht zwar vordergründig darüber, dass wissenschaftliche Lehren, Methoden und Erkenntnisse im Interesse der Wissenschaftsfreiheit nicht monopolisiert werden dürfen.⁶⁹ Wissenschaftliche Leistungen werden dennoch nicht per se vom Urheberschutz ausgenommen. Soweit die Fachsprache und -methode Gestaltungsspielräume zulassen, können Forschende mit ihren Formulierungen, ihrer Gedankenführung und ihren Beispielen individuelle Werke schaffen.⁷⁰ Selbst den Abstract eines wissenschaftlichen Gutachtens hielt das LG Köln für schutzfähig. Dessen Autor habe schließlich ein „umfangreiches wissenschaftliches Werk in einer prägnanten, auch für Laien verständlichen Weise“ zusammengefasst und nachvollziehbar dargestellt.⁷¹

Abgesehen von der Arbeit mit einzelnen Wörtern und kurzen N-Grammen können Forschende, die im Rahmen von TDM-Projekten automatisiert Textbestände nutzen, praktisch kaum vermeiden, in die Nutzungsrechte an fremden Sprachwerke einzugreifen. Je größer die genutzten Textausschnitte und Textbestände sind, desto stärker verdichtet sich die Möglichkeit zur Gewissheit. Verlässliche Grenzen gibt es hier jedoch kaum. Die Individualität – der Kern des Werkbegriffs – ist schillernd und entzieht sich schließlich schematischen Bewertungen. Nötig ist die gezeigte normative Gesamtschau verschiedener Kriterien im Einzelfall.

⁶⁸ LG Köln, Urt. v. 2.10.2014 – 14 O 333/13, GRUR-RR 2015, 55, 56; bestätigt durch OLG Köln, Urt. v. 12.6.2015 – 6 U 5/15, GRUR-RR 2016, 59 Tz. 25 ff. Mit berechtigten Zweifeln dagegen BGH, Vorlagebeschluss v 1.6.2017 – I ZR 139/15 = GRUR 2017, 901 Tz. 13 – *Afghanistan Papiere*, und EuGH, Urt. v. 29.7.2019 – C-469/17, ECLI:EU:C:2019:623 Tz. 24 – *Funke Medien NRW GmbH*. Kritisch zum Urheberschutz auch *Hoeren*, MMR 2017, 684.

⁶⁹ *Loewenbeim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 85 mwN.

⁷⁰ *Loewenbeim/Leistner*, in: S/L (Fn. 10), UrhG, § 2 Rn. 86.

⁷¹ LG Köln, Urt. v. 15.12.2016 – 14 O 302/15, BeckRS 2016, 133275 Tz. 58, bestätigt durch OLG Köln, Urt. v. 6.12.2017 – 6 U 8/17, BeckRS 2017, 151600 – *Glyphosat Gutachten*.

2. Wissenschaftliche Ausgaben (§ 70 UrhG) und nachgelassene Werke (§ 71 UrhG)

Neben Eingriffen in fremde Urheberrechte drohen durch TDM-Projekte zudem Eingriffe in die Leistungsschutzrechte für wissenschaftliche Ausgaben (§ 70 UrhG) und nachgelassene Werke (§ 71 UrhG).

Wissenschaftliche Rekonstruktionen historischer Originaltexte sind aufwändig, in der Regel aber nicht schöpferisch.⁷² § 70 UrhG belohnt deshalb die Verfasser solcher Editionen für ihre wissenschaftlich sichtende Tätigkeit mit einem Leistungsschutzrecht für die Dauer von 25 Jahren nach dem Erscheinen der Edition.⁷³ Das Leistungsschutzrecht umfasst diejenigen Teile der Edition, die auf einer wissenschaftlich sichtenden Tätigkeit beruhen.⁷⁴ Den Verfassern stehen die gleichen persönlichkeitsrechtlichen und verwertungsrechtlichen Befugnisse zu, wie Urhebern (§ 70 I UrhG).⁷⁵ Hierzu gehören vor allem das Vervielfältigungsrecht (§ 16 UrhG) und das Recht der öffentlichen Zugänglichmachung (§ 19a UrhG), in die IT-gestützte Forschung regelmäßig eingreift.⁷⁶

Das Leistungsschutzrecht aus § 71 I UrhG belohnt denjenigen, der ein bislang *nicht erschienenes* Werk – dessen Schutzfrist inzwischen abgelaufen ist – erscheinen lässt oder öffentlich wiedergibt. Die Schutzdauer beträgt 50 Jahre nach Erscheinen des nachgelassenen Werkes. Nach herrschender Meinung steht das Recht dem Herausgeber zu⁷⁷ und umfasst die urheberrechtlichen Verwertungsrechte (§ 71 I 3 UrhG).

3. Leistungsschutzrecht der Presseverleger, §§ 87 f-h UrhG

Im Auge behalten müssen Forschende auch die aktuellen Entwicklungen zum Leistungsschutzrecht der Presseverleger.

De lege lata beschränkt das in §§ 87 f-h UrhG geregelte Leistungsschutzrecht⁷⁸ die Arbeit Forschender zwar nicht. Denn der deutsche Gesetzgeber hatte diese

⁷² *Lauber-Rönsberg*, in: M/N (Fn. 19), UrhG, § 70 Rn. 1.

⁷³ Zum Schutzzumfang *Thum*, in: W/B (Fn. 21), UrhG, § 70 Rn. 20; und BGH, Urt. v. 23.05.1975 – I ZR 22/74, GRUR 1975, 667 – *Reichswehrprozess*.

⁷⁴ *Loewenheim*, in: S/L (Fn. 10), UrhG, § 70 Rn. 10.

⁷⁵ *Schack*, (Fn. 10) Rn. 736.

⁷⁶ Siehe oben II.

⁷⁷ Str. *Lauber-Rönsberg*, in: M/N (Fn. 19), UrhG, § 71 Rn. 30 mwN.; *Schack*, (Fn. 10) Rn. 738 ff.; ausführlich *Langer*, *Der Schutz nachgelassener Werke*, Göttingen 2012, 147 ff.

⁷⁸ Zum rechtspolitischen Hintergrund *Stieper*, in: S/L (Fn. 10), UrhG, vor §§ 87f ff. Rn. 2 ff.

Regeln unter Verstoß gegen die unionsrechtliche Notifizierungspflicht des Art. 8 Transparenz-RL⁷⁹ erlassen,⁸⁰ sodass die §§ 87 f-h UrhG bis auf Weiteres keine Anwendung finden. Die Nachfolgeregeln werden jedoch nicht lange auf sich warten lassen. Da nationale Alleingänge gegenüber *Google* und Co. wenig erfolgversprechend sind, haben sich besonders die deutschen Presseverleger für ein Pendant auf EU-Ebene eingesetzt.⁸¹ Sie meinen, nur auf diese Weise ihr Ertragsmodell gegenüber Nachrichtenaggregatoren wie *Google News* bewahren zu können.⁸² Das Ergebnis ist Art. 15 der DSM-RL.⁸³ Die Bundesregierung will die Richtlinie an dieser Stelle rasch umsetzen und hat hierzu bereits einen Diskussionsentwurf vorgelegt.⁸⁴ Gegenstand des europäischen Leistungsschutzrechts sind Presseveröffentlichungen iSv. Art. 2 Nr. 4 DSM-RL, zu denen journalistische Beiträge in Tageszeitungen, Zeitschriften und auf Nachrichtenwebseiten gehören.⁸⁵ Neben wissenschaftlichen Fachzeitschriften sind vom Schutz ausgenommen Blog-Beiträge, die unter keiner redaktionellen Verantwortung und Aufsicht „eines Dienstleisters“ entstanden sind (ErwGr. 56 aE DSM-RL), wie auch immer das in der Praxis IT-gestützt ermittelt werden soll. Die Schutzdauer wird mindestens zwei Jahre nach der erstmaligen Veröffentlichung der Presseerzeugnisses betragen.⁸⁶ Umfassen wird das neue Leistungsschutzrecht neben dem Recht der öffentlichen Zugänglichmachung auch vorbereitende Vervielfältigungen für die spätere Online-Nutzung.⁸⁷ Im Ausgangspunkt kann das neue Leistungsschutzrecht daher für Forschende bei TDM-Prozessen ein Thema

⁷⁹ RL 98/34/EG v. 22.6.1998 über ein Informationsverfahren auf dem Gebiet der Normen und technischen Vorschriften, geändert durch RL 98/48/EG v. 20.7.1998; abgelöst durch RL (EU) 2015/1535 v. 9.9.2015.

⁸⁰ EuGH, Urt. v. 12.9.2019 – C-299/17, ECLI:EU:C:2019:716, GRUR 2019, 1188 Tz. 39 f. – *VG Media*, ausführlich die Anm. *Stieper*, GRUR 2019, 1264, 1265 ff.

⁸¹ *Spindler* CR 2019, 277, 281.

⁸² Vgl. ErwGr. 54 DSM-RL.

⁸³ Richtlinie (EU) 2019/790 v. 17.4.2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG, ABl. 2019 L 130, 92.

⁸⁴ Entwurf des BMJV v. 15.1.2020 für ein „Erstes Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarkts“, mit grundlegender Kritik *Schack*, ZUM 2020, 165f.

⁸⁵ Der Begriff der Presseveröffentlichung ist unabhängig vom Medium, ErwGr. 56 DSM-RL.

⁸⁶ § 87j UrhG-E iVm § 69 UrhG und Art. 15 Abs. 4 DSM-RL.

⁸⁷ § 87g UrhG-E, der Art. 15 I DSM-RL umsetzt. Werden Online-Presseveröffentlichungen im Rahmen von TDM-Prozessen vervielfältigt, ohne aber selbst später öffentlich zugänglich gemacht zu werden, fallen sie dagegen nicht in den Schutzbereich des neuen Leistungsschutzrechts, *Stieper*, ZUM 2020, 166, 168.

sein. In einem gewissen Rahmen sollten Forschende jedoch von den Bereichsausnahmen profitieren: Ausgenommen vom Leistungsschutzrecht sind etwa einzelne Wörter und „sehr kurze Auszüge einer Presseveröffentlichung“,⁸⁸ zu denen der deutsche Umsetzungsgesetzgeber Überschriften zählt.⁸⁹ Vom Schutz ausgeklammert sind außerdem private und nicht-kommerzielle Nutzungen durch einzelne Nutzer.⁹⁰ Darauf werden sich wohl jedenfalls echte nicht-gewinnorientiert arbeitende einzelne Forscher stützen können.

IV. Schutz von Textbeständen

Da Forschende im Rahmen von TDM-Prozessen meist mit großen Datenmengen arbeiten, müssen sie zudem beachten, dass neben den Schutzrechten an einzelnen Texten auch solche an Textbeständen in Betracht kommen.

1. Sammel- und Datenbankwerk

Die eigenschöpferische Auswahl oder Anordnung von Werken und Daten kann etwa als Sammel- und Datenbankwerk (§ 4 UrhG) Urheberschutz genießen. Hierzu zählen z.B. Anthologien, Lexika und Enzyklopädien, soweit deren Herausgeber nicht rein handwerklich vorgehen oder nach Vollständigkeit streben. Gleiches gilt für einzelne Ausgaben von Tageszeitungen oder Zeitschriften.⁹¹ Solche Sammelwerke werden zu Datenbankwerken, wenn die Elemente strukturiert und mit elektronischen Mitteln oder auf andere Weise zugänglich gemacht werden, § 4 II UrhG.⁹² Gegenstand des Schutzes ist die schöpferische Auswahl oder Anordnung der einzelnen Elemente, nicht jedoch die Inhalte der Datenbanken selbst (vgl. § 4 I UrhG).⁹³ In das Urheberrecht am Sammelwerk greift daher ein, wer diejenigen Strukturen übernimmt, die durch Auswahl oder Anordnung den schöpferischen Gehalt des Sammelwerkes ausmachen.⁹⁴ Viele TDM-Prozesse lassen sich so gestalten, dass sie sich außerhalb des Schutzbereichs vorbestehender Sammelwerke bewegen. Übernehmen Forschende die

⁸⁸ Dazu *Stieper*, ZUM 2020, 166, 168.

⁸⁹ § 87g Abs. 2 Nr. 3 iVm. Abs. 3 Nr. 1 UrhG-E. Beruht auf Art. 15 Abs. 1 Uabs. 4 DSM-RL.

⁹⁰ § 87g Abs. 2 Nr. 1 UrhG-E und Art. 15 Abs. 1 Uabs. 2 DSM-RL.

⁹¹ *Schack*, (Fn. 10) Rn. 289; *Leistner*, in: S/L (Fn. 10), UrhG, § 4 Rn. 28 mwN.

⁹³ *Czychowski*, in: F/N (Fn. 16), UrhG, § 4 Rn. 40; *Leistner*, in: S/L (Fn. 10), UrhG, § 4 Rn. 34.

⁹⁴ BGH, Urt. v. 27. 3. 2013 – I ZR 9/12, GRUR 2013, 1213 Tz. 57 mwN. – *SUMO*.

einzelnen Inhalte einer Datenbank und führen diese mit weiteren Texten zusammen, geht spätestens im Korpus diejenige Leistung verloren, die § 4 UrhG schützt.⁹⁵ Forschende sollten daher die TDM-Prozesse so aufbauen, dass die Strukturen vorbestehender Datenbestände möglichst in keinem Stadium der Verarbeitung übernommen werden.

2. Leistungsschutzrecht des Datenbankherstellers, §§ 87a ff. UrhG

Schwieriger zu vermeiden ist dagegen der *sui generis* Schutz von Datenbankherstellern (§ 87b I 1 UrhG). Dieses Recht schützt die in einer Datenbank verkörperte wesentliche Investition, die der Hersteller für die Beschaffung, Überprüfung oder Darstellung der Inhalte aufgebracht hat, § 87a I 1 UrhG.⁹⁶ Ebenso wie § 4 UrhG verschafft auch dieses Leistungsschutzrecht dem Hersteller kein Monopol an den einzelnen in der Datenbank gespeicherten Inhalten oder Informationen.⁹⁷ Datenbankhersteller können Dritten aber verbieten, die Datenbank oder wesentliche Teile davon zu entnehmen und weiterzuverwenden, § 87b I 1 UrhG iVm Art. 7 I Datenbank-RL.⁹⁸ Wesentlich sind solche Teile, deren Nutzung in qualitativer oder quantitativer Hinsicht einen erheblichen Schaden an der Investition verursacht.⁹⁹ Das ist wieder eine Frage des Einzelfalls.¹⁰⁰ Eine Entnahme ist jede Übertragung wesentlicher Teile der Datenbank auf einen anderen Datenträger, Art. 7 II lit. a Datenbank-RL. Der EuGH legt diesen Begriff weit aus und fasst darunter jede „unerlaubte Aneignung“ relevanter Inhalte, unabhängig von der Art und Form des angewandten Verfahrens.¹⁰¹ Wenn For-

⁹⁵ Vgl. *Spindler*, GRUR 2016, 1112, 1114.

⁹⁶ Art. 7 I RL Datenbank-RL (Richtlinie 96/9/EG v. 11.3.1996 über den rechtlichen Schutz von Datenbanken). Unberücksichtigt bleiben dabei die Kosten für die Erzeugung der Inhalte selbst, EuGH, Urt. v. 9.11.2004 – C-46/02, ECLI:EU:C:2004:694, GRUR Int 2005, 244 Tz. 36 – *Fixtures Marketing III*; *Hermes*, in: W/B (Fn. 21), UrhG, § 87a Rn. 36 mwN.

⁹⁷ ErwGr. 45 und 46 Datenbank-RL; *Hermes*, in: W/B (Fn. 21), UrhG, § 87b Rn. 4; *Czychowski*, in: F/N (Fn. 16), UrhG, § 87a Rn. 6.

⁹⁸ Umgesetzt durch § 87b I 1 UrhG. Dem gleichgestellt sind Entnahmen und Weiterverwendungen *unwesentlicher* Teile der Datenbank, die der normalen Nutzung der Datenbank entgegenstehen oder die Interessen des Datenbankherstellers unzumutbar beeinträchtigen, Art. 7 V Datenbank-RL (§ 87 I 2 UrhG).

⁹⁹ ErwGr. 42 Datenbank-RL; EuGH, Urt. v. 9.11.2004 – C-203/02, ECLI:EU:C:2004:695, GRUR Int 2005, 247 Tz. 69 – *The British Horseracing Board*; *Dreier*, in: D/S (Fn. 18), UrhG, § 87b Rn. 6;

¹⁰⁰ *Dreier*, in: D/S (Fn. 18), UrhG, § 87b Rn. 6.

¹⁰¹ EuGH, Urt. v. 5.3.2009 – C-545/07, ECLI:EU:C:2009:132, GRUR 2009, 572 Tz. 40 mwN. – *Apis-Hristovich*.

schende für die Korpuserstellung wesentliche Teile fremder Datenbanken crawlen, greifen sie also in das Leistungsschutzrecht ein.¹⁰² Eine relevante Weiterverwendung iSv. Art. 7 II lit. b Datenbank-RL liegt ferner vor, wenn jemand wesentliche Teile der Datenbank öffentlich verbreitet.¹⁰³ Das geschieht etwa, wenn Forschende in einem Korpus die entsprechenden Inhalte aus einer fremden Datenbank Dritten zur Verfügung stellen.

3. Umgehungsschutz für technische Schutzmaßnahmen, § 95a UrhG

Viele Texte befinden sich in zugangsbeschränkten Bereichen des Internets. Daher ist das Umgehungsverbot in § 95a UrhG ein weiterer Stolperstein für Forschende bei der Beschaffung von Inhalten. § 95a UrhG untersagt, wirksame technische Schutzmaßnahmen ohne Zustimmung des Rechtsinhabers zu umgehen. Technische Schutzmaßnahmen in diesem Sinne sind z.B. Zugangskontrollen¹⁰⁴ und DRM,¹⁰⁵ die viele Webangebote heute einsetzen. Das Umgehungsverbot gilt selbst dann, wenn die urheberrechtlichen Schranken an sich die angestrebten Verwertungshandlungen erlauben würden. In diesen Fällen dürfen Werknutzer nicht selbst die Schutzmaßnahmen überwinden, sondern sie können vom Rechtsinhaber nur verlangen, dass er ihnen die nötigen Mittel zur Verfügung stellt, § 95b II UrhG.¹⁰⁶

V. Fazit

Die individuelle Schöpfung ist das Kernelement des urheberrechtlichen Werkbegriffs. Die Länge und die Art des einzelnen Textes liefern erste Hinweise auf einen möglichen Werkschutz. Ob ein Text iSv. § 2 II UrhG hinreichend individuell ist, lässt sich jedoch nicht schematisch, sondern nur in einer wertenden Gesamtbetrachtung weiterer Kriterien ermitteln. Abgesehen von der Arbeit mit einzelnen Worten und kurzen N-Grammen macht diese komplexe Einzelfallprüfung es praktisch schwer, die urheberrechtlich geschützten Bestandteile von

¹⁰² *Spindler*, GRUR 2016, 1112, 1114; *Dreier*, in: D/S (Fn. 18), UrhG, § 87b Rn. 4; vgl. *Raue*, ZUM 2019, 684, 685.

¹⁰³ EuGH, Urt. v. 5.3.2009 – C-545/07, ECLI:EU:C:2009:132, GRUR 2009, 572 Tz. 49 mwN. – *Apis-Hristovich*.

¹⁰⁴ *Specht*, in: D/S (Fn. 18), UrhG, § 95a Rn. 14.

¹⁰⁵ *Wandtke/Obst*, in: W/B (Fn. 21), UrhG, § 95b Rn. 38. Ausführlich zum Einsatz von DRM bei E-Books *Henke*, E-Books im Urheberrecht, Göttingen 2018, 175 ff.

¹⁰⁶ *Wandtke/Obst*, in: W/B (Fn. 21), UrhG, § 95b Rn. 44.

Texten rechtssicher per Software zu erkennen und herauszufiltern.¹⁰⁷ Forschende müssen beim Einsatz von TDM-Anwendungen daher in der Regel damit rechnen, urheberrechtlich geschützte Inhalte zu vervielfältigen (§ 16 UrhG) und im Korpus unter Umständen öffentlich zugänglich zu machen (§ 19a UrhG). Eingriffe drohen auch in die Leistungsschutzrechte für nachgelassene Werke (§ 71 UrhG) und wissenschaftliche Ausgaben (§ 70 UrhG). Erfassen Forschende vorbestehende Textbestände, wie etwa digitale Archive, ist das Leistungsschutzrecht des Datenbankherstellers (§ 87a UrhG) ein weiteres Thema. Wie einleitend beschrieben, zeigt dieser Befund jedoch nur die eine Seite der Medaille. Eingriffe in Urheber- und Leistungsschutzrechte Dritter sind für die TDM-Forschung keine unüberwindbaren Hürden. Die Rechtsordnung gewährt diese Schutzrechte nur innerhalb der Schranken. Die neuen forschungsspezifischen TDM-Schranken haben hier den Raum der Möglichkeiten zwar spürbar erweitert. Mit Blick auf Streitige Auslegungsfragen verbleiben jedoch insbesondere bei der Weitergabe von Textkorpora an Dritte und bei Anschlussnutzungen aus Sicht von Forschenden Defizite. Eine interessante Alternative sind daher die im Rahmen dieses Projekts entwickelten Formate. *Grisse* zeigt im zweiten juristischen Beitrag des Projekts,¹⁰⁸ wie die Primärtexte unter Ausnutzung der Schranken des UrhG in solche Formate konvertiert werden können, die keine urheberrechtlich geschützten Bestandteile mehr enthalten, und die dennoch als wertvolle Ressource für viele Forschungsfelder taugen.

¹⁰⁷ *Grisse*, Teil 3, B. in diesem Band, 141, 147.

¹⁰⁸ *Grisse*, Teil 3, B. in diesem Band, 141, 144 ff.

C. Die Nutzung wissenschaftlicher Ausgaben für Textanalysen

Katharina Erler-Fridgen

Version 1.0 (27.05.2021), CC BY-SA 4.0.

Wer jüngere Editionen von Primärtexten in Text und Data Mining-Verfahren einsetzt, kann die maschinenlesbare Aufbereitung wesentlich vereinfachen.¹ Daneben können wissenschaftliche Editionen Quellen für zusätzliche Informationen sein, beispielsweise für Metadaten, orthographische Erklärungen und Interpretationen sowie vorhandene fachliterarische Analysen des Editors.² Werden diese unterschiedlichen Teile einer Edition genutzt, kann das in das Leistungsschutzrecht für wissenschaftliche Ausgaben nach § 70 UrhG eingreifen.³ Im Folgenden wird gezeigt, welche Teile einer wissenschaftlichen Ausgabe frei benutzt werden können und welche Teile urheberrechtlichen Restriktionen unterliegen.

I. Voraussetzungen und Anwendbarkeit des Leistungsschutzrechts

Nach § 70 UrhG wird die wissenschaftliche Leistung geschützt, gemeinfreie Werke oder urheberrechtlich nicht geschützte Texte in einer Ausgabe zusammenzuführen und zu editieren.⁴ Das Leistungsschutzrecht steht dem Verfasser einer solchen Ausgabe nach § 70 UrhG zu, wenn sie das Ergebnis wissenschaftlich sichtender Tätigkeit darstellt und sich wesentlich von den bisher bekannten Ausgaben der Werke unterscheidet. § 70 UrhG billigt dem Verfasser einer wissenschaftlichen Ausgabe also ein Leistungsschutzrecht unterhalb des Werk-schutzes zu.⁵

¹ Der Einsatz solcher jüngerer Textquellen kann etwa den Einsatz des Optical Recognition Verfahrens erleichtern.

² Ein Beispiel für eine solche wissenschaftliche Ausgabe ist etwa: *Annie Rivara* (Hrsg.), *Du Laurens, Imirce ou la Fille de la Natur*.

³ Zur Abgrenzung von eigenschöpferischen Leistungen des Editors siehe unten.

⁴ Aus welchen Gründen die in der Ausgabe enthaltenen Werke oder Texte nicht urheberrechtlich geschützt sind ist dabei unerheblich, siehe *Loewenheim*, in Schrickler/Loewenheim, *Urheberrecht*, 6. Aufl. 2020, § 70 Rn. 5.

⁵ *A. Nordemann*, in *Fromm/Nordemann, Urheberrecht*, 12. Aufl. 2018, § 70 Rn. 1.

Dafür muss die Edition das Ergebnis wissenschaftlich sichtender Tätigkeit sein, also über das reine Auffinden eines Schriftstücks hinaus gehen. Erforderlich ist eine sichtende, ordnende und abwägende Tätigkeit unter Anwendung wissenschaftlicher Methoden.⁶ Häufig liegt diese in einer text- und quellenkritischen Tätigkeit zur Rekonstruktion der Originalfassung.⁷ Es existieren darüber hinaus aber auch weitere Fälle, die eine wissenschaftlich sichtende Tätigkeit darstellen. Eine solche wurde beispielsweise in einem Fall angenommen, in dem der Wortlaut von Verhandlungen vor dem Reichsgericht rekonstruiert und dazugehörige Zeitungsartikel aufgearbeitet wurden.⁸ Auch wenn Rechtschreibung und Zeichensetzung modernisiert, Übersetzungen eingefügt und Textunterschiede herausgearbeitet werden, kann eine wissenschaftlich sichtende Tätigkeit vorliegen.⁹ Ein Fußnotenapparat ist dabei keine Schutzvoraussetzung,¹⁰ kann jedoch zum einen ein Beweisanzeichen für das wissenschaftliche Arbeiten¹¹ und zum anderen ein Teil der eigenschöpferischen Tätigkeit des Editors sein.¹² Für eine wissenschaftlich sichtende Tätigkeit reicht es hingegen nicht aus, wenn aufgefundene alte Werke wahllos veröffentlicht werden.¹³

Die Ausgabe muss sich dem Gesetzestext in § 70 UrhG nach zusätzlich von bisher bekannten Ausgaben wesentlich unterscheiden. Daher sind erneute Ausgaben desselben Werkes möglich. Sie begründen aber nur dann ein eigenes Leistungsschutzrecht, wenn sie sich wesentlich von der Vorausgabe unterscheiden.¹⁴ Diese Voraussetzung soll der Rechtssicherheit dienen und bedeutet, dass sich bei einer Verwertung leicht feststellen lassen muss, auf welcher Ausgabe die vorliegende Ausgabe beruht.¹⁵ Dies führt in Auslegung des § 70 UrhG dazu, dass

⁶ BGH GRUR 1975, 667, 668 – Reichswehrprozess, *Loewenheim*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 70 Rn. 6.

⁷ *Loewenheim*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 70 Rn. 6.

⁸ BGH GRUR 1975, 667, 668 – Reichswehrprozess.

⁹ KG GRUR 1991, 596, 597 – Schopenhauer-Ausgabe.

¹⁰ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 734.

¹¹ *Thum*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 70 Rn. 10.

¹² Zu eigenschöpferischen Teilen seitens des Editors und deren urheberrechtlichen Schutzfähigkeit siehe unten.

¹³ OLG Braunschweig GRUR 1974, 411 – Zinn-Stadtmarken.

¹⁴ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 735.

¹⁵ Unter der Annahme, dass für diesen Zweck keine hohen Anforderungen an die Wesentlichkeit des Unterschieds gestellt werden, siehe ausführlich *Thum*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 70 Rn. 11.

auch relativ geringfügige bzw. wenige Änderungen dafür ausreichen können, einen wesentlichen Unterschied zu bejahen.¹⁶ Der Schutz ist aber dann abzusprechen, wenn kaum ein Unterschied zu einer Vorfassung besteht, die Edition also mit einer Vorausgabe im Wesentlichen übereinstimmt.¹⁷ Beispielsweise unterscheidet sich eine Ausgabe nicht wesentlich, wenn bereits bekannte Einzelwerke in einer Gesamtausgabe zusammengefasst werden oder lediglich ihre Reihenfolge geändert wird.¹⁸

§ 70 UrhG ist auch dann anwendbar, wenn Verletzungen des Leistungsschutzrechts in Rede stehen, weil beispielsweise französische wissenschaftliche Ausgaben in Deutschland für Textanalysen genutzt werden. § 124 UrhG stellt klar, dass auch für den Schutz wissenschaftlicher Ausgaben die fremdenrechtliche Regelung des § 120 UrhG Anwendung findet. Das bedeutet, dass aufgrund des allgemeinen Diskriminierungsverbots in Art. 18 AEUV auch ein EU-Bürger als Verfasser einer wissenschaftlichen Ausgabe nach § 120 Abs. 2 UrhG wie ein Deutscher gegen die Verletzung des Leistungsschutzrechtes aus § 70 UrhG geschützt ist.¹⁹ Bürger ausländischer Drittstaaten (EU-Ausländer) können, da die Staatsverträge auf dem Gebiet des Urheberrechts²⁰ das Leistungsschutzrecht für wissenschaftliche Ausgaben als verwandtes Schutzrecht nicht erfassen und keine Bekanntmachung nach § 121 Abs. 4 S. 2 UrhG vorliegt, den Schutz des § 70 UrhG dann nutzen, wenn die Ausgabe erstmals – oder spätestens 30 Tage nach Erscheinen im Ausland – in Deutschland erschienen ist.²¹

II. Gegenstand des Leistungsschutzrechts

Inhaltlich ist das Leistungsschutzrecht nach § 70 UrhG dem Urheberrecht gleichgestellt, es stehen dem Verfasser alle Befugnisse der §§ 11 ff. UrhG zu.²² Die kürzere Schutzfrist beträgt nach § 70 Abs. 3 UrhG 25 Jahre ab Erscheinen

¹⁶ *Thum*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 70 Rn. 13.

¹⁷ *Loewenheim*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 70 Rn. 8.

¹⁸ *Loewenheim*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 70 Rn. 8.

¹⁹ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 120 Rn. 8.

²⁰ Ausführlich zu den einzelnen Staatsverträgen auf dem Gebiet des Urheberrechts nach § 121 Abs. 4 UrhG: *Katzenberger/Metzger*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 121 Rn. 11 ff.

²¹ *Katzenberger/Metzger*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 124 Rn. 3, *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 124 Rn. 2.

²² *Thum*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 70 Rn. 18.

der Ausgabe.²³ Das Recht erlischt 25 Jahre nach Herstellung der Ausgabe, wenn die Ausgabe nicht innerhalb von 25 Jahren erscheint.

Dieses Leistungsschutzrecht entsteht jedoch nicht mit Blick auf das zugrundeliegende Primärwerk und dessen Originaltext.²⁴ Es umfasst lediglich die Teile der Ausgabe, die Ergebnis der wissenschaftlich sichtenden Tätigkeit sind.²⁵ Wer lediglich das Primärwerk nutzt, das selbst nicht wissenschaftlich-sichtend überarbeitet wurde, verletzt das Leistungsschutzrecht an der wissenschaftlichen Ausgabe grundsätzlich nicht.²⁶ Es stellt daher dementsprechend keinen Eingriff dar, wenn allein die in der wissenschaftlichen Ausgabe enthaltene, unüberarbeitete, Primärquelle genutzt wird (siehe Abbildung 1 unten, links, grün markiert).²⁷ Anderes gilt, wenn der Primärtext selbst das Ergebnis wissenschaftlich-sichtender Tätigkeit ist (hierzu unten).

Verletzt wird das Leistungsschutzrecht nur dann, wenn diejenigen Teile der Ausgabe übernommen werden, die das Ergebnis der wissenschaftlich sichtenden Tätigkeit waren und sich von bisher bekannten Ausgaben wesentlich unterscheiden.²⁸ Daher kann die Übernahme etwa von Erläuterungen, Fußnoten des Verfassers der Ausgabe, Nachworten usw. (siehe Beispiel in Abbildung 1 unten, rechts, orange markiert) in das Leistungsschutzrecht des Verfassers eingreifen.

²³ Das Urheberrecht erlischt hingegen 70 Jahre nach dem Tode des Urhebers nach § 64 UrhG.

²⁴ Für den Schutz des Originalwerks kommt lediglich der Schutz als nachgelassenes Werk nach § 71 UrhG in Betracht, falls das Werk bisher noch nicht erschienen oder öffentlich wiedergegeben ist und gemeinfrei ist, siehe *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 6. Aufl. 2018, § 70 Rn. 2.

²⁵ *Thum*, in *Wandtke/Bullinger*, Urheberrecht, 5. Aufl. 2019, § 70 Rn. 19.

²⁶ *A. Nordemann*, in *Fromm/Nordemann*, Urheberrecht, 12. Aufl. 2018, § 70 Rn. 18.

²⁷ Für den Fall, dass das Primärwerk selbst das Ergebnis wissenschaftlich-sichtender Tätigkeit ist, siehe unten.

²⁸ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 6. Aufl. 2018, § 70 Rn. 9.



Abbildung 1: Ein wissenschaftlich-sichtend unüberarbeiteter Primärtext (links, grün) ist (inklusive gegebenenfalls enthaltener Anmerkungen des Originalautors links unten, grün) nutzbar, ohne das Leistungsschutzrecht zu verletzen. Erläuterungen des Verfassers der wissenschaftlichen Ausgabe (rechts, orange) können dem Leistungsschutzrecht unterfallen.²⁹ Ein Beispiel für enthaltene Erläuterungen des Verfassers der wissenschaftlichen Ausgabe wäre etwa die Ausgabe von Michel De-lon, Gabriel Sénac de Meilhan, *L'Émigré*, Paris 2004.

Wichtig ist aus diesem Grund, dass Erläuterungen und Fußnoten des Verfassers der wissenschaftlichen Ausgabe getrennt vom Fußnotenapparat des Primärwerks betrachtet werden. Dabei ist zu beachten, dass Anmerkungen und Fußnoten des Verfassers der wissenschaftlichen Ausgabe gegebenenfalls auch getrennt vom Primärtext und dessen Anmerkungen aufgeführt werden (siehe *Abbildung 1*, rechts unten in orange und links unten in grün). In manchen Fällen aber befinden sich Fußnoten und Anmerkungen des Verfassers der wissenschaftlichen Ausgabe in direkter Kombination mit dem Primärtext (siehe *Abbildung 2* unten links, orange). Die Übernahme solcher Anmerkungen und Fußnoten des Verfassers der wissenschaftlichen Ausgabe kann in sein Leistungsschutzrecht eingreifen.



Abbildung 2: Hier befinden sich Anmerkungen/Fußnoten des Verfassers der wissenschaftlichen Ausgabe direkt am Primärtext (links unten, orange). Es muss beachtet werden, dass an diesen Erläuterungen das Leistungsschutzrecht bestehen kann. Ein Beispiel für Anmerkungen des Verfassers der wissenschaftlichen Ausgabe direkt am Primärtext wäre etwa die Ausgabe von Maurice Lévy, Louis-François-Marie, Célestine, ou les Époux sans l'être, Bellin de la Liborlière, 2011.

²⁹ Die Abbildungen und Beispiele sind in Zusammenarbeit mit Dr. Maria Hinzmann, Projektkoordinatorin des Projekts MiMoText (Universität Trier), entstanden.

Mit Blick auf die Nutzung des Primärtextes stellt sich die Situation anders dar, wenn auch der Primärtext selbst das Ergebnis wissenschaftlich-sichtender Tätigkeit ist (siehe Abbildung 3). Dann kann auch die Nutzung des Primärwerkes das Leistungsschutzrecht verletzen. Wurde nämlich der Primärtext wissenschaftlich fundiert aufgearbeitet, etwa um die Originalfassung herzustellen, so stellt auch dies eine wissenschaftlich sichtende Tätigkeit dar.³⁰ Beispielsweise werden in manchen historisch-kritischen Ausgaben mittels wissenschaftlicher Kriterien unterschiedliche Fassungen des Primärtextes aus verschiedenen Ausgaben kombiniert, um eine möglichst originalgetreue Version des Primärtextes („besten Text“) zu generieren³¹. In diesen Fällen kann auch am wissenschaftlich erstellten Primärtext das Leistungsschutzrecht entstehen.



Abbildung 3: Der Primärtext selbst (links oben) ist in diesem Fall auch das Ergebnis wissenschaftlich-sichtender Tätigkeit, beispielsweise wurde ein „bester Text“ aus verschiedenen Primärtextversionen generiert. Das Leistungsschutzrecht kann daher auch mit Blick auf den erarbeiteten Primärtext entstehen. Beispielsweise findet sich ein „bester Text“ in der Ausgabe von Lukasz Szkopinski, François Guillaume Ducray-Duminil, Victor ou l'enfant de la forêt, ed. Classiques Garnier 2019.

III. Eigenschöpferische Leistungen

Durch eigenschöpferische Leistungen des Verfassers kann, über das Leistungsschutzrecht nach § 70 UrhG hinaus, urheberrechtlicher Schutz an diesen Leistungen mit einer Schutzfrist³² von 70 Jahren nach dem Tod des Urhebers entstehen (siehe Abbildung 4).³³ Ergänzungen des Originaltexts, ein ergänzter Fußnotenapparat sowie Anmerkungen oder Berichte des Verfassers der Ausgabe können – neben dem Leistungsschutzrecht an den sonstigen Teilen der

³⁰ *Loewenbeim*, in Schrickler/Loewenbeim, Urheberrecht, 6. Aufl. 2020, § 70 Rn. 6.

³¹ Dieses Verfahren wird insbesondere dann angewandt, wenn das Original verloren ist.

³² Die Schutzfrist bestimmt sich nach § 64 UrhG.

³³ *Loewenbeim*, in Schrickler/Loewenbeim, Urheberrecht, 6. Aufl. 2020, § 70 Rn. 3.

Edition – eigenständig als Schriftwerk nach § 2 Abs. 1 Nr. 1 geschützt sein.³⁴ So können beispielsweise textkritische Nachworte, ein allgemeines Nachwort oder ausnahmsweise auch ein Register³⁵ eigenständigen Schutz als Sprachwerk erfahren.³⁶ Auch kann Werkschutz der Ausgabe durch eigenschöpferische Auswahl oder Anordnung der Textteile als Sammelwerk nach § 4 UrhG entstehen.³⁷ Dabei begrenzt sich der Urheberrechtsschutz auf die Teile der Ausgabe, die eine eigene schöpferische Leistung des Verfassers nach § 2 Abs. 2 UrhG darstellen.³⁸



Abbildung 4: Erläuterungen wie Nachworte oder textkritische Ausführungen des Verfassers der wissenschaftlichen Ausgabe (rechts) können neben dem Leistungsschutzrecht der wissenschaftlichen Ausgabe (orange) auch als eigenschöpferische Leistungen urheberrechtlichen Schutz erfahren (rot). Beispielsweise könnte in Erwägung gezogen werden, dass in der „Introduction“ der Ausgabe von Lukasz Szkopinski, François Guillaume Ducray-Duminil, Victor ou l'enfant de la forêt, ed. Classiques Garnier 2019 eine eigenschöpferische Leistung liegen könnte.

IV. Ergebnis

Ohne weiteres können das gemeinfreie Primärwerk sowie zu diesem Primärtext gehörende Fußnoten und Anmerkungen des Originalautors übernommen

³⁴ BGH GRUR 1980, 227, 231 – Monumenta Germanicae Historica, KG GRUR 1991, 596, 597 – Schopenhauer Ausgabe, siehe auch *Lauber-Rönsberg*, in BeckOK Urheberrecht, 27. Edition 2019, § 70 Rn. 2.

³⁵ KG GRUR 1991, 596, 598 – Schopenhauer-Ausgabe: Nach der Rechtsprechung des BGH genügt für die Schutzfähigkeit eines Registers die bloße Zusammenstellung einzelner Fakten nicht. Die Anordnung und Darbietung des Registers müssen sich als schöpferische geistige Leistung darstellen. Dies kann durch besondere Konzeption des Registers oder durch schöpferische Auswahl, Sammlung, Einteilung oder Anordnung des vorhandenen Stoffs geschehen.

³⁶ KG GRUR 1991, 596, 598 – Schopenhauer-Ausgabe.

³⁷ Zur Abgrenzung und dem Verhältnis des Leistungsschutzrechts zum urheberrechtlichen Schutz und bestehender Alternativität siehe *Lauber-Rönsberg*, in BeckOK Urheberrecht, 27. Edition 2019, § 70 Rn. 2, ausführlich hierzu auch: *Thum*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 70 Rn. 32 ff., für ein Nebeneinander der Schutzvorschriften siehe *Loewenheim*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 70 Rn. 3.

³⁸ *Loewenheim*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 70 Rn. 3.

werden, sofern der Primärtext nicht selbst das Ergebnis wissenschaftlich-sichtender Tätigkeit ist. Wenn der Primärtext selbst das Ergebnis wissenschaftlich-sichtender Tätigkeit ist, also beispielsweise ein „bester Text“ hergestellt wurde, dann kann das Leistungsschutzrecht der wissenschaftlichen Ausgabe nach § 70 UrhG der Nutzung entgegenstehen.

Urheberrechtlich schutzfähig können in jedem Fall diese Teile der wissenschaftlichen Ausgabe sein, die das Ergebnis wissenschaftlich sichtender Tätigkeit sind, also beispielsweise Erläuterungen, Nachworte oder Fußnoten des Verfassers der wissenschaftlichen Ausgabe. Diese können dem Leistungsschutzrecht der wissenschaftlichen Ausgabe nach § 70 UrhG mit einer Schutzfrist von 25 Jahren ab Erscheinen der Ausgabe unterliegen. Eigenschöpferische Leistungen des Verfassers können urheberrechtlichen Schutz nach § 2 UrhG mit einer Schutzfrist von 70 Jahren ab dem Tod des Urhebers auslösen. Übersetzungen können nach § 3 S. 1 UrhG eigenständig geschützt sein.³⁹

In diesen Fällen muss entweder die Zustimmung des Rechtsinhabers einholen oder auf urheberrechtliche Schranken (§§ 44a ff UrhG) zurückgreifen, wer urheberrechtlich relevante Nutzungshandlungen vornehmen möchte.

Leseempfehlungen zur vertiefenden Lektüre: *Loewenheim*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 70 und *Thum*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 70.

³⁹ Diese können wie selbstständige Werke nach § 3 S. 1 UrhG geschützt sein: *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 269, *Loewenheim*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 3 Rn. 35.

D. Datenbanken als Quelle oder Ergebnis von Textanalysen

Datenbankenwerkschutz und das Leistungsschutzrecht des Datenbankherstellers

Katharina Erler-Fridgen

Version 1.0 (02.06.2022), CC BY-SA 4.0.

Werden Texte gesammelt und Informationen extrahiert und zusammengeführt, so gewinnt der Schutz von Datenbanken in zweierlei Weise an Bedeutung: Zum einen kann die Entnahme von Material aus Datenbanken urheberrechtlichen Restriktionen unterliegen, wenn Datenbanken etwa als Quelle für Textanalysen genutzt werden. Zum anderen kann aber auch eine neu entstandene Sammlung – etwa als Ergebnis von Text und Data Mining-Analysen – als Datenbank eigenständigen Schutz erfahren. In beiden Fällen können sowohl der urheberrechtliche Schutz als Datenbankwerk nach § 4 Abs. 2 UrhG als auch das Leistungsschutzrecht des Datenbankherstellers nach § 87a UrhG berührt werden. Zur Schaffung oder Abfrage der Datenbanken erforderliche Computerprogramme können außerdem nach § 69a ff. UrhG geschützt sein.¹

I. Perspektiven auf den Datenbankschutz bei Textanalysen

Wie bereits beschrieben, spielt der Datenbankschutz im Forschungsprozess aus zwei unterschiedlichen Perspektiven eine Rolle bei dem Text und Data Mining. Einerseits können Datenbanken als Quelle für Ausgangstexte für Textanalysen genutzt werden. Hierfür ist relevant, ob eine Quelle als Datenbankwerk (II.) oder durch das *sui-generis* Leistungsschutzrecht des Datenbankherstellers

¹ Zu solchen Datenbankverwaltungs- oder Datenbankmanagementsystemen siehe *Leistner*, in *Schricker/Loewenheim*, *Urheberrecht*, 6. Aufl. 2020, § 4 Rn. 58; siehe auch *Ahlberg/Lauber-Rönsberg*, in *Ahlberg/Götting/Lauber-Rönsberg*, *BeckOK Urheberrecht*, 33. Edition 2022, § 4, Rn. 24.

(III.) geschützt ist. Insbesondere im Verfahrensschritt² der Sammlung von Textquellen kann als urheberrechtlich relevante Handlung bspw. die Entnahme (IV.) aus geschützten Datenbanken vorgenommen werden. Andererseits kann das Ergebnis von Textanalyseverfahren, zum Beispiel eine Sammlung gewonnener Erkenntnisse, möglicherweise selbst als Datenbankwerk (II.) oder durch das Leistungsschutzrecht des Datenbankherstellers (III.) geschützt sein. Hierfür müssten die jeweiligen Voraussetzungen nach § 4 Abs. 2 UrhG (Datenbankwerk, II.) oder § 87a ff. UrhG (Leistungsschutzrecht des Datenbankherstellers, III.) erfüllt sein.

Im Ergebnis beeinflussen also die Voraussetzungen des Datenbankrechts sowohl die Frage nach der Entnahme und Weiterverwendung, wenn Ausgangstexte gesammelt werden, als auch die Frage nach der Rechtsinhaberschaft, wenn etwa der Schutz der Sammlung von Ergebnissen von Textanalyseverfahren in den Blick genommen wird.

II. Urheberrechtlicher Schutz als Datenbankwerk³

1. Schutzvoraussetzungen

Sammlungen von Werken oder Daten können als *Datenbankwerk* urheberrechtlich geschützt sein. Als Unterfall des Schutzes als Sammelwerk erfasst § 4 Abs. 1 UrhG auch den Schutz von Datenbankwerken⁴ nach § 4 Abs. 2 UrhG. Mit Umsetzung der Datenbank-Richtlinie (96/9/EG)⁵ und dessen Artikel 3 sind auch Zusammenstellungen von reinen Fakten (§ 4 Abs. 1 UrhG: „Daten oder sonstige Elemente“), also Sammlungen von Daten und sonstigen Elementen, unabhängig vom schöpferischen Wert dieser Elemente als Datenbankwerke schutzfähig.⁶

² Zu den iterativen Verfahrensschritten im Rahmen des Text und Data Mining siehe *Erler-Fridgen*, Teil 1, C. in diesem Band.

³ Dank gilt Prof. Dr. Christof Schöch und Dr. Maria Hinzmann (Projekt MiMoText, Universität Trier) für den Austausch zu Beispielen aus den Digital Humanities.

⁴ Die Legaldefinition eines Datenbankwerkes findet sich in § 4 Abs. 2 UrhG: Datenbankwerk im Sinne dieses Gesetzes ist ein Sammelwerk, dessen Elemente systematisch oder methodisch angeordnet und einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich ist.

⁵ Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11. März 1996 über den rechtlichen Schutz von Datenbanken.

⁶ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 4 Rn. 2.

Für solche Datenbankwerke begründet – wie bei Sammelwerken⁷ – der schöpferische Wert der Auswahl oder Anordnung der Elemente den urheberrechtlichen Schutz unter § 4 Abs. 1 UrhG.⁸ Voraussetzung ist dabei ein gewisser Entscheidungsspielraum für die individuelle Auswahl oder Anordnung, der vorhanden sein muss, sodass auch reine Datensammlungen als Datenbankwerke geschützt sein können.⁹ Diese für den Schutz von Datenbankwerken erforderliche *persönliche geistige Schöpfung* nach § 2 Abs. 2 UrhG bildet den wesentlichen Unterschied zum *sui-generis*-Leistungsschutzrecht von Datenbanken nach § 87a UrhG. Bloßer Aufwand oder eine besondere Sachkenntnis sind für den Datenbankwerkschutz nach § 4 UrhG von keiner Bedeutung.¹⁰ Im Gegensatz dazu erfordert der *sui-generis*-Datenbankschutz nach § 87a UrhG eine nach Art und Umfang wesentliche Investition bei der Beschaffung, Überprüfung oder Darstellung des Datenbankinhalts.¹¹

Wie auch bei Sammelwerken sind die *Auswahl und Anordnung* der Elemente bei schöpferischer Leistung schutzbegründende Bezugspunkte. Auswahl bedeutet dabei das Sammeln und Aufnehmen und Anordnung meint die Einteilung und Präsentation von Elementen.¹² Durch den wahrgenommenen Entscheidungsspielraum bei Auswahl oder Anordnung zeigt sich dann die schöpferische Leistung. Beispielsweise führt die individuelle Auswahl von Gedichtstiteln nach eigenen Kriterien bei einer Sammlung der wichtigsten Gedichte aus der Zeit von 1730-1900 zu einer dementsprechend schöpferischen Eigentümlichkeit.¹³ Ein weiteres Beispiel für eine individuelle Auswahl könnte eine Datenbank sein, die gemeinfreie Romane in mehreren Korpora und in feststehender Anzahl nach eigenen Kriterien zusammenstellt. Diese Kriterien sollen etwa sicherstellen, dass die Auswahl der gesammelten Romane für einen gewissen Zeitraum ausgewogen und durch feststehende Auswahlkriterien balanciert ist. Beispielsweise sollen unterrepräsentierte Sprachen, das Geschlecht der Autorin-

⁷ Zum Schutz von Sammelwerken siehe *Erler-Fridgen*, Teil 2, A. II. in diesem Band.

⁸ Siehe oben; *Ahlberg/Lauber-Rönsberg*, in *Ahlberg/Götting/Lauber-Rönsberg*, BeckOK Urheberrecht, 33. Edition 2022, § 4 Rn. 18.

⁹ EuGH C-604/10, ECLI:EU:C:2012:115, GRUR 2012, 386 Rn. 38 ff. – Football Dataco/Yahoo; BGH GRUR 2007, 685 Rn. 21 – Gedichttitelliste.

¹⁰ *Vobwinkel*, in *Ahlberg/Götting/Lauber-Rönsberg*, BeckOK Urheberrecht, 33. Edition 2022, § 87a Rn. 8; *Vogel*, in *Schricker/Loewenheim*, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 30.

¹¹ *Vogel*, in *Schricker/Loewenheim*, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 30.

¹² *Leistner*, in *Schricker/Loewenheim*, Urheberrecht, 6. Aufl. 2020, § 4 Rn. 23.

¹³ BGH GRUR 2007, 685 Rn. 19 ff. – Gedichttitelliste.

nen und Autoren sowie das Erscheinungsdatum der Werke bei der Auswahl Berücksichtigung finden und hierdurch etwa auch in der Forschung weniger stark beachtete Werke aufgenommen werden.

Ebenso wie für Sammelwerke gilt bei Schutz von Datenbankwerken der Schutz der sogenannten kleinen Münze¹⁴.¹⁵ Der Begriff der „kleinen Münze“ beschreibt geistige Schöpfungen, die gerade noch als urheberrechtlich schutzfähig zu beurteilen sind und an der untersten Schutzuntergrenze liegen.¹⁶ § 4 UrhG greift auf den *Werkbegriff* aus § 2 Abs. 2 UrhG zurück und schützt die „persönliche geistige Leistung“. Dabei entspricht der *Werkbegriff* in § 4 UrhG nicht dem in Art. 3 der Datenbank-RL (96/9/EG)¹⁷, der von „eigener geistigen Leistung“ spricht. Ein sachlicher Unterschied liegt hierin jedoch nicht.¹⁸ Denn einerseits dürfen bei der Schutzbestimmung nach Art. 3 der Datenbank-RL (96/9/EG) keine anderen Kriterien als die Originalität im Sinne einer geistigen Schöpfung herangezogen werden.¹⁹ Und andererseits war bei Datenbanken auch aufgrund der deutschen Rechtsprechung seit jeher bei § 2 Abs. 2 UrhG der Schutz der kleinen Münze gegeben.²⁰ Das bedeutet hier im Ergebnis, dass Datensammlungen dann als Datenbankwerke geschützt sind, wenn ein gewisser Entscheidungsspielraum für eine individuelle Auswahl oder Anordnung vorliegt und dieser ausgenutzt wurde.²¹

Der Schutz als Datenbankwerk kann nicht allein deshalb abgesprochen werden, weil bei Vollständigkeit der Sammlung keine individuelle, eigenschöpferische

¹⁴ Zum Schutz der kleinen Münze unter § 2 Abs. 2 UrhG siehe auch *Erler-Fridgen*, Teil 2, A, I. 1. d. in diesem Band; zur ausführlichen Erläuterung der kleinen Münze siehe *Schulze*, Die kleine Münze und ihre Abgrenzungsproblematik bei den Werkarten des Urheberrechts, S. 1 ff. und *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 293 ff.

¹⁵ *Leistner*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 4 Rn. 50; *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 290; *Ahlberg/Lauber-Rönsberg*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 4 Rn. 25; *Marquardt*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 4 Rn. 5.

¹⁶ Siehe bspw. BGH GRUR 1995, 581, 582 – Silberdistel; zum Begriff der „kleinen Münze“ und dessen Herkunft siehe *Schulze*, Die kleine Münze und ihre Abgrenzungsproblematik bei den Werkarten des Urheberrechts, 1 ff. und *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 293 ff.

¹⁷ Richtlinie 96/9/EG über den rechtlichen Schutz von Datenbanken.

¹⁸ *Leistner*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 4 Rn. 50; zum Streit: *Berger*, Der Schutz elektronischer Datenbanken nach der EG-Richtlinie vom 11.3.1996, GRUR 1997, 169, 170.

¹⁹ Erwägungsgrund 16 der Richtlinie 96/9/EG über den rechtlichen Schutz von Datenbanken; *Ahlberg/Lauber-Rönsberg*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 4 Rn. 25.

²⁰ *Leistner*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 4 Rn. 50.

²¹ EuGH C-604/10, ECLI:EU:C:2012:115, GRUR 2012, 386 Rn. 38 ff. – Football Dacato, siehe *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 290.

Auswahl der Einzelemente vorliegt.²² In diesen Fällen kann sich die Schutzfähigkeit der Datenbank auch aus der Anordnung des Stoffs begründen.²³ Gerade für investitionsintensive, auf Vollständigkeit angelegte Datenbanken kann jedoch die Schutzfähigkeit nach § 4 Abs. 2 UrhG mangels Entscheidungsspielraums abgesprochen werden, wenn diese sich strikt an generelle Ordnungskriterien halten.²⁴ Bei einer Datenbank, die beispielsweise bibliographische Daten aus den zwei existierenden Werken zur entsprechenden Epoche übernimmt und die damit auf Vollständigkeit angelegt ist, könnte demzufolge lediglich eine individuelle Anordnung, etwa eine individuelle Kategorisierung, der Elemente einen Schutz als Datenbankwerk begründen.

Zusätzlich zu den Voraussetzungen an ein Sammelwerk fordert § 4 Abs. 2 UrhG, dass die Elemente systematisch oder methodisch angeordnet und einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sind. Elemente einer Sammlung sind dann systematisch angeordnet, wenn sie sich an einem System, einer Klassifizierung oder einem Ordnungsschema orientieren.²⁵ Methodisch angeordnet sind sie dann, wenn ihre Zusammenstellung ordnenden Handlungsanweisungen oder einem bestimmten Plan folgt.²⁶ Die Anordnung kann beispielsweise alphabetischen, numerischen oder chronologischen Prinzipien folgen.²⁷ Dabei reicht es aus, wenn die Zusammenstellung nach solchen Ordnungsgesichtspunkten erfolgt, die den Zugriff auf die einzelnen Elemente ermöglichen.²⁸ Der EuGH fordert hierzu, dass eine Datenbank eine Methode oder ein System zur Verarbeitung der einzelnen Elemente beinhaltet, mit der bzw. dem die einzelnen Bestandteile der Sammlung wiederzufinden sind.²⁹ Beispielsweise wurde eine Gedichtstitelliste als systematisch angeordnet qualifiziert, die Urheber, Titel, Anfangszeile und Erscheinungsdatum der Gedichte in Gruppen nach Zahl der Nennungen in zugrundeliegenden Sammlungen und in sich geordnet nach den Anfangsbuchstaben der Namen der Dichtenden aufführt.³⁰

²² BGH GRUR 2011, 79 Rn. 38 – Markenheftchen.

²³ BGH GRUR 2011, 79 Rn. 38 – Markenheftchen.

²⁴ BGH GRUR 1999, 923, 924 – Tele-Info-CD; *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 290; so *Marquardt*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 4 Rn. 9.

²⁵ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 4 Rn. 17.

²⁶ *Ahlberg/Lauber-Rönsberg*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 4 Rn. 20.

²⁷ OLG München, GRUR-RR 2001, 228, 229 – Stellenanzeigen.

²⁸ *Czychowski*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 4 Rn. 34.

²⁹ EuGH C-444/02, ECLI:EU:C:2004:697, GRUR 2005, 254 Rn. 30 ff. – Fixtures-Fußballspielpläne II.

³⁰ BGH GRUR 2007, 685 Rn. 17 – Gedichttitelliste.

Eine Datensammlung ist hingegen nicht systematisch oder methodisch geordnet, wenn ein unstrukturierter „Datenhaufen“ vorliegt.³¹ Ein solcher unstrukturierter Datenhaufen beinhaltet zwar Informationen, es fehlt ihm jedoch an einem Mittel zur Verarbeitung der einzelnen Elemente.³² Beispielsweise stellt eine Darstellung von Stellenanzeigen keine Datenbank dar, die kein erkennbares Gliederungsschema aufweist, wenn ihre einzelnen Anzeigen lediglich grob eingeteilt und nach der Größe der Anzeigen geordnet sind.³³

Dass die Elemente einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sein müssen, soll einem zusätzlichen Schutz bereits rechtlich geschützter Einheiten vorbeugen.³⁴ Beispielsweise ist ein Buch als literarisches Werk aus einzelnen Worten geschützt und soll nicht zusätzlich den Schutz als Datenbankwerk erfahren.³⁵ Zugänglichkeit bedeutet, dass auf die Elemente unter Berücksichtigung der Anordnungskriterien zugegriffen werden kann.³⁶ Mit dem Zusatz „mit Hilfe elektronischer Mittel oder auf andere Weise“ wird deutlich gemacht, dass auch nicht-elektronische Datenbanken mitumfasst sind.³⁷ Beispielsweise sind solche Elemente nach § 4 Abs. 2 UrhG auf andere Weise zugänglich, die rein optisch wahrnehmbar sind, etwa gedruckte Bücher oder Zeitungen.³⁸

2. Gegenstand

Das Datenbankwerk wird in seiner Struktur als Sammlung geschützt, wohingegen der Inhalt der Datenbank selbst nicht vom Schutz umfasst ist.³⁹ Der Schutz betrifft daher allein die in der Auswahl oder Anordnung liegende Leistung.⁴⁰ Beispielsweise wurde eine Auflistung der wichtigsten Gedichte aus der Zeit von 1730-1900 aufgrund der individuellen Auswahl von Gedichtstiteln nach eigenen Kriterien als Datenbankwerk charakterisiert.⁴¹ Das Urheberrecht am

³¹ Für § 87a UrhG: OLG München GRUR-RR 2001, 228, 229 – Stellenanzeigen; *Czychowski*, in: Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 4 Rn. 35.

³² OLG München MMR 2007, 525 – Subito.

³³ KG GRUR-RR 2001, 102, 102 – Stellenanzeigen.

³⁴ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 4 Rn. 18.

³⁵ *Czychowski*, in *Fromm/Nordemann*, Urheberrecht, 12. Aufl. 2018, § 4 Rn. 36.

³⁶ *Czychowski*, in *Fromm/Nordemann*, Urheberrecht, 12. Aufl. 2018, § 4 Rn. 36.

³⁷ *Ahlberg*, in *Ahlberg/Götting, Möhring/Nicolini*, Urheberrecht, 4. Aufl. 2018, § 4 Rn. 22.

³⁸ *Ahlberg/Lauber-Rönsberg*, in *Ahlberg/Götting/Lauber-Rönsberg*, BeckOK Urheberrecht, 33. Edition 2022, § 4 Rn. 22.

³⁹ *Leistner*, in *Schricker/Loewenheim*, Urheberrecht, 6. Aufl. 2020, § 4 Rn. 59.

⁴⁰ *Marquardt*, in *Wandtke/Bullinger*, Urheberrecht, 5. Aufl. 2019, § 4 Rn. 18.

⁴¹ BGH GRUR 2007, 685 Rn. 19 ff. – Gedichttitelliste.

Datenbankwerk wird nur dann verletzt, wenn diejenigen Strukturen hinsichtlich der Auswahl oder Anordnung des Stoffs übernommen werden, die die persönliche geistige Schöpfung des Datenbankwerks ausmachen.⁴² Die Übernahme nur einzelner Elemente löst daher regelmäßig keine Verletzung des Urheberrechts am Datenbankwerk aus.⁴³ Beispielsweise kann etwa die Übernahme strukturbegründender Abfrageelemente wie Thesaurus oder Index- und Abfragesysteme eine Verletzung des Datenbankwerkschutzes begründen.⁴⁴ Abfragesysteme, die jedoch üblich oder aus üblichen Zweckmäßigkeitserwägungen hervorgebracht werden, genießen mangels ausreichender Individualität (s.o.) keinen Schutz.⁴⁵ Sind die einzelnen Elemente der Datenbank für sich genommen urheberrechtlich geschützt, so ist zwischen diesem Schutz und dem Schutz des Datenbankwerks zu unterscheiden und beide bestehen nebeneinander.⁴⁶ Beispielsweise ist bei einer Plattform, die digitalisierte Bilder aus Kunst und Kultur bereitstellt, zwischen der möglichen Qualifikation der Plattform als Datenbankwerk⁴⁷ einerseits und dem Schutz der einzelnen Elemente, der Bilder⁴⁸, andererseits zu unterscheiden.

Rechteinhaber bei einem Datenbankwerk ist derjenige, auf dessen persönlichen geistigen Schöpfung die Konzeption beruht, die im Datenbankwerk verkörpert ist.⁴⁹ Der Urheber muss für die Datenbank erforderlichen nicht-schöpferische Leistungen jedoch nicht selbst geleistet haben.⁵⁰ Der wesentliche Unterschied zur Rechteinhaberschaft bei dem Leistungsschutzrecht des Datenbankherstellers (siehe III. 2.) ist, dass die schöpferische Leistung und nicht etwa die Investition der relevante Anknüpfungspunkt für die Rechteinhaberschaft ist.⁵¹ Im Gegensatz dazu ist somit Rechteinhaber des Leistungsschutzrechts nach § 87a Abs.

⁴² BGH GRUR 1992, 382, 384 – Leitsätze; *Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 4 Rn. 59.

⁴³ *Czychowski*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 4 Rn. 40.

⁴⁴ *Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 4 Rn. 59.

⁴⁵ OLG Frankfurt a.M. GRUR-RR 2005, 299, 301 – Online-Stellenmarkt.

⁴⁶ *Marquardt*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 4 Rn. 18.

⁴⁷ Hierzu müsste eine schöpferische Leistung in Auswahl oder Anordnung der Elemente liegen (s.o.).

⁴⁸ Zu denken ist an den Schutz als Lichtbild § 72 UrhG bzw. Lichtbildwerk § 2 Abs.1 Nr. 5 UrhG; zudem können die abgebildeten Objekte ihrerseits urheberrechtlich geschützt sein § 2 Abs. 1 Nr. 4 UrhG.

⁴⁹ *Marquardt*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 4 Rn. 18.

⁵⁰ BGH GRUR 2007, 658 Rn. 23 – Gedichttitelliste.

⁵¹ Ausführlich dazu unten III. 2.

2 UrhG, wer die Investition nach § 87a Abs. 1 UrhG vorgenommen hat und damit die Person, die in den Aufbau einer Datenbank investiert.⁵²

Haben mehrere Personen ein Werk gemeinsam geschaffen, sind sie Miturheber des Werks nach § 8 UrhG, wenn sich ihre Anteile nicht gesondert verwerten lassen. Dabei ist Voraussetzung, dass ein gemeinsames Schaffen der Beteiligten vorliegt, bei dem jeder Beteiligte einen schöpferischen Beitrag leistet, der in das gemeinsame Werk miteinfließt.⁵³ Die schöpferische Mitwirkung kann auch in einem Vorstadium der Werkerstellung erbracht werden, wenn sie als unselbstständiger Beitrag zu einem einheitlichen Schöpfungsprozess der Werkvollendung geleistet wird.⁵⁴ Sowohl vertikale Arbeitsteilung, bei der die Beteiligten nacheinander tätig werden, als auch horizontale Arbeitsteilung, in der gleichzeitig verschiedene Abschnitte eines Werks geschaffen werden, sind für eine Miturheberschaft grundsätzlich geeignet.⁵⁵ Wesentlich ist dabei, dass sich die Beteiligten einer Gesamtidee unterordnen.⁵⁶ Zum Beispiel könnte bei einem Datenbankwerk die gemeinsame Entwicklung von Auswahlkriterien, bei der mehrere Beteiligte einen schöpferischen Beitrag leisten, grundsätzlich zu einer Miturheberschaft führen, wenn die übrigen Voraussetzungen der Miturheberschaft vorliegen. Entscheidend für die Miturheberschaft ist außerdem nach § 8 Abs. 1 UrhG, dass sich die einzelnen Anteile nicht gesondert verwerten lassen. Keine Miturheberschaft liegt regelmäßig vor, wenn die Anteile an einem Werk herauslösbar sind und es theoretisch denkbar ist, dass diese wieder Verwendung finden können, ohne dass diese unvollständig werden.⁵⁷ Es kommt dabei nicht auf einen möglichen kommerziellen Absatz an, sondern auf die selbstständige Verkehrsfähigkeit.⁵⁸ Auch bei einer Beteiligung im Vorstadium der Werkerstellung muss die gesonderte Verwertbarkeit der Werkstufen fehlen.⁵⁹ Schließlich muss der Wille der Beteiligten zur Zusammenarbeit bestehen, im anderen Falle läge in Abgrenzung dazu eine Bearbeitung nach § 3 UrhG vor.⁶⁰ Mit der Werk-

⁵² *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 36; ausführlich dazu unten III. 2.

⁵³ BGH GRUR 1994, 39, 40 – Buchhaltungsprogramm.

⁵⁴ BGH GRUR 1994, 39, 40 – Buchhaltungsprogramm; BGH NJW 1986, 192, 196 – Inkasso-Programm.

⁵⁵ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 8 Rn. 3.

⁵⁶ BGH GRUR 2005, 860, 863 – Fash 2000.

⁵⁷ *Loewenheim/Peifer*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 8 Rn. 5.

⁵⁸ Das bedeutet, dass auch eine Verwertung bei „freier Software“ möglich sein kann: *Loewenheim/Peifer*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 8 Rn. 5.

⁵⁹ *Loewenheim/Peifer*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 8 Rn. 7.

⁶⁰ *Loewenheim/Peifer*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 8 Rn. 7.

schöpfung entsteht kraft Gesetzes eine Verwertungsgemeinschaft, eine Gesamthandsgemeinschaft mit den Rechten und Pflichten aus § 8 Abs. 2 und 4 UrhG.⁶¹ Das bedeutet, dass die Veröffentlichung und Verwertung des Werkes die vorherige Zustimmung aller Miturheber benötigt.⁶²

Bei internationaler Zusammenarbeit entscheidet im Rahmen der Miturheberschaft nicht für jeden Beteiligten das Recht des jeweiligen Heimatlandes, ob und in welcher Form urheberrechtlicher Schutz und eine Miturheberschaft besteht.⁶³ Vielmehr kommt es nach dem Schutzlandprinzip⁶⁴ auf das Recht des Schutzlandes an, in dessen Bereich der Urheberrechtsschutz geltend gemacht wird.⁶⁵ Das bedeutet auch, dass zwar ausländische Urheber grundsätzlich nach Art. 5 Abs. 1 RBÜ⁶⁶ den gleichen Schutz wie Inländer beanspruchen können, dennoch die Miturheberschaft in jedem Schutzland andere Voraussetzungen und Rechtsfolgen haben kann.⁶⁷ Für die Berufung auf das UrhG im Schutzland Deutschland genügt es in personaler Hinsicht nach § 120 Abs. 1 S. 2 UrhG, wenn ein Miturheber deutscher Staatsangehöriger oder nach § 120 Abs. 2 UrhG Deutschen gleichgestellt ist, beispielsweise als Staatsangehöriger eines europäischen Mitgliedsstaates (§ 120 Abs. 2 Nr. 2 UrhG).⁶⁸

Hinsichtlich der Schranken des Datenbankwerkschutzes gelten die allgemeinen Schrankenbestimmungen der §§ 44a ff. UrhG. Berücksichtigt werden müssen die durch die Datenbank-RL (96/9/EG) umgesetzten Sonderregelungen in § 53 Abs. 5 UrhG, wonach die Vervielfältigung zum privaten Gebrauch bei elektronischen Datenbankwerken eingeschränkt ist, und § 55a UrhG, der die normale Benutzung des Datenbankwerkes durch den rechtmäßigen Benutzenden sicherstellen soll.⁶⁹

⁶¹ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 8 Rn. 12.

⁶² *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 8 Rn. 16.

⁶³ *Spindler*, Rechtsfragen der Open Source Software, 65.

⁶⁴ Das Schutzlandprinzip besagt, dass das Recht desjenigen Landes, für dessen Gebiet Rechtsschutz begehrt wird, für die Beurteilung der Verletzung von Urheberrechten und verwandten Schutzrechten maßgeblich ist: *Katzenberger/Metzger*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, Vor § 120 Rn. 121; siehe auch *Raue*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, Vor § 120 Rn. 48.

⁶⁵ *Spindler*, Rechtsfragen der Open Source Software, 65.

⁶⁶ Berner Übereinkommen zum Schutz von Werken der Literatur und Kunst.

⁶⁷ *Spindler*, Rechtsfragen der Open Source Software, 65.

⁶⁸ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 120 Rn. 10.

⁶⁹ *Leistner*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 4 Rn. 46.

III. Leistungsschutzrecht des Datenbankherstellers – Sui-generis-Datenbankschutz

Neben dem Datenbankwerkschutz sollen die Hersteller von Datenbanken durch den *sui-generis*-Datenbankschutz der §§ 87a ff. UrhG in Bezug auf die widerrechtliche Aneignung der Ergebnisse⁷⁰ von Investitionen geschützt werden, die für die Beschaffung und Sammlung des Datenbankinhalts getätigt wurden.⁷¹ Der Zweck dieses Schutzes besteht darin, dass die Person, die den Datenbankaufbau initiiert und das Investitionsrisiko trägt, gegen eine solche unerlaubte Aneignung der Ergebnisse dieser Investition geschützt wird.⁷² Das Recht des Urhebers an einem Datenbankwerk und das Leistungsschutzrecht des Datenbankherstellers nach §§ 87a ff. UrhG bestehen unabhängig voneinander mit verschiedenem Schutzgegenstand.⁷³ Eine Datenbank kann also als Datenbankwerk nach § 4 UrhG geschützt sein,⁷⁴ aber auch zugleich den *sui-generis*-Datenbankschutz nach §§ 87a ff. UrhG erfahren, wenn eine wesentliche Investition in die Erstellung dieser Datenbank und die Zusammenführung deren Inhalts getätigt wurde.⁷⁵ Im Unterschied zum Schutzgegenstand der §§ 87a ff. UrhG sind Datenbankenwerke nach § 4 Abs. 2 UrhG schöpferischer Natur und weisen eine Struktur auf, die aufgrund der Auswahl und Anordnung ihrer Elemente einen individuellen Charakter hat.⁷⁶ Datenbanken nach § 87a UrhG erfordern hingegen eine nach Art und Umfang wesentliche Investition bei der Beschaffung, Überprüfung oder Darstellung ihres Inhalts.⁷⁷ Computerprogramme, mit denen die Datenbank erstellt bzw. verwaltet wird, etwa Zugangs- und Recherchesoftware, sind allein über § 69a ff. UrhG geschützt.⁷⁸

⁷⁰ Der Schutz führt dazu, dass die Gesamtheit oder wesentliche Teile der Datenbank gegen bestimmte Handlungen („Entnahme“ und „Weiterverwendung“) geschützt werden, hierzu ausführlich unten unter 2.

⁷¹ Erwägungsgrund (39) der Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11. März 1996 über den rechtlichen Schutz von Datenbanken.

⁷² EuGH C-762/19, ECLI:EU:C:2021:434, GRUR 2021, 1075 Rn. 22 – CV Online Latvia.

⁷³ BGH GRUR 2007, 658 Rn. 27 – Gedichttitelliste; unter Betonung der steten kumulativen Anwendung der Schutzrechte und klarer Differenzierung der dahinterliegenden Schutzzwecke: *Leistner*, CR 2018, 17, 18.

⁷⁴ Zu den Voraussetzungen des Datenbankschutzes nach § 4 Abs. 2 UrhG siehe oben.

⁷⁵ *Dreier*, in: *Dreier/Schulze, Urheberrechtsgesetz*, 7. Aufl. 2022, Vor § 87a Rn. 1; *Czychowski*, in: *Fromm/Nordemann, Urheberrecht*, 12. Aufl. 2018, § 4 Rn. 49.

⁷⁶ *Vogel*, in: *Schricker/Loewenheim, Urheberrecht*, 6. Aufl. 2020, § 87a Rn. 30.

⁷⁷ *Vogel*, in: *Schricker/Loewenheim, Urheberrecht*, 6. Aufl. 2020, § 87a Rn. 30.

⁷⁸ *Vohwinkel*, in: *Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht*, 33. Edition 2022, § 87a Rn. 10.

1. Voraussetzungen des Leistungsschutzrechts

Eine Datenbank liegt nach § 87a UrhG – in Teilen überschneidend mit den in § 4 Abs. 2 UrhG genannten Voraussetzungen –⁷⁹ dann vor, wenn es sich um eine Sammlung von Werken, Daten oder anderen unabhängigen Elementen handelt, die systematisch oder methodisch angeordnet sind, einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sind und zu deren Beschaffung, Überprüfung oder Darstellung eine nach Art und Umfang wesentliche Investition erforderlich ist.⁸⁰

Elemente einer Datenbank können Texte, Bilder, Zahlen, Fakten, Daten sowie urheberrechtlich geschützte Werke etwa aus dem literarischen, musikalischen oder künstlerischen Bereich sein.⁸¹ Dabei ist nicht relevant, aus welcher Quelle die Elemente stammen oder ob sie selbst erzeugt sind.⁸² Dass diese Elemente einer Datenbank unabhängig sein müssen, bedeutet, dass diese einen eigenständigen Informationsgehalt aufweisen müssen.⁸³ Die Elemente müssen sich voneinander trennen lassen, ohne dass der Wert ihres informativen, literarischen oder musikalischen Inhalts beeinträchtigt wird.⁸⁴ Dies soll urheberrechtlich geschützte Werke wie Texte, Musikstücke oder Bilder und ihre Bestandteile vom Schutz als Datenbanken ausschließen, obwohl deren Bestandteile gegebenenfalls auch einzeln aufrufbar sind.⁸⁵ Außerdem müssen die Elemente der Datenbank systematisch oder methodisch angeordnet sein. Systematik heißt dabei eine Ordnung nach logischen oder sachlichen Kriterien (etwa alphabetisch, numerisch, chronologisch oder thematisch),⁸⁶ Methodik setzt eine planmäßige, auf einen bestimmten Zweck gerichtete Strukturierung voraus.⁸⁷

Hauptmerkmal einer das Leistungsschutzrecht nach §§ 87a ff. UrhG auslösenden Datenbank ist eine nach Art und Umfang wesentliche Investition.⁸⁸ Diese

⁷⁹ Gemeinsame Merkmale sind dabei die Sammlung von Elementen, die systematische oder methodische Anordnung sowie die einzelne Zugänglichkeit der Elemente: *Vogel*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 5.

⁸⁰ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87a Rn. 3.

⁸¹ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 9.

⁸² EuGH C-444/02, ECLI:EU:C:2004:697, GRUR 2005, 254 Rn. 25 – Fixtures Marketing.

⁸³ *Vobwinkel*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 87a Rn. 20.

⁸⁴ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 12.

⁸⁵ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 12.

⁸⁶ *Schmidt/Zech*, CR 2017, 417, 420.

⁸⁷ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87a Rn. 7.

⁸⁸ Art. 7 Datenbank-RL (96/9/EG) spricht von einer in qualitativer oder quantitativer Hinsicht wesentlichen Investition.

Investition muss bei der Beschaffung, Überprüfung oder Darstellung der Datenbankelemente umgesetzt werden. Das bedeutet, dass die Investition beispielsweise beim Auffinden von Daten, ihrer Auswahl, Prüfung, Anordnung und Pflege vorgenommen werden muss.⁸⁹ Zur relevanten Investition zählen demnach jedoch nicht die Mittel, die eingesetzt werden, um den Datenbankinhalt, die Datenbankelemente, zu erzeugen.⁹⁰ Nach der sogenannten Spin-Off-Theorie sollen daneben solche Vorinvestitionen nicht umfasst sein, die eine unabhängige vorherige Leistung für eine Datenbank als reines Nebenprodukt einsetzen.⁹¹

Die Datenbank muss in der Gesamtbetrachtung aller erforderlichen Aufwendungen eine wesentliche Investition erfordern haben.⁹² Es ist also beispielsweise unschädlich, wenn sich die Investition ungleich über die Erstellungsprozesse verteilt. Der Investitionsbegriff ist grundsätzlich weit auszulegen: Erwägungsgrund (40) der Datenbank-Richtlinie (96/9/EG)⁹³ hält fest, dass die Investition keine finanzielle Leistung sein muss, sondern auch im Aufwenden von Zeit, Arbeit oder Energie liegen kann.⁹⁴ Teile des Investitionsaufwands können auch von Dritten erbracht werden, die im Auftrag des Datenbankherstellers tätig sind, beispielsweise Annotationsleistungen zur Vernetzung von Datenbankelementen (etwa über das sog. Crowdfunding).⁹⁵

Zu berücksichtigen sind in diesem Kontext die Investitionen, die bei der Beschaffung, Überprüfung oder Darstellung der Datenbankelemente erbracht werden. In der Phase der Beschaffung des Datenbankinhalts kann eine solche Investition etwa in der Auffindung und Sichtung der Daten, in den Kosten deren Erwerbs oder des Erwerbs von Nutzungsrechten liegen.⁹⁶ Beispielsweise

⁸⁹ Vogel, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 42.

⁹⁰ EuGH C-203/02, ECLI:EU:C:2004:695, GRUR 2005, 244 Rn. 42 – The British Horseracing Board.

⁹¹ So Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87a Rn. 13; ausführlich dazu Ehmann, K&R 2014, 394, 397; siehe auch Sagstetter, Digitaler Strukturwandel und Privatrecht, in Strukturwandel und Privatrecht, 275 mwN; der EuGH dazu bspw. in EuGH C-338/02, ECLI:EU:C:2004:696, GRUR 2005, 252 Rn. 29 – Fixtures Fußballspielpläne I wird diskutiert; zur Auslegung des EuGH siehe auch ausführlich: Hermes, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 41.

⁹² Vogel, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 48.

⁹³ Datenbank-RL: Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11. März 1996 über den rechtlichen Schutz von Datenbanken.

⁹⁴ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87a Rn. 12.

⁹⁵ Siehe etwa zu solchen Vorarbeiten: BGH GRUR 1999, 923, 925 f. – Tele-Info-CD.

⁹⁶ Vogel, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 44.

können Lizenzkosten für die Erschließung von Werken eine zu beachtende Investition darstellen.⁹⁷ Investitionen in die Datenerzeugung sind jedoch nicht erfasst.⁹⁸ Auch Aufwendungen für die Darstellung des Inhalts einer Datenbank können relevant sein wie etwa solche, die in die Anordnung der Datenbankelemente fließen.⁹⁹ Ebenso beispielsweise Personalkosten und sonstige Kosten bei der Aktualisierung und Pflege des Datenbankinhalts werden umfasst.¹⁰⁰ Weiterhin können Kosten der fortlaufenden Pflege der Datenbank und der Mittel erfasst sein, mit denen die Datenbank ihre Funktion¹⁰¹ der Informationsverarbeitung verliehen werden soll.¹⁰² Ein Beispiel für relevante Aktualisierungskosten in den Digital Humanities könnten die Personalkosten sein, die durch die Anpassung von Datenbankinhalt an weiterentwickelte Modellierungsstandards entstehen. Werden beispielsweise die sogenannten TEI-Standards der Text Encoding Initiative¹⁰³ verändert bzw. weiterentwickelt, so müssen etwa Texte in Datenbanken an diese Standards angepasst werden und unter dem Einsatz von Personalkosten in die neue Version dieses Standards konvertiert werden. Durch diese Aktualisierung des Datenbankinhalts wird die weitere Nutzbarkeit der Texte in der Datenbank sichergestellt.

Das Merkmal der Wesentlichkeit der Investition ist ein unbestimmter Rechtsbegriff, den die Datenbank-RL (96/9/EG) nicht näher konkretisiert¹⁰⁴ und dessen Ausfüllung der Umsetzungsgesetzgeber der Rechtsprechung überlassen hat.¹⁰⁵ Als flexibles Kriterium soll die Wesentlichkeit es den Gerichten ermöglichen, einer Vielzahl verschiedener Datenbanken gerecht zu werden.¹⁰⁶ Ausschlaggebend für diese Auslegung des Begriffes der Wesentlichkeit sind vornehmlich die Beschreibung „in qualitativer oder quantitativer Hinsicht“ aus der

⁹⁷ Siehe auch *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 44.

⁹⁸ EuGH C-203/02, ECLI:EU:C:2004:695, EuZW 2004, 757 Rn. 42 – The British Horseracing Board; EuGH C-46/02, ECLI:EU:C:2004:694, GRUR Int 2005, 244 Rn. 49 – Fixtures Marketing; BGH GRUR 2005, 857, 858 – Hit Bilanz.

⁹⁹ Siehe auch *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 44.

¹⁰⁰ Siehe beispielsweise: LG Köln MMR 2002, 689, 690.

¹⁰¹ EuGH C-444/02, ECLI:EU:C:2004:697, GRUR 2005, 254 Rn. 43 – Fixtures Marketing.

¹⁰² Zu weiteren Beispielen siehe *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 44.

¹⁰³ <https://tei-c.org>, zuletzt abgerufen am 05.04.2022.

¹⁰⁴ Art. 7 Datenbank-RL (96/9/EG) spricht von einer „in qualitativer oder quantitativer Hinsicht wesentlichen Investition“.

¹⁰⁵ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 87a Rn. 11; amtliche Begründung der Umsetzungsgesetzgebung, BT-Drs. 13/7385.

¹⁰⁶ *Hermes*, in *Wandtke/Bullinger*, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 52.

Richtlinie (96/9/EG) in einer Gesamtbetrachtung sowie deren Erwägungsgründe und der Zweck der Vorschriften, der Investitionsschutz.¹⁰⁷ Dabei bezieht sich die quantitative Beurteilung auf bezifferbare Mittel, die eingesetzt werden.¹⁰⁸ In der Regel wird es sich dabei um finanzielle Aufwendungen handeln, beispielsweise um Kaufpreis- oder Lohnzahlungen an Dritte sowie eigenen (fiktiven) Lohnaufwand oder die Investition in einen Mitarbeitendenstab¹⁰⁹ etwa für die Beschaffung der Daten.¹¹⁰ In qualitativer Hinsicht stehen dagegen nicht quantifizierbare Anstrengungen, wie geistiger Aufwand oder Verbrauch von Energie in Rede.¹¹¹ Dabei ist zu beachten, dass eine schöpferische Leistung, die zum urheberrechtlichen Schutz als Datenbankwerk führt (s.o.), nicht automatisch allein zu einer qualitativ wesentlichen Investition führt. Zur Begründung des eigenständigen Leistungsschutzrechtes des Datenbankherstellers ist darüber hinaus immer eine berücksichtigungsfähige wesentliche Investition notwendig, die jedoch auch im Einsatz von Zeit, Arbeit und Energie bestehen kann.¹¹² Unbeachtlich ist in jedem Fall, ob die jeweilige Investition objektiv erforderlich ist oder sie sich rentiert.¹¹³ Ist eine Investition also berücksichtigungsfähig (s.o.), dann kann sie in voller Höhe einbezogen werden.¹¹⁴ Generell ist die Wesentlichkeit ein Standard, der Datenbanken, die ohne entsprechenden wirtschaftlichen Aufwand erstellt worden sind, vom Schutz nach § 87a ff. UrhG ausschließen soll.¹¹⁵

Der Maßstab der Bestimmung der Wesentlichkeit der Investition wird uneinheitlich bewertet: Während teilweise vertreten wird, keinen strengen Maßstab an diese Schutzwelle zu legen,¹¹⁶ wird anderer Ansicht nach vorgebracht, dass

¹⁰⁷ *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 51.

¹⁰⁸ EuGH C-338/02, ECLI:EU:C:2004:696, GRUR 2005, 252 Rn. 28 – Fixtures Fußballspielpläne I.

¹⁰⁹ LG München MMR 2002, 58, 59 – Schlagzeilensammlungen im Internet.

¹¹⁰ *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 58.

¹¹¹ EuGH C-444/02, ECLI:EU:C:2004:697, GRUR 2005, 254 Rn. 44 – Fixtures Fußballspielpläne II.

¹¹² *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 68.

¹¹³ *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 52.

¹¹⁴ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 63.

¹¹⁵ *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 42.

¹¹⁶ *Benecke*, CR 2004, 608, 611; *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 12; *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 54; ähnlich *Berger*, GRUR 1997, 169, 173; AG Rostock MMR 2001, 631, 632 – Linksammlung als Datenbank.

diese Schwelle ernstgenommen und nicht ausgehöhlt werden dürfe.¹¹⁷ Der BGH setzt die Schutzwelle für wesentliche Investitionen nun nicht zu hoch an.¹¹⁸ Nach seiner Rechtsprechung reiche es aus, wenn bei objektiver Betrachtung keine ganz unbedeutenden, von allen leicht zu erbringende Aufwendungen erforderlich waren.¹¹⁹ Dies entspreche auch dem Ziel der Datenbank-Richtlinie (96/9/EG), einen Schutz zu schaffen, der einen Anreiz für die Einrichtung von Datenbanken bietet.¹²⁰ Nicht notwendig für die Wesentlichkeit sei eine Investition von substantiellem Gewicht.¹²¹ Generell gilt, dass je höher die Investitionsleistung einzuordnen, also je mehr Arbeit, Zeit oder Geld aufgewendet worden ist, desto eher ist vom Leistungsschutz der § 87a ff. UrhG auszugehen.¹²² Grundsätzlich gilt, dass eine wesentliche Investition abstrakt nicht bezifferbar ist: In der Rechtsprechung gehen die Zahlen auseinander, da die Art der Datenbank bei der Beurteilung der Wesentlichkeit eine nicht unerhebliche Rolle spielt.¹²³ Bejaht wurde beispielsweise die Wesentlichkeit bei Kosten von 34.900 Euro bei der Erarbeitung einer Gedichtstitelliste „der 1.100 wichtigsten Gedichte der deutschen Literatur zwischen 1730 und 1900“ u.a. für die Auswahl der Gedichte, die Vereinheitlichung der Titel und Anfangszeilen der Gedichte und Ermittlung der jeweiligen Entstehungsdaten.¹²⁴ Bei hohen Kosten (ca. 50 Mio Euro Beschaffungskosten¹²⁵ oder Gesamtkosten von 10 Mio Euro für eine Veranstalterdatenbank¹²⁶) wurde eine wesentliche Investition ohne große Zweifel begründet, auch etwa bei Aktualisierungskosten von 50.000 Euro über einen

¹¹⁷ *Schack* etwa betont, dass der Schutz um der Investition willen hier einen strengeren Maßstab als allein den Ausschluss von „Allerweltsinvestitionen“ bedeutet: *Schack*, MMR 2001, 9, 12; auch für einen strengen Maßstab: *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 745; für eine Investition von substantiellem Gewicht: LG Köln ZUM-RD 2000, 304, 306 – Linksammlung babynet.de; *Milbradt*, CR 2002, 710, 713.

¹¹⁸ So auch: *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87a Rn. 14; *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 54; *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 54.

¹¹⁹ BGH GRUR 2011, 724 Rn. 23 – Zweite Zahnarztmeinung II; so auch: BGH GRUR 2011, 1018 Rn. 30 – Automobil-Onlinebörse.

¹²⁰ BGH GRUR 2011, 724 Rn. 23 – Zweite Zahnarztmeinung II.

¹²¹ BGH GRUR 2011, 724 Rn. 23 – Zweite Zahnarztmeinung II unter Verweis auf OLG Köln ZUM 2007, 548, 550 – Wetterdaten für Luftfahrzeugführer: nur der Ausschluss ganz unbedeutender und objektiv minimaler Aufwendungen, da auch Hersteller kleiner Datenbanken geschützt werden sollen.

¹²² *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 54.

¹²³ *Vogel*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 58.

¹²⁴ BGH GRUR 2007, 685, 686 und Rn. 27 – Gedichttitelliste I; zu weiteren Beispielen siehe *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87a Rn. 15.

¹²⁵ BGH GRUR 1999, 923, 926 – Tele-Info CD.

¹²⁶ KG MMR 2001, 171, 172 – Ticketverkaufssystem.

Zeitraum von 5 Jahren¹²⁷ wurde die Wesentlichkeit bejaht.¹²⁸ Doch auch ein wöchentlicher Arbeitsaufwand von 20 bis 30 Stunden zur Beschaffung und Pflege von Daten kann quantitative Wesentlichkeit begründen.¹²⁹ Eine Grenze der quantitativ wesentlichen Investition nach unten muss von der Rechtsprechung geklärt werden.¹³⁰ Es gibt Fälle, in denen die Schutzwelle relativ niedrig gezogen wurde:¹³¹ Beispielsweise die Kosten für die Beschaffung, Betreuung und Weiterentwicklung einer Datenbankssoftware von 3.500 - 4.000 Euro hat der BGH (jedenfalls in Verbindung mit erheblichem Personalaufwand) bereits als quantitativ wesentlich eingeordnet.¹³² Diese Kosten würden beispielsweise bei dem Personaleinsatz in Forschungsprojekten in etwa den Personalkosten einer studentischen Hilfskraft für ein Jahr entsprechen.¹³³ Verneint wurde die Wesentlichkeit der Investition hingegen beispielsweise bei der Beschaffung öffentlich leicht zugänglicher Informationen, etwa aus den Gelben Seiten oder Telefonbüchern, unter Hinweis auf die zu beachtende Gemeinfreiheit von Informationen.¹³⁴

2. Gegenstand und Schutzzumfang des Leistungsschutzrechts

Schutzgegenstand der §§ 87a ff. UrhG ist die Datenbank in Gesamtheit des unter wesentlichem Investitionsaufwand gesammelten Inhalts oder wesentliche Teile ihres Inhalts, nicht der Inhalt der Datenbank selbst.¹³⁵

Rechteinhaber des Leistungsschutzrechts ist nach § 87a Abs.2 UrhG, wer die Investition nach § 87a Abs. 1 UrhG vorgenommen hat und damit die Person, die in den Aufbau einer Datenbank, also die Beschaffung, Überprüfung und Darstellung ihres Inhalts investiert.¹³⁶ Hersteller einer Datenbank ist nach Erwägungsgrund (41) der Datenbank-RL (96/9/EG) insbesondere die Person, die

¹²⁷ Hier handelte es sich bei einzelnen topografischen Landkarten jeweils um eine analoge Datenbank: LG München I GRUR 2006, 225, 227 – Topografische Kartenblätter.

¹²⁸ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 64 mwN.

¹²⁹ OLG Köln ZUM-RD 2014, 433, 435 – Photovoltaik-Datenbanken.

¹³⁰ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 65.

¹³¹ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 65 mwN auch zu Fällen, in denen die Wesentlichkeit verneint wurde, siehe etwa auch AG Rostock MMR 2001, 631, 632 – Linksammlung als Datenbank.

¹³² BGH GRUR 2011, 724 Rn. 25 – Zweite Zahnarztmeinung II; dies jedenfalls in Verbindung mit Personalaufwand siehe Rn. 24: *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 65.

¹³³ Für eine studentische Hilfskraft bei 20h/Monat.

¹³⁴ LG Düsseldorf ZUM 2002, 65, 66 – Übernahme einer Datensammlung.

¹³⁵ *Vogel*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 34.

¹³⁶ *Vogel*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87a Rn. 36.

zwei Kriterien erfüllt: jene, die die Initiative ergreift und das Investitionsrisiko trägt.¹³⁷ Das bedeutet, dass rein passive Geldgebende in der Regel keine Hersteller im Sinne des § 87a Abs. 2 UrhG sind.¹³⁸ Umgekehrt sind nach Erwägungsgrund (41 S. 3) der Datenbank-RL (96/9/EG) Auftragnehmende sowie sonstige Personen, die lediglich Hilfsdienste erbringen, keine Hersteller.¹³⁹ Datenbankhersteller ist hiernach insgesamt jede natürliche oder juristische Person, die wesentliche Investitionen vorgenommen hat und das wirtschaftliche und organisatorische Risiko trägt, das in dem Aufbau der Datenbank steckt.¹⁴⁰ Es kommt also beispielsweise darauf an, wer die Finanzierungs-, Beschaffungs-, oder Personalverträge in seinem Namen abschließt und Nutzungs- und Eigentumsrechte am Datenbankinhalt erwirbt.¹⁴¹ Das Kriterium der Investition prägt die Eigenschaft als Datenbankhersteller und erfasst damit regelmäßig nicht die natürlichen Personen, die die Daten sammeln oder ordnen.¹⁴² Da jedoch auch der Einsatz von Zeit, Arbeit und Energie eine relevante Investition darstellen kann (s.o.), sind in Abgrenzung jedenfalls solche persönlichen sichtenden und ordnenden Tätigkeiten ausgenommen, die im Rahmen eines Angestelltenverhältnisses oder Werkvertrags und ohne eigenes Investitionsrisiko vorgenommen werden, da sie entgeltlich getätigt werden.¹⁴³ Auch die Ideen und das Konzept der Datenbank allein begründet keinen leistungsrechtlichen Schutz.¹⁴⁴ Ob die Grundsätze im universitären Umfeld bei wissenschaftlichen Tätigkeiten möglicherweise modifiziert werden müssen, ist noch nicht abschließend geklärt.¹⁴⁵ Jedoch haben EuGH und BGH auch im universitären Umfeld bei wissenschaftlichen Tätigkeiten allein die genannten Kriterien angewandt: Danach haben sie

¹³⁷ Siehe etwa BGH GRUR 2011, 1018 Rn. 32 – Automobil-Onlinebörse; siehe auch *Czychowski*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 87a Rn. 25.

¹³⁸ *Czychowski*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 87a Rn. 25.

¹³⁹ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 131.

¹⁴⁰ KG MMR 2001, 171, 172 – Vervielfältigung von Daten mwN.

¹⁴¹ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 131.

¹⁴² *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 132.

¹⁴³ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a Rn. 132.

¹⁴⁴ *Vohwinkel*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 87a, Rn. 61: Das bedeutet, dass eine solche, möglicherweise schöpferische Leistung zwar zum Datenbankwerkschutz s.o., aber ggf. nicht zum sui generis Leistungsschutz führen kann.

¹⁴⁵ Vgl. etwa *Baumann/Krahn/Lauber-Rönsberg*, Forschungsdaten-Management und Recht, 2021, 56 f. und für die Einräumung von Nutzungsrechten an Werken BGH GRUR 1991, 523, 525 – Grabungsmaterialien.

die Universität Freiburg als Datenbankherstellerin angesehen, weil sie die Kosten¹⁴⁶ für die in einem Forschungsprojekt zusammengestellten, überprüften und dargestellten Inhalte einer Gedichtstitelliste getragen hatte.¹⁴⁷ Wirken bei der Erstellung der Datenbank mehrere Personen zusammen und erfüllen die oben genannten Kriterien – Ergreifen der Initiative sowie Risikotragung bei der Investition – so steht Ihnen das Schutzrecht gemeinsam zu.¹⁴⁸

Das Leistungsschutzrecht des Herstellers der Datenbank entsteht unmittelbar kraft Gesetzes, nach Abschluss der Herstellung der Datenbank.¹⁴⁹ Die Schutzdauer erstreckt sich nach § 87d UrhG auf 15 Jahre nach Veröffentlichung bzw. 15 Jahre nach Herstellung, wenn die Datenbank bis dahin nicht veröffentlicht wurde. Jede weitere wesentliche Investition kann jedoch bei nachfolgenden Änderungen der Datenbank zu einem neuen Leistungsschutzrecht führen.¹⁵⁰

Diesem Datenbankhersteller stehen die in § 87b UrhG aufgezählten Verwertungsrechte zu, nämlich das Vervielfältigungsrecht nach § 16 UrhG, das Verbreitungsrecht in § 17 UrhG sowie das Recht auf öffentliche Wiedergabe (§ 15 Abs. 2, §§ 19 ff. UrhG) und auf öffentliche Zugänglichmachung (§ 19a UrhG). Der in § 87b UrhG umgesetzte Artikel 7 Abs. 1 der Datenbank-RL (96/9/EG)¹⁵¹ spricht hingegen vom Recht des Datenbankherstellers, die Entnahme und/oder Weiterverwendung der Gesamtheit oder eines in qualitativer oder quantitativer Hinsicht wesentlichen Teils des Inhalts der Datenbank zu untersagen. Beispielsweise können durch das Leistungsschutzrecht geschützte Datenbanken, die

¹⁴⁶ Aus welchen Mitteln (Drittmittel oder Haushaltsmittel) die Kosten durch die Universität Freiburg getragen wurden, war nicht Gegenstand der Erwägungen des Urteils BGH GRUR 2007, 685 Rn. 27 – Gedichttitelliste I.

¹⁴⁷ EuGH C-304/07, ECLI:EU:C:2008:552, GRUR 2008, 1077, Rn. 24, 59, 60 – Directmedia/Albert-Ludwigs-Universität Freiburg; ebenso BGH GRUR 2007, 688 Rn. 13 – Gedichttitelliste II. S. auch BGH GRUR 2007, 685 Rn. 27 – Gedichttitelliste I zum Auseinanderfallen von Datenbankherstellereigenschaft der Universität und dem das Forschungsprojekt leitenden Professor als Urheber des Datenbankwerks.

¹⁴⁸ Zum dogmatischen Hintergrund siehe ausführlich: *Czychowski*, in: Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 87a Rn. 26.

¹⁴⁹ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 746.

¹⁵⁰ LG München I, GRUR 2006, 225, 228 – Topografische Kartenblätter.

¹⁵¹ Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11. März 1996 über den rechtlichen Schutz von Datenbanken.

zwar gemeinfreie Werke zusammenstellen, auf dieser Basis aber die Nutzung wesentlicher Teile¹⁵² der Datenbank für die Schutzdauer von 15 Jahren nach Veröffentlichung bzw. Herstellung nach § 87d UrhG (s.o.)¹⁵³ untersagen. Die Auslegung des § 87b UrhG ist vorrangig an der Richtlinienvorschrift des Art. 7 Datenbank-RL (96/9/EG) und dessen Entnahme- und Weiterverwendungsrecht zu orientieren.¹⁵⁴ Im Zweifel ist stets auf die abschließende Regelung in Art. 7 Datenbank-RL (96/9/EG) und deren Erwägungsgründe in Art. 42 ff. zurückzugreifen.¹⁵⁵ Hiernach sind die Begriffe der Entnahme und der Weiterverwendung aufgrund der Formulierung der Richtlinie (96/9/EG) „ungeachtet der dafür verwendeten Mittel und der Form der Entnahme“ und „jede Form öffentlicher Verfügbarmachung“ weit auszulegen.¹⁵⁶

Beide Begriffe sind nach der Rechtsprechung des EuGH dahingehend auszulegen, dass sie sich auf jede Handlung beziehen, mit der sich jemand ohne Zustimmung des Datenbankherstellers die Ergebnisse dessen Investition aneignet bzw. sie öffentlich verfügbar macht.¹⁵⁷ Zugleich müssen damit Einkünfte entzogen werden, mit denen die Kosten dieser Investition amortisiert werden sollen.¹⁵⁸ Solche Einkünfte können beispielsweise Lizenzeinnahmen sein, die bei konkurrierender Nutzung wesentlicher Teile der Datenbank beeinträchtigt werden könnten. Eine Beeinträchtigung der Investition wurde beispielsweise bei einer spezialisierten Metasuchmaschine angenommen, da sie ein Konkurrenzprodukt darstelle, dass die Gefahr berge, dass entsprechende Einkünfte von Datenbanken entzogen werden.¹⁵⁹ Das Interesse des Datenbankherstellers an der Amortisierung der Investition ist im Fall von Konkurrenzprodukten – etwa bei Suchmaschinen – mit dem Interesse abzuwägen, Zugang zu den Informationen aus den Datenbanken zu erhalten und hieraus innovative Produkte zu erstellen.¹⁶⁰ Dabei ist das Hauptkriterium für die Beeinträchtigung der Investition die Frage

¹⁵² Nicht aber einzelner schutzfreier Elemente.

¹⁵³ Jede weitere wesentliche Investition kann jedoch bei nachfolgenden Änderungen der Datenbank zu einem neuen Leistungsschutzrecht führen: LG München I, GRUR 2006, 225, 228 – Topografische Kartenblätter.

¹⁵⁴ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 2.

¹⁵⁵ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 23.

¹⁵⁶ EuGH C-203/02, ECLI:EU:C:2004:695, GRUR Int 2005, 247 Rn. 51 – The British Horseracing Board.

¹⁵⁷ EuGH C-762/19, ECLI:EU:C:2021:434, GRUR 2021, 1075 Rn. 31 – CV Online Latvia.

¹⁵⁸ EuGH C-762/19, ECLI:EU:C:2021:434, GRUR 2021, 1075 Rn. 31 – CV Online Latvia.

¹⁵⁹ EuGH C-202/12, ECLI:EU:C:2013:850, GRUR Int. 2014, 279 Rn. 37 – Innoweb.

¹⁶⁰ EuGH C-762/19, ECLI:EU:C:2021:434, GRUR 2021, 1075 Rn. 41 f. – CV Online Latvia.

danach, ob die vorliegende Handlung eine Gefahr für die Möglichkeit darstellt, dass die Investition durch den normalen Betrieb der Datenbank amortisiert werden kann.¹⁶¹

Der weit auszulegende Begriff der „Entnahme“ bedeutet nach Art. 7 Abs. 2 lit. a Datenbank-RL (96/9/EG) die ständige oder vorübergehende Übertragung der Gesamtheit oder eines wesentlichen Teils des Inhalts einer Datenbank auf einen anderen Datenträger, ungeachtet der dafür verwendeten Mittel und der Form der Entnahme. Sie umfasst jede unerlaubte Aneignung der Gesamtheit oder eines Teils des Inhalts einer Datenbank.¹⁶² Das bedeutet, dass das Entnahmerecht zur ausschließlichen Verwertung unabhängig von formalen, technischen oder physischen Kriterien,¹⁶³ also online wie offline, befähigt.¹⁶⁴ Eine physische Kopie der übernommenen Daten ist nicht nötig.¹⁶⁵ Auch vorübergehende Übertragung von Elementen, wie das Abspeichern der Daten im Arbeitsspeicher oder auf der Festplatte des Computers, ist nach der Rechtsprechung des BGH vom Begriff der Entnahme umfasst.¹⁶⁶ Beispielsweise hat der EuGH entschieden, dass eine auf Stellenanzeigen spezialisierte Internet-Suchmaschine, die auf ihrem eigenen Server den Inhalt von Webseiten indiziert und als Kopie abspeichert, den Inhalt der auf den Webseiten dargestellten Datenbanken auf einen anderen Datenträger überträgt.¹⁶⁷ Hierdurch werde das Recht auf Entnahme beeinträchtigt.¹⁶⁸

„Weiterverwendung“ ist in Art. 7 Abs. 2 lit. b Datenbank-RL (96/9/EG) definiert als jede Form öffentlicher Verfügbarmachung der Gesamtheit oder eines wesentlichen Teils des Inhalts der Datenbank durch die Verbreitung von Vielfältigungsstücken, durch Vermietung, durch Online-Übermittlung oder

¹⁶¹ EuGH C-762/19, ECLI:EU:C:2021:434, GRUR 2021, 1075 Rn. 44 f. – CV Online Latvia.

¹⁶² EuGH C-203/02, ECLI:EU:C:2004:695, GRUR Int 2005, 247 Rn. 34 – The British Horseracing Board.

¹⁶³ EuGH C-304/07, ECLI:EU:C:2008:552, MMR 2008, 807 Rn. 38 – Directmedia Publishing.

¹⁶⁴ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87b Rn. 2.

¹⁶⁵ EuGH C-304/07, ECLI:EU:C:2008:552, MMR 2008, 807 Rn. 38 – Directmedia Publishing; BGH NJW 2010, 778 Rn. 17 – Gedichttitelliste III.

¹⁶⁶ BGH GRUR 2011, 1018 Rn. 39 – Automobil-Onlinebörse.

¹⁶⁷ EuGH C-762/19, ECLI:EU:C:2021:434, GRUR 2021, 1075 Rn. 36 – CV Online Latvia.

¹⁶⁸ EuGH C-762/19, ECLI:EU:C:2021:434, GRUR 2021, 1075 Rn. 36 – CV Online Latvia; weiter zur Entnahme aus Datenbanken allgemein siehe III.

durch andere Formen der Übermittlung.¹⁶⁹ Konkret interpretiert der EuGH jede Handlung als Weiterverwendung, die darin besteht, die Ergebnisse seiner Investition ohne die Zustimmung des Datenbankherstellers öffentlich verfügbar zu machen und damit die Einkünfte zu entziehen, die es ihm ermöglichen sollen, die Kosten seiner Investition zu amortisieren.¹⁷⁰ Damit ist jede nicht erlaubte Handlung mitumfasst, mit der der Inhalt einer geschützten Datenbank oder wesentlicher Teile dieser in der Öffentlichkeit verbreitet wird.¹⁷¹ Der EuGH hat entsprechend etwa festgestellt, dass eine Internet-Suchmaschine, die es ermöglicht, im Internet frei zugängliche Stellenanzeigen-Datenbanken zu durchsuchen und ihren Nutzenden Zugang zu deren gesamten Inhalt auf einem anderen Weg¹⁷² als den von ihren Herstellern vorgesehenen gewährt, eine solche Zurverfügungstellung der Daten an die Öffentlichkeit darstellt.¹⁷³ Reine Verlinkungen auf Datenbanken, die mit Zustimmung des Rechteinhabers frei verfügbar sind, werden hingegen hiervon in der Regel nicht erfasst.¹⁷⁴

Diese, vom Leistungsschutzrecht umfassten Verwertungsrechte sind jedoch beschränkt auf die Nutzung der Datenbank insgesamt oder eines nach Art und Umfang wesentlichen Teils der Datenbank.¹⁷⁵ Die Bezugnahme der Entnahme und Weiterverwendung auf die Datenbank insgesamt oder wesentlicher ihrer Teile ist damit Anwendungsvoraussetzung des sui-generis-Leistungsschutzrechts.¹⁷⁶

Das sui-generis-Datenbankrecht verweist nicht auf die allgemeinen Schrankenregelungen der §§ 44a ff. UrhG, sondern verfügt über eine eigene, abschließende

¹⁶⁹ Art. 7 Abs. 2 lit. b S. 2 Datenbank-RL (96/9/EG): Mit dem Erstverkauf eines Vervielfältigungsstücks einer Datenbank in der Gemeinschaft durch den Rechtsinhaber oder mit seiner Zustimmung erschöpft sich in der Gemeinschaft das Recht, den Weiterverkauf dieses Vervielfältigungsstücks zu kontrollieren.

¹⁷⁰ EuGH C-202/12, ECLI:EU:C:2013:850, GRUR Int. 2014, 279 Rn. 37 – Innoweb.

¹⁷¹ EuGH C-202/12, ECLI:EU:C:2013:850, MMR 2014, 185 Rn. 37 – Innoweb.

¹⁷² Durch gleichzeitige Recherchemöglichkeit in verschiedenen Datenbanken nach vom Nutzenden wählbaren Kriterien, siehe EuGH C-762/19, ECLI:EU:C:2021:434, GRUR 2021, 1075 Rn. 34 – CV Online Latvia.

¹⁷³ EuGH C-762/19, ECLI:EU:C:2021:434, GRUR 2021, 1075 Rn. 25 ff. – CV Online Latvia; weiter zur Entnahme aus Datenbanken allgemein siehe III.

¹⁷⁴ *Vogel*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87b Rn. 55; zur Differenzierung nach der Art der Linksetzung im Falle der öffentlichen Zugänglichmachung und mwN. siehe v. *Ungern-Sternberg*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 19a Rn. 91 ff.

¹⁷⁵ *Vogel*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87b Rn. 1.

¹⁷⁶ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 25.

Schrankenregelung in § 87c UrhG.¹⁷⁷ Nach § 87c UrhG sind demnach Vielfältigungen eines nach Art und Umfang wesentlichen Teils einer Datenbank für den privaten Gebrauch bei analogen Datenbanken (Nr. 1), zu Zwecken der wissenschaftlichen Forschung gemäß §§ 60c (Nr. 2), der Veranschaulichung des Unterrichts und der Lehre nach §§ 60a und 60b (Nr. 3), des Text und Data Mining gemäß § 44b (Nr. 4) sowie des Text und Data Mining für Zwecke der wissenschaftlichen Forschung nach § 60d (Nr. 5) und schließlich der Erhaltung einer Datenbank nach § 60e Abs. 1 und 6 und § 60f Abs. 1 und 3 (Nr. 6) zulässig. Die Schranke umfasst jedoch generell nicht die Datenbank als Ganzes, sondern immer nur wesentliche Teile der Datenbank, die in den genannten Fällen vielfältigt werden dürfen.¹⁷⁸

IV. Entnahme oder Weiterverwendung bzw. Verwertung nach § 87b UrhG von Material aus Datenbanken

Wie beschrieben beschränkt sich der *sui-generis*-Schutz der Datenbank bei der Entnahme oder Weiterverwendung¹⁷⁹ bzw. Verwertung nach § 87b UrhG von Material auf die Datenbank insgesamt oder wesentliche Datenbankteile. Generell bleibt daher die Nutzung unwesentlicher Teile einer Datenbank frei.¹⁸⁰ Werden unwesentliche Teile einer Datenbank aber wiederholt und systematisch genutzt und läuft dies der normalen Auswertung einer Datenbank zuwider oder beeinträchtigt die berechtigten Interessen des Datenbankherstellers unzumutbar, so wird dies nach § 87b Abs. 1 S. 2 UrhG auch dem Leitungsschutzrecht unterstellt (Umgehungsschutz). Vor diesem Hintergrund sind jedoch vertragliche Vereinbarungen, die über diesen Umgehungsschutz hinaus die Nutzung unwesentlicher Teile einer Datenbank einschränken, nach § 87e UrhG unwirksam.¹⁸¹

Das bedeutet, dass das Schutzrecht bei der Entnahme von Material erst dann greift, wenn diese Wesentlichkeitsschwelle überschritten wird. Auch die Auslegung dieses Wesentlichkeitsbegriffs wurde – mangels Hilfestellung durch den Richtliniengeber und deutschen Umsetzungsgesetzgeber – der Rechtsprechung überlassen.¹⁸² Maßgeblich für die Auslegung ist dabei der Wortlaut der

¹⁷⁷ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87c Rn. 1.

¹⁷⁸ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87c Rn. 3.

¹⁷⁹ Näher zu den Begriffen der Entnahme und Weiterverwendung siehe oben III. 2.

¹⁸⁰ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87b Rn. 5.

¹⁸¹ Hermes, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 4.

¹⁸² Czychowski, in: Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 87b Rn. 5.

Datenbank-RL (96/9/EG), die in Art. 7 von einem in „qualitativer oder quantitativer Hinsicht wesentlichen“ Teil spricht.

Der EuGH hat für den quantitativ wesentlichen Teil des Inhalts einer Datenbank eine relative Betrachtung festgesetzt. Er hat festgelegt, dass sich der quantitativ wesentliche Anteil ausschließlich auf das aus der Datenbank entnommene Datenvolumen bezieht und im Verhältnis zum Gesamtvolumen des Datenbankinhalts zu beurteilen ist.¹⁸³ Hintergrund dieser relativen Bestimmung ist, dass, wenn bei einer Datenbank, die unter wesentlichen Investitionen geschaffen worden ist, ein quantitativ wesentlicher Teil entnommen werde, ebenso die Investitionen, die sich auf den entnommenen Teil bezieht, proportional auch erheblich seien.¹⁸⁴ Je kleiner also beispielsweise die betrachtete Datenbank, desto kleinere Datenbankvolumina als solche bei großen Datenbanken können bereits wesentlich sein und umgekehrt.¹⁸⁵ Hier zeigt sich, dass die beiden Faktoren – Wesentlichkeit der Investition und des wesentlichen Teils bei der Entnahme – bei der quantitativen Wesentlichkeit miteinander verbunden sind.¹⁸⁶ Das bedeutet aber nicht, dass größere Investitionen in größere Datenbanken mit einem geringeren Schutzniveau belegt sind, denn hier kann die qualitative Wesentlichkeit zum Tragen kommen.¹⁸⁷ Diese berücksichtigt die tatsächlich im entnommenen Datenbankteil verkörperte Investition.¹⁸⁸

Ob der Anteil einer Datenbank als quantitativ wesentlich einzustufen ist, lässt sich nur anhand der Umstände des Einzelfalls beurteilen.¹⁸⁹ Teilweise wird angeführt, dass jedenfalls Anteile einer Datenbank von mindestens 50 % des gesamten Datenvolumens als quantitativ wesentlich einzuordnen sein dürften.¹⁹⁰ Der BGH hat bei der Verwendung von 75 % der Titel aus einer Gedichtstitelliste mit 1.100 Titeln eine Entnahme in quantitativ wesentlicher Weise angenommen.¹⁹¹ Eine Übernahme von 10 % des Gesamtvolumens einer Datenbank er-

¹⁸³ EuGH C-203/02, ECLI:EU:C:2004:695, GRUR 2005, 244 Rn. 70 – The British Horseracing Board; siehe auch BGH GRUR 2010, 1004 Rn. 29 – Autobahnmaut.

¹⁸⁴ EuGH C-203/02, ECLI:EU:C:2004:695, GRUR 2005, 244 Rn. 70 – The British Horseracing Board.

¹⁸⁵ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 12.

¹⁸⁶ *Benecke*, CR 2004, 608, 613.

¹⁸⁷ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 13.

¹⁸⁸ BGH GRUR 2011, 1018 Rn. 50 – Automobil-Onlinebörse.

¹⁸⁹ *Leistner*, GRUR Int. 1999, 819, 832.

¹⁹⁰ *Raue/Bensinger*, MMR 1998, 507, 511; *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 15.

¹⁹¹ BGH GRUR-RR 2010, 232, Rn. 18 – Gedichttitelliste III.

fülle hingegen laut BGH nicht die Anforderungen an einen quantitativ wesentlichen Teil.¹⁹² Beispielsweise wäre daher die Entnahme eines einzelnen Werks aus einer Korpusdatenbank unproblematisch, wohingegen größere Anteile einer Datenbank entsprechend der Rechtsprechung des BGH (s.o.) je nach den Umständen des Einzelfalls als wesentlich eingestuft werden können.

Die Entnahme eines qualitativ wesentlichen Anteils einer Datenbank kann zu einem Schaden für die Investition des Datenbankherstellers führen, auch wenn der Teil quantitativ nur geringfügig ist.¹⁹³ Denn auch ein quantitativ geringfügiger Teil des Inhalts einer Datenbank kann bezogen auf die Beschaffung, Überprüfung oder Darstellung des Inhalts eine erhebliche menschliche, technische oder finanzielle Investition erfordern.¹⁹⁴ Bezugspunkt des qualitativ wesentlichen Teils einer Datenbank ist daher die Bedeutung und der Umfang der Investition, die mit der Beschaffung, Überprüfung oder Darstellung des Inhalts des entnommenen Datenbankteils verbunden war.¹⁹⁵ Diese qualitative Einordnung geschieht unabhängig davon, ob der Teil in quantitativer Hinsicht einen wesentlichen Teil des Inhalts der geschützten Datenbank darstellt.¹⁹⁶ Denn auch ein quantitativ geringfügiger Teil des Inhalts einer Datenbank kann eine ganz erhebliche menschliche, technische oder finanzielle Investition erfordern.¹⁹⁷ Grundsätzlich verbietet sich hier eine schematische Herangehensweise für die Beurteilung der qualitativen Wesentlichkeit.¹⁹⁸ Unberührt bleibt diese Einordnung von dem schöpferischen oder informativen Wert des Datenbankteils, denn dies sind ja gerade keine relevanten Kriterien für den sui-generis-Schutz des Datenbankherstellers.¹⁹⁹ Auch subjektive Verwendungszwecke oder Funktionen sowie der Weiterverwendungswert der in Frage stehenden Datenbankteile sind ohne Bedeutung, es kommt allein auf die in den Datensätzen verkörperte Investition an.²⁰⁰ Beispielsweise war die Tatsache, dass aus einer Pferdesportda-

¹⁹² BGH GRUR 2011, 724 Rn. 29 – Zweite Zahnarztmeinung II.

¹⁹³ BGH GRUR 2011, 724 Rn. 30 – Zweite Zahnarztmeinung II.

¹⁹⁴ BGH GRUR 2011, 724 Rn. 30 – Zweite Zahnarztmeinung II.

¹⁹⁵ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 16.

¹⁹⁶ EuGH C-203/02, ECLI:EU:C:2004:695, GRUR 2005, 244 Rn. 71 – The British Horseracing Board.

¹⁹⁷ EuGH C-203/02, ECLI:EU:C:2004:695, GRUR 2005, 244 Rn. 71 – The British Horseracing Board.

¹⁹⁸ *Vogel*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87b Rn. 32.

¹⁹⁹ *Vogel*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87b Rn. 32.

²⁰⁰ *Vogel*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 87b Rn. 32; *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 17.

tenbank entnommene Daten für die Veranstaltung von Pferderennen unbedingt erforderlich sind, unerheblich für die Beurteilung der Wesentlichkeit.²⁰¹ Kommt es bei einer Datenbank auf Aktualität an, kann der Aktualisierungsaufwand etwa durch Personalkosten eine wesentliche Investition und damit die Aktualisierungen qualitativ wesentliche Teile der Datenbank darstellen.²⁰² Eine qualitative Wesentlichkeit kann jedenfalls dann angenommen werden, wenn ein erheblicher Schaden für die Amortisation der Investition des Datenbankherstellers durch die Entnahme droht, etwa wenn das Rückgrat einer Datenbank übernommen wird.²⁰³ Auch hier zeigt sich die Wechselwirkung zwischen Investition und Schutz der Datenbank bei der Entnahme: Es werden daher insgesamt bei investitionsintensiveren Datenbanken bereits kleinere Datenbankteile die Schwelle der qualitativen Wesentlichkeit überschreiten als bei investitionsärmeren Datenbanken.²⁰⁴

Ein unwesentlicher Teil des Inhalts einer Datenbank ist jeder Teil, der dem Begriff wesentlicher Teil sowohl in quantitativer als auch in qualitativer Hinsicht nicht entspricht.²⁰⁵ Einzelne Datensätze sind jedenfalls keine wesentlichen Teile einer Datenbank.²⁰⁶ Beispielsweise ist die Nutzung von einzelnen kleinen Bestandteilen von Zeitungsartikeln, die in einer Datenbank gespeichert sind, zur Verschaffung eines Überblicks, ob der Abruf der Volltexte der Artikel sinnvoll wäre, keine Entnahme wesentlicher Teile einer Datenbank.²⁰⁷ Etwa auch die Entnahme von Daten eines Renntages aus einer Pferdesportdatenbank einer ganzen Saison stellen keine Entnahme wesentlicher Teile dar.²⁰⁸ Keine unwesentliche Entnahme ist hingegen der Abgleich von zwei Datenbanken zur Erstellung einer Abgleichsliste, da die Daten umfänglich zum Datenabgleich entnommen werden mussten.²⁰⁹ Schließlich ist die hinter einer Datenbank stehende

²⁰¹ EuGH C-203/02, ECLI:EU:C:2004:695, GRUR 2005, 244 Rn. 78 – The British Horseracing Board.

²⁰² BGH GRUR 2009, 852 Rn. 47 – Elektrischer Zolltarif; *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 17.

²⁰³ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 19.

²⁰⁴ *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87b Rn. 19.

²⁰⁵ EuGH C-203/02, ECLI:EU:C:2004:695, GRUR 2005, 244 Rn. 82 – The British Horseracing Board.

²⁰⁶ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 87b Rn. 7.

²⁰⁷ BGH GRUR 2003, 958, 962 – Paperboy..

²⁰⁸ EuGH C-203/02, ECLI:EU:C:2004:695, GRUR 2005, 244 Rn. 92 – The British Horseracing Board.

²⁰⁹ BGH GRUR 2009, 852 Rn. 45 – Elektronischer Zolltarif.

Struktur bzw. das Ordnungssystem kein wesentlicher Bestandteil einer Datenbank unter dem sui-generis-Leistungsschutz.²¹⁰ Werden entnommene Daten andersartig strukturiert, folgt dem aber nicht, dass diese ihre Eigenschaft als wesentlicher Teil einer Datenbank verlieren.²¹¹

V. Nachnutzung von geschützten Datenbanken

Dem Ansatz von Open Science folgend, kann es Forschenden ein Anliegen sein, die Nutzung von erzeugten Datenbanken für die Nachnutzung beispielsweise durch die Allgemeinheit freizugeben. Hier ist zunächst anzumerken, dass sowohl der urheberrechtliche Schutz als Datenbankwerk – mit Entstehung des schutzfähigen Werkes²¹² – als auch das Leistungsschutzrecht des Datenbankherstellers – mit Abschluss der Herstellung der Datenbank – unmittelbar kraft Gesetzes entstehen.²¹³ Zum Umfang des Schutzes als Datenbankwerk bzw. durch das Leistungsschutzrecht des Datenbankherstellers siehe bereits oben II. 2. und III. 2. Für den Schutz der Datenbank gilt dabei grundsätzlich: Im Gegensatz zum entstandenen Urheberrecht²¹⁴ und damit verbundenen Verwertungsrechten am Datenbankwerk ist das Leistungsschutzrecht des Datenbankherstellers und die entsprechenden Rechte nach § 87b UrhG frei übertragbar.²¹⁵ Soll die Nachnutzung von Datenbanken ermöglicht werden, so sind zum einen für die Nutzenden die gesetzlichen Schranken – für Datenbankwerke §§ 44a ff. UrhG und für das Leistungsschutzrecht die Schrankenregelung in § 87c UrhG (s.o.) – in Betracht zu nehmen. Zum anderen können die Rechteinhaber der Nutzung zustimmen.²¹⁶

Rein schuldrechtliche Vereinbarungen über die Nutzung sind zwischen den Parteien möglich, wirken dann jedoch nur inter partes, also zwischen den Vertragsparteien und nicht gegenüber Dritten.²¹⁷ Sowohl bei Datenbankwerken als

²¹⁰ OLG Frankfurt a. M. MMR 2003, 45, 48 – IMS Health, *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 87b Rn. 7.

²¹¹ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 87b Rn. 7.

²¹² *Loewenheim/Leistner*, in *Schricker/Loewenheim*, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 92.

²¹³ Für das Leistungsschutzrecht siehe *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 746.

²¹⁴ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 593.

²¹⁵ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 658; *Czychowski*, in: *Fromm/Nordemann*, Urheberrecht, 12. Aufl. 2018, § 87b Rn. 34.

²¹⁶ *Euler*, RuZ 2020, 56, 70 unter einer CC BY-SA Lizenz; zur schlichten Einwilligung siehe BGH GRUR 2010, 628 – Vorschaubilder I.

²¹⁷ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 604.

auch im Falle des Leistungsschutzrechts ist es jedoch auch möglich, Nutzungsrechte nach § 31 Abs. 1 UrhG mit (quasi-)dinglicher Wirkung einzuräumen.²¹⁸ Forschende können beispielsweise mit Hilfe von sogenannten Open-Content-Lizenzen²¹⁹ jedermann einfache Nutzungsrechte einräumen.²²⁰ Open-Content Lizenzen, wie beispielsweise die bekannten Lizenzen der Organisation Creative Commons²²¹, können eine kontrollierte Nutzungsfreigabe an die Allgemeinheit ermöglichen.²²² Seit Einführung der Version 4.0 der Creative Commons Lizenzen sind diese neben urheberrechtlich geschützten Werken auch ausdrücklich auf das Leistungsschutzrecht des Datenbankherstellers anwendbar.²²³ Wichtig ist, dass eine Lizenzierung unter den Creative Commons Lizenzen nur möglich ist, wenn und soweit die betroffenen Nutzungsrechte noch nicht anderweitig eingeräumt wurden.²²⁴ Insbesondere ist eine Lizenzierung dann ausgeschlossen, wenn ausschließliche Nutzungsrechte bereits anderweitig eingeräumt wurden.²²⁵ Zu beachten ist außerdem, dass der Datenbankinhalt und die Datenbank als solche regelmäßig einen unterschiedlichen Schutz erfahren können.²²⁶ Auch bei der Einräumung von Nutzungsrechten sind demzufolge zwischen Datenbankinhalt und dem Schutz der Datenbank als solcher (wie oben beschrieben) zu unterscheiden.

VI. Ergebnis

Sammlungen von Werken oder Daten können als Datenbankwerk nach § 4 Abs. 2 UrhG urheberrechtlich geschützt sein, wenn in Auswahl oder Anordnung der Datenbankelemente eine schöpferische Leistung liegt. Außerdem

²¹⁸ *Obly*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 31 Rn. 6 ff.; *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 31 Rn. 7; zum dogmatischen Hintergrund: *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 593 ff.

²¹⁹ Zum rechtsdogmatischen Hintergrund von Open-Content-Lizenzen siehe *Jaeger/Metzger*, MMR 2003, 431.

²²⁰ *Euler*, RuZ 2020, 56, 72 unter einer CC BY-SA Lizenz.

²²¹ Zum Inhalt der Creative Commons Lizenzen 4.0 siehe *Paul*, in Hoeren/Sieber/Holznel, Handbuch Multimediarecht, 57. EL 2021, Teil 7.4 Rn. 124 ff.

²²² *Paul*, in Hoeren/Sieber/Holznel, Handbuch Multimediarecht, 57. EL 2021, Teil 7.4 Rn. 121.

²²³ *Paul*, in Hoeren/Sieber/Holznel, Handbuch Multimediarecht, 57. EL 2021, Teil 7.4 Rn. 123; siehe <https://creativecommons.org/Version4/>, zuletzt abgerufen am 13.04.2022.

²²⁴ *Paul*, in Hoeren/Sieber/Holznel, Handbuch Multimediarecht, 57. EL 2021, Teil 7.4 Rn. 139.

²²⁵ *Paul*, in Hoeren/Sieber/Holznel, Handbuch Multimediarecht, 57. EL 2021, Teil 7.4 Rn. 139.

²²⁶ Siehe oben II. 2. und III. 2.

müssen für den Schutz nach § 4 Abs. 2 UrhG die Datenbankelemente systematisch oder methodisch angeordnet und einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sein. Wer urheberrechtlich relevante Nutzungshandlungen bei Datenbankwerken vornehmen möchte, muss die Zustimmung des Rechtsinhabers einholen oder auf die urheberrechtlichen Schranken der § 44a ff UrhG zurückgreifen.

Unabhängig von diesem Schutz kann eine Datenbank den *sui-generis*-Datenbankschutz nach §§ 87 a ff. UrhG mit Schutzdauer von 15 Jahren nach § 87d UrhG erfahren, also dem Leistungsschutzrecht des Datenbankherstellers unterliegen. Hierzu bedarf es – im Gegensatz zur schöpferischen Leistung bei § 4 Abs. 2 UrhG – einer nach Art und Umfang wesentlichen Investition bei der Beschaffung, Überprüfung oder Darstellung der Datenbankelemente.

Die Entnahme oder Weiterverwendung bzw. Verwertung nach § 87b UrhG von Elementen aus einer Datenbank kann einen Eingriff in das Leistungsschutzrecht des Datenbankherstellers nach §§ 87a ff. UrhG begründen. Hier gilt: Die Übernahme des Datenbankinhalts insgesamt oder wesentlicher Teile des Datenbankinhalts greift in den Leitungsschutz nach § 87b UrhG ein. Werden unwesentliche Teile einer Datenbank aber wiederholt und systematisch genutzt, kann dies in bestimmten Fällen nach § 87b Abs. 1 S. 2 UrhG auch dem Leitungsschutzrecht unterstellt sein (Umgehungsschutz).

Wer wesentliche Teile einer Datenbank entnehmen oder weiterverwenden möchte, muss die Zustimmung des Rechtsinhabers einholen oder insofern auf die gesonderte Schrankenregelung in § 87c UrhG zurückgreifen.

Leseempfehlungen zur vertiefenden Lektüre: *Leistner*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 4, *Schmidt/Zech*, Datenbankherstellerschutz für Rohdaten?, Computer und Recht (CR) 2017, 417, *Hermes*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87a und b.

E. Die internationale Anwendbarkeit des (deutschen) Urheberrechts

Karolina Benedyk

Version 1.0 (05.01.2024), CC BY-SA 4.0.

Urheberrechtlich geschützte Gegenstände¹ sind immateriell. Sie können damit auf der Welt gleichzeitig genutzt werden. Daraus leitet sich bei grenzüberschreitenden Sachverhalten die Frage ab, welches Recht anzuwenden ist. Aus diesen Gründen erläutert die Handreichung im Folgenden in Grundzügen, welches Urheberrecht auf grenzüberschreitendes Forschen, Publizieren und Präsentieren anwendbar ist.

I. Einleitung

1. Spezifische Eigenschaften von Urheberrechten

Immaterielle Güter haben keinen Belegenheitsort.² Mit Belegenheitsort ist der Ort gemeint, an dem sich eine Sache aktuell befindet. Immaterielle Güter sind dagegen keine Sachen. Es handelt sich Gedanken und Ideen, die geschützt werden. Aufgrund ihrer ubiquitären Art können Werke und Leistungsschutzrechte auf der ganzen Welt gleichzeitig genutzt werden.³ Zudem gibt es kein einheitliches internationales Urheberrecht. Es besteht aus einem Bündel nationaler, territorial begrenzter Urheberrechte.⁴ Die jeweiligen nationalen Urheberrechte beruhen auf (weitgehend im Detail) unterschiedlichen Vorstellungen über die Voraussetzungen und die Ausgestaltung des Urheberschutzes. Das führt zu variierenden Schutzvoraussetzung und einem voneinander abweichenden Schutzzumfang.

¹ Zu den Kriterien der urheberrechtlichen Schutzfähigkeit siehe *Erler-Fridgen*, Teil 2, A. in diesem Band.

² *Raue*, in: Dreier/Schulze, *Urheberrechtsgesetz*, 7. Auflage 2022, Vor § 120 Rn. 3.

³ *Raue*, in: Dreier/Schulze, *Urheberrechtsgesetz*, 7. Auflage 2022, Vor § 120 Rn. 3.

⁴ EuGH C-173/11, ECLI:EU:C:2012:642 Rn. 27 – *Football Dataco/Sportradar*; BGH GRUR 2007, 691 Rn. 18 – *Staatsgeschenk*.

Viele Länder wenden auf das Urheberrecht das Recht des jeweiligen Staates an, für dessen Hoheitsgebiet der Schutz beansprucht wird (*lex loci protectionis*). Nach dem Territorialitätsprinzip beschränkt sich die räumliche Wirkung des Urheberrechts auf das Territorium des Staates, der sie zugunsten des Rechtsinhabers im Regelfall *ipso iure* anerkannt hat.⁵ Demnach erzeugt jede nationale urheberrechtliche Regelung im Grundsatz lediglich Wirkung auf dem eigenen Staatsgebiet.⁶ Diese Zusammensetzung national gewährter, territorial begrenzter Urheberrechte, die den grenzüberschreitenden Schutz ihrer Urheber bewirken, wird als „Bündel“, „Urheberrechts-Mosaik“ oder auch „Flickenteppich“ bezeichnet.⁷

2. Grenzüberschreitende Sachverhalte

Die internationale Anwendbarkeit des (deutschen) Urheberrechts ist nur bei grenzüberschreitenden Sachverhalten fraglich. Deutsche Urheber berufen sich regelmäßig auf deutsches Urheberrecht. Wenn Urheber in Deutschland eine Verletzungshandlung rügen wollen, die in Deutschland stattgefunden hat, ist unproblematisch deutsches Urheberrecht an einem deutschen Gericht anzuwenden. Handelt es sich um eine Verletzungshandlung im Ausland, die sich jedoch (nur) in Deutschland auswirkt, ist wiederum (jedenfalls auch) deutsches Urheberrecht anwendbar.

Allerdings können sich urheberrechtliche Verletzungshandlungen auf viele Staaten gleichzeitig auswirken. Das wird durch den grenzüberschreitenden Internetauftritt zusätzlich verstärkt.⁸ Die Frage, welches Recht bei grenzüberschreitenden Sachverhalten anzuwenden ist, ist im internationalen Privatrecht angesiedelt (vgl. Art. 3 EGBGB). Sind Verletzungshandlungen auf einer Webseite auffindbar, die überall zugänglich ist, ist zu klären, das Recht welchen Staates anzuwenden ist.⁹ Problematisch ist hierbei insbesondere, den Eingriffsort zu lokalisieren.¹⁰ Eingriffsort ist der Ort, an dem die Verletzung der Rechtsposition

⁵ *Lauber-Rönsberg*, in: Möhring/Nicolini, Urheberrecht, 4. Auflage 2018, Kollisionsrecht Rn. 4.

⁶ *Dreier/Schulze*, Urheberrechtsgesetz, 7. Auflage 2022, Einl. Rn. 42.

⁷ *Lauber-Rönsberg*, in: Möhring/Nicolini, Urheberrecht, 4. Auflage 2018, Kollisionsrecht Rn. 4.

⁸ Vgl. hierzu *Geller*, GRUR Int. 2000, 659.

⁹ v. *Welser*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, Vor §§ 120 Rn. 15.

¹⁰ v. *Welser*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, Vor §§ 120 Rn. 15.

des Betroffenen erfolgt ist. Der Handlungs- sowie der Erfolgsort können nämlich auseinanderfallen. Ohne Einschränkungen wäre dabei jede Rechtsordnung anwendbar.

3. Prüfung im Vierschritt

Hierbei erfolgt die Prüfung sinnvollerweise in einem Vierschritt.¹¹ Zunächst ist die internationale Zuständigkeit der Gerichte zu klären. Urheber bestimmen im Grundsatz, welches Gericht sie anrufen wollen. Die Regeln über die internationale Zuständigkeit bestimmen, welches Gericht bei grenzüberschreitenden Sachverhalten anzurufen ist, wenn die Gerichte mehrerer Staaten dafür in Frage kommen. Die Gerichtszuständigkeit hat Auswirkungen auf das Kollisionsrecht, weil jedes Gericht das internationale Privatrecht anwendet, das am Gerichtsort gilt (lex fori-Prinzip).

Ist ein deutsches Gericht zuständig, ist im zweiten Schritt deutsches (bzw. europäisches) Kollisionsrecht anwendbar. Hierbei bestimmen die Regeln des internationalen Privatrechts das anzuwendende Recht bei Sachverhalten mit einer Verbindung zu einem ausländischen Staat, vgl. Art. 3 EGBGB. Sie regeln also, welche Rechtsordnung für den Schutz maßgeblich ist.¹²

Verweist das internationale Privatrecht auf das deutsche Recht, ist im dritten Schritt das Fremdenrecht zu beachten, vgl. §§ 120 ff. UrhG. Das Fremdenrecht regelt, ob Ausländer und Ausländerinnen für ihre Werke und sonstigen Schutzgegenstände in Deutschland Rechtsschutz genießen.¹³

Im vierten und letzten Schritt prüfen Gerichte, ob durch die grenzüberschreitende Handlung materielles deutsches Urheberrecht verletzt ist. Hierbei ist das Auswirkungs- und Handlungsprinzip heranzuziehen.

II. Internationale Zuständigkeit

Das internationale Zuständigkeitsrecht regelt, welches Gericht welchen Staates bei Sachverhalten mit Auslandsbezug zuständig sind.

¹¹ Vgl. *Raue*, in Dreier/Schulze, Vor § 120 Rn. 1.

¹² v. *Welser*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, Vor §§ 120 Rn. 2.

¹³ v. *Welser*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, Vor §§ 120 Rn. 2.

1. Beklagtengerichtsstand

Die internationale Zuständigkeit deutscher Gerichte richtet sich bei Beklagten mit Wohn- oder Geschäftssitz in der EU nach der EuGVO¹⁴, Art. 4 I EuGVO. Sie sind grundsätzlich vor den Gerichten in ihrem Wohnsitzmitgliedstaat zu verklagen (Beklagtenwohnsitz). Dabei ist die Staatsangehörigkeit unerheblich. Die EuGVO regelt alle Zivil- und Handelssachen (Art. 1 I EuGVO) und ist somit auch auf das Urheberrecht anzuwenden.

2. Zusätzlicher Gerichtsstand bei Urheberrechtsverletzungen in der EU

Urheberrechtsverletzungen sind unerlaubte Handlungen, Art. 7 Nr. 2 EuGVO.¹⁵ Dadurch gelten ergänzende Regelungen zum Beklagtenwohnsitz. Es können die Gerichte des Handlungs- oder Erfolgsort angerufen werden. Der Handlungsort richtet sich danach, wo der Beklagte das ursächliche Geschehen vorgenommen hat. Dabei handelt es sich um das Ereignis, auf dem die behauptete Urheberrechtsverletzung beruht.¹⁶ Die Suche nach dem ursächlichen Geschehen gestaltet sich bei Internetsachverhalten schwierig. Dabei kann auf die Entscheidungen des EuGH zurückgegriffen werden. Beruht die Urheberrechtsverletzung darauf, dass Webseitenbetreiber geschützte Werke zugänglich machen, sieht der EuGH das maßgebliche Verhalten im Auslösen des technischen Vorgangs, der zum Erscheinen der geschützten Inhalte auf der Webseite führte.¹⁷ Der Serverstandort ist nicht maßgeblich, da er zufällig und schwer zu erkennen ist.¹⁸ Vielmehr ist der Ort entscheidend, an dem die Verletzer über die Veröffentlichung des Angebots auf der Webseite entschieden haben.¹⁹ Bei mehraktigen Delikten oder gestreckten Handlungsverläufen bedarf es einer Schwerpunktsetzung.²⁰ Ist die Urheberrechtsverletzung einem Unternehmen zuzurechnen, verortet der EuGH die Handlung der natürlichen Person am Sitz des Unternehmens.²¹

¹⁴ VO (EU) Nr. 1215/2012 vom 12. Dezember 2012 über die gerichtliche Zuständigkeit und die Anerkennung und Vollstreckung von Entscheidungen in Zivil- und Handelssachen.

¹⁵ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 24.

¹⁶ EuGH C-441/13, ECLI:EU:C:2015:28 Rn. 23 – *Hejduk*.

¹⁷ EuGH C-441/13, ECLI:EU:C:2015:28 Rn. 24 – *Hejduk*.

¹⁸ EuGH C-523/10, ECLI:EU:C:2012:220 Rn. 36 – *Wintersteiger*.

¹⁹ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 28.

²⁰ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 28.

²¹ EuGH C-441/13, ECLI:EU:C:2015:28 Rn. 25 – *Hejduk*; *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 28.

Der Kläger kann auch an dem Ort klagen, an dem das schädigende Ereignis eingetreten ist oder einzutreten droht (Erfolgsort).²² Hierfür ist ausreichend, dass die Webseite mit dem urheberrechtlich geschützten Gegenstand in dem jeweiligen Mitgliedstaat der EU abrufbar und das geltend gemachte Urheberrecht dort geschützt ist.²³ Forschende können mithin nicht nur an ihrem Wohnsitz, sondern auch am Handlungs- oder Erfolgsort verklagt werden. Macht eine Forschungsgruppe urheberrechtlich geschützte Gegenstände über eine Webseite weltweit zugänglich, können die Mitglieder der Gruppe (jedenfalls theoretisch) vor Gerichten in der gesamten EU²⁴ verklagt werden, soweit sie ein dort geschütztes Urheberrecht verletzen.

3. Gerichtsstand bei Urheberrechtsverletzungen außerhalb der EU

Für Beklagte mit Sitz in Island, Norwegen und der Schweiz ist das Lugano-Abkommen (LugÜ) und bei allen Beklagten mit Sitz im übrigen Ausland sind in entsprechender Anwendung die §§ 12ff. ZPO anwendbar, die bei den hier interessierenden Sachverhalten im Wesentlichen dieselben Kriterien anwenden.²⁵

III. Internationales Privatrecht

Das internationale Privatrecht regelt, welches Recht auf Sachverhalte mit Auslandsbezug anwendbar ist. In der Europäischen Union ist das internationale Privatrecht in den hier interessierenden Konstellationen weitgehend durch kollisionsrechtliche Verordnungen vereinheitlicht. Damit gilt das nachfolgend Beschriebene gleichermaßen in allen EU-Staaten.

1. Schutzlandprinzip (Art. 8 I Rom II-VO)

Das dem europäischen Urheberrecht zugrundeliegende Kollisionsprinzip ist das Schutzlandprinzip, Art. 8 I Rom II-VO. Art. 8 I Rom II-VO regelt, dass auf

²² EuGH C-441/13, ECLI:EU:C:2015:28 Rn. 18 – *Hejduk; Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 24, 30.

²³ EuGH C-441/13, ECLI:EU:C:2015:28 Rn. 29, 34 – *Hejduk; Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 32 f.

²⁴ Und möglicherweise auch vor außereuropäischen Gerichten. Das liegt aber außerhalb des Untersuchungsgegenstands dieses Beitrags.

²⁵ BGH GRUR 2020, 647 Rn. 35 f. – *Club Hotel Robinson; Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 41 f.

Verletzungen von Rechten des geistigen Eigentums das Recht des Staates anzuwenden ist, für den der Schutz beansprucht wird. Geistige Eigentümer sind das Urheberrecht, verwandte Schutzrechte sowie das sui-generis Schutzrecht für Datenbanken.²⁶ Nach Art. 1 I Rom II-VO ist der Anwendungsbereich auf alle außervertraglichen Schuldverhältnisse in Zivil- und Handelssachen eröffnet. Ein außervertragliches Schuldverhältnis definiert der EuGH als solches, das nicht aus einem Vertrag oder Ansprüchen aus einem Vertrag erwächst.²⁷ Lizenzvereinbarungen unterfallen also nicht der Rom II-VO, sondern dem Vertragsstatut.²⁸ Die Reichweite der erfassten Rechtsfragen regelt Art. 15 I Rom II-VO.²⁹ Eine freie Rechtswahl schließt Art. 8 III Rom II-VO dagegen aus. Mit Ausnahme des Urhebervertragsrechts knüpfen also alle urheberrechtlichen Fragestellungen an das Schutzlandprinzip an.³⁰ Nach dem Schutzlandprinzip ist das Recht des Staates anzuwenden, für den der Klagende Schutz beansprucht.³¹ Dieses Prinzip berücksichtigt die kulturellen und wirtschaftlichen Gegebenheiten des in Anspruch genommenen Staates.³² Beispielsweise ist auf die Frage, ob wissenschaftliche Ausgaben *in Frankreich* Schutz genießen, französisches materielles Urheberrecht anzuwenden, für den Schutz *in Deutschland* deutsches Urheberrecht. Möchte der Verfasser die Verwertung der Ausgabe *in Frankreich* untersagen, ist französisches Recht heranzuziehen, für die Verwertung *in Deutschland* dagegen deutsches Recht.³³ Vielschichtige Sachverhalte werden dabei nicht pauschal nach ihrem Kern beurteilt. Vielmehr müssen Anwender und Anwenderinnen die einzelnen Rechtsfragen aufteilen und der zutreffenden Kollisionsnorm³⁴ zuordnen.³⁵

²⁶ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 55.

²⁷ EuGH C-359/14 und C-475/14, ECLI:EU:C:2016:40 Rn. 45 – *ERGO Insurance*.

²⁸ OLG Köln ZUM 2011, 574, 575.

²⁹ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 57.

³⁰ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 46.

³¹ BGH GRUR 2018, 178 Rn. 13 – *Vorschaubilder III*.

³² BGH GRUR 2003, 328, 330 – *Sender Felsberg*; *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 48.

³³ BGH GRUR 2003, 328, 330 – *Sender Felsberg*; *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 48.

³⁴ Kollisionsnormen sind Verweisungsregeln. Sie bestimmen die Rechtsordnung, die auf den Sachverhalt anwendbar ist. Sie betreffen die Rechtslage mithin nicht unmittelbar.

³⁵ Dazu mit Beispielen *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 44.

IV. Anwendungsbereich des UrhG bei internationalen Sachverhalten

Nimmt der Kläger Schutz für das Inland in Anspruch, muss der persönliche sowie der sachliche Anwendungsbereich des UrhG eröffnet sein.³⁶

1. Persönlicher Anwendungsbereich. §§ 120 ff. UrhG

Nach § 120 I 1 UrhG ist der Anwendungsbereich für deutsche Staatsangehörige eröffnet. Nach § 120 I 2 UrhG reicht es bei mehreren Urhebern aus, wenn mindestens einer der Beteiligten deutscher Staatsangehöriger ist. Nach § 120 II UrhG stehen deutschen Staatsangehörigen Deutsche im Sinne des Artikels 116 Abs. 1 GG und Staatsangehörige eines anderen Mitgliedstaates der Europäischen Union oder eines anderen Vertragsstaates des Abkommens über den Europäischen Wirtschaftsraum gleich. Auch sonstige ausländische Urheber und Leistungsschutzberechtigte genießen unter bestimmten Voraussetzungen Schutz, vgl. §§ 121 ff. UrhG.³⁷ Bei der Frage nach der Staatsangehörigkeit ist die Person maßgeblich, der das Recht ursprünglich zustand.³⁸

2. Sachlicher Anwendungsbereich

Aufgrund des formalen Schutzlandprinzips ist das UrhG stets anwendbar, solange Schutz für das Inland in Anspruch genommen wird.³⁹ Somit verlagert sich die Prüfung des räumlichen Anwendungsbereichs in die materielle Prüfung.⁴⁰ Mithin sind die Regeln des UrhG und nicht des Kollisionsrechts entscheidend, ob grenzüberschreitende Handlungen gegen das (deutsche) Urheberrecht verstoßen.⁴¹

Für diese Prüfung ist wiederum das Territorialitätsprinzip heranzuziehen. Eine Handlung verletzt danach das UrhG, soweit ein ausreichender Inlandsbezug vorhanden ist.⁴² Heranzuziehen sind das Auswirkungs- und Handlungsprinzip, welche die Interessen der Rechteinhaber am relevanten Markt berücksichtigen. Damit ist nicht gefordert, dass der Beklagte eine natürliche Handlung in

³⁶ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 88.

³⁷ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 91.

³⁸ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 95.

³⁹ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 99.

⁴⁰ EuGH C-173/11, ECLI:EU:C:2012:642 Rn. 27 f. – *Football Dataco/Sportradar*.

⁴¹ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 99.

⁴² EuGH C-173/11, ECLI:EU:C:2012:642 Rn. 27, 36 ff. – *Football Dataco/Sportradar*; BGH GRUR 1994, 798, 799 – *Folgerecht bei Auslandsbezug*.

Deutschland vornahm. Vielmehr muss sich die Handlung auf das Inland auswirken (Auswirkungsprinzip).⁴³ Bei einer Webseite ist das beispielsweise anzunehmen, wenn diese auf Deutschland ausgerichtet ist.⁴⁴ Anhaltspunkt kann Werbung sein, die im Inland ausgestrahlt wird.⁴⁵

Andererseits lösen Handlungen, die im Inland erfolgen und alle Tatbestandsvoraussetzungen deutscher Verwertungsrechte erfüllen, eine Urheberrechtsverletzung nach dem UrhG aus, auch wenn der tatbestandliche Erfolg im Ausland herbeigeführt wird (Handlungsprinzip).⁴⁶

V. Die Wahrung des Urheberrechts im digitalen Zeitalter

In der EU besteht ein Bündel nationaler, territorial begrenzter Urheberrechte. Dadurch entstehen unterschiedliche Schutzzumfänge und Schutzlängen. Durch das Auswirkungsprinzip besteht die Gefahr einer „uferlose[n] Ausdehnung“.⁴⁷ Grundsätzlich ist mit dem Upload eines Werkes das Recht der öffentlichen Zugänglichmachung aller Staaten betroffenen, in denen der Upload zugänglich ist.⁴⁸ Nach dem EuGH muss dagegen das Online-Angebot die Bewohner eines Staates „gezielt“ ansprechen bzw. es muss an diese „gerichtet“ sein.⁴⁹ Um diese zu beschränken, können Webseitenbetreiber die Technik des Geoblockings heranziehen.⁵⁰ Dadurch können sie bestimmte Inhalte regional sperren.

VI. Ergebnis

Urheberrechte sind immaterielle Rechte. Die Prüfung erfolgt in einem Vierschritt, wobei die gerichtliche Zuständigkeit entscheidet, welches Kollisionsrecht anwendbar ist. Ist ein deutsches Gericht zuständig, bestimmen die Regeln

⁴³ BGH GRUR 2007, 871 Rn. 31 – *Wagenfeld-Leuchte*.

⁴⁴ Vgl. EuGH C-173/11, ECLI:EU:C:2012:642 Rn. 33 ff. – *Football Dataco/Sportradar*.

⁴⁵ EuGH C-516/13, ECLI:EU:C:2015:315 Rn. 30 ff. – *Dimensione Direct Sales und Labianca*; weitere Kriterien bei *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn.118.

⁴⁶ BGH GRUR 2003, 328 – *Sender Felsberg*; *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 106.

⁴⁷ BGH GRUR 2020, 647 Rn. 28 – *Club Hotel Robinson* zum Markenrecht; *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 116.

⁴⁸ BGH GRUR 2016, 1048 Rn. 26 – *An Evening with Marlene Dietrich*.

⁴⁹ EuGH C-173/11, ECLI:EU:C:2012:642 Rn. 39 – *Football Dataco/Sportradar*; ebenso OLG München GRUR-RR 2011, 1, 2 – *Videodateien*; LG Hamburg ZUM 2016, 887, 890; AG Köln GRUR-RS 2021, 9276 Rn. 42 – *Parkettboden*.

⁵⁰ *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 105, 118 aE.

des deutschen inklusive des europäischen internationalen Privatrechts das anzuwendende Recht bei grenzüberschreitenden Sachverhalten, vgl. Art. 3 EG-BGB.

Im internationalen Urheberrecht ist das Schutzlandprinzip zu beachten. Hiernach ist formal das Recht des Staates anzuwenden, für den der Kläger Schutz beansprucht.⁵¹ Aufgrund des formalen Schutzlandprinzips ist das UrhG stets anwendbar, solange Schutz für das Inland in Anspruch genommen wird.⁵²

Ist demnach deutsches Recht anwendbar, ist im dritten Schritt das Fremdenrecht und damit der persönliche Anwendungsbereich zu prüfen, §§ 120 ff. UrhG. In einem vierten Schritt ist zu prüfen, ob nach materiellem deutschem Urheberrecht eine Urheberrechtsverletzung vorliegt. Um eine uferlose Anwendbarkeit zu vermeiden, beschränkt das materielle Urheberrecht diese nach dem Auswirkungs- und Handlungsprinzip. Durch den digitalen Zugang sind Werke grundsätzlich überall aufrufbar. Durch technische Maßnahmen, wie Geoblocking, können Webseitenbetreiber bestimmte Inhalte regional sperren.

Leseempfehlung zur vertiefenden Lektüre: *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120.

⁵¹ BGH GRUR 2018, 178 Rn. 13 – *Vorschaubilder III*.

⁵² *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 120 Rn. 99.

Teil 3: Urheberrechtsfreie Korpora durch abgeleitete Textformate

A. Zugang zu großen Textkorpora des 20. und 21. Jahrhunderts mit Hilfe abgeleiteter Textformate

Versöhnung von Urheberrecht und textbasierter Forschung*

Benjamin Raue/Christof Schöch

Version (2020), CC BY-SA 4.0.

I. Bestandsaufnahme

Ausschließlichkeitsrechte der Urheber können gerade die Analyse großer Datenmengen behindern, die zeitgenössische, urheberrechtlich noch geschützte Texte enthalten. Sowohl die neue Text und Data Mining-Schranke in § 60 d UrhG als auch Lizenzierungsmodelle von Verlagen oder *closed room*-Zugänge von Gedächtniseinrichtungen haben für die wissenschaftliche Praxis entscheidende Nachteile (dazu II.). Deswegen haben wir das Potenzial abgeleiteter Textformate untersucht, die durch Informationsreduktion aus dem Schutzbereich des Urheberrechts gelangen und dennoch eine ausreichende Informationsgrundlage für eine Reihe von digitaler Analysemethoden sein können.

* Dieser Beitrag wurde zuerst veröffentlicht als Raue/Schöch RuZ 2020, 118-127 unter einer CC BY-SA Lizenz.

1. Das weitgehend analoge 20. und 21. Jahrhundert

Als Folge des Urheberschutzes wird etwa in den Computational Literary Studies (CLS)¹ eine erhebliche Einschränkung der digital analysierbaren Textbestände beklagt.² Für diese gebe es ein sehr kurzes *window of opportunity*: Es öffnet sich um 1800, weil sich frühere Schriften mit den verfügbaren Texterkennungstechniken nur unzureichend erschließen lassen. Wegen der urheberrechtlichen Restriktionen enden die verfügbaren Textbestände um das Jahr 1920, weil bei später veröffentlichten Werken die Wahrscheinlichkeit steigt, dass die lange urheberrechtliche Schutzfrist von 70 Jahren p.m.a. noch nicht abgelaufen ist. Die betroffenen Forscherinnen und Forscher empfinden es als starke Beeinträchtigung ihrer Wissenschaftsfreiheit, dass sie ihre Forschungsschwerpunkte nicht primär aufgrund ihres Erkenntnisinteresses bestimmen können, sondern von rechtlichen Faktoren abhängig machen müssen, wenn das Forschungsinteresse nicht mit verfügbaren Textkorpora aus Literatur des 19. Jahrhunderts befriedigt werden kann. Aus literaturwissenschaftlicher Perspektive sind das 20. und 21. Jahrhundert zwar keine dunklen, aber weitgehend analoge Jahrhunderte.

Es ist zudem ein erhebliches Problem, wenn Modelle des Maschinellen Lernens ohne die zum Trainieren eingesetzten Daten publiziert werden müssen. Das gilt insbesondere für komplexe (Sprach-)Modelle der Computerlinguistik. Wer die Trainingsdaten aus urheberrechtlichen Gründen nicht offenlegen darf, schränkt die Reproduzierbarkeit der Ergebnisse und damit deren Anerkennung im internationalen Forschungsdialog ein. Darüber hinaus erschwert das Urheberrecht die Etablierung von Referenzdatensätzen, sogenannter „Benchmarks“, die ihre Funktion ebenfalls nur ausüben können, wenn sie international frei zugänglich gemacht werden können.

2. DFG-Expertenworkshops „Strategien für die Nutzbarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte“

Diese Bestandsaufnahme war Anlass für zwei DFG-Expertenworkshops „Strategien für die Nutzbarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte“.³ Im vorliegenden Beitrag fassen wir die

¹ Die CLS sind ein Teilbereich der Digital Humanities, in dem quantitative Methoden aus Statistik, Informatik und Natural Language Processing für die Bearbeitung literaturwissenschaftlicher Fragestellungen eingesetzt werden.

² Dazu *Schöch* et al, Teil 3, C. in diesem Band.

³ Tagungsbericht von *Erler*, RuZ 2020, 108, 108 ff. unter einer CC BY-SA Lizenz.

Ergebnisse der beiden Workshops zusammen, die in den Beiträgen von *Jotzo* (Teil 2, B.), *Grisse* (Teil 3, B.) und *Schöch et al* (Teil 3, C.) in diesem Sammelband weiter vertieft werden. Der Workshop hat das Potenzial abgeleiteter Textformate im Hinblick darauf untersucht, ob mit ihrer Hilfe rechtssicher große Textkorpora aktueller, urheberrechtlich noch geschützter Textbestände öffentlich zugänglich gemacht werden können. Hinter dem Konzept der abgeleiteten Textformate steht die Idee, urheberrechtlich geschützte Texte durch eine gezielte Informationsreduktion so zu transformieren, dass die urheberrechtlich geschützte Textstruktur irreversibel verloren geht. Ein Werkgenuss im eigentlichen Sinne wird dadurch unmöglich. Weil den Urhebern die in ihren Werken enthaltenen Informationen nicht zugewiesen sind,⁴ ist deren Verwertungsinteresse durch ein öffentliches Zugänglichmachen der abgeleiteten Textformate nicht weiter betroffen, solange die ursprüngliche Struktur nicht mit verhältnismäßigem Aufwand rekonstruiert werden kann. Die Herausforderung ist dabei, Reduktionsverfahren zu entwickeln, die zuverlässig die urheberrechtlich relevanten Elemente entfernen, aber den Korpora so viel Informationen belassen, dass mit ihnen in den digitalen Literaturwissenschaften verbreitete Analyseverfahren weiterhin möglich sind.

II. Lösungsansätze

Für den urheberrechtskonformen Zugang zu großen Textkorpora können vier Lösungsansätze identifiziert werden.⁵

1. Text und Data Mining Schranke

Sowohl der deutsche als auch der europäische Gesetzgeber haben erkannt, welchen Wettbewerbsnachteil das Urheberrecht bei der Analyse großer Textmengen und sonstiger Datenkorpora für die deutsche bzw. europäische Forschungslandschaft haben kann.⁶ Deswegen hat der deutsche Gesetzgeber in § 60 d UrhG eine Urheberrechtsschranke für das nicht-kommerzielle Text und Data Mining eingeführt, die das Erstellen, Aufbereiten und Nutzen eines Textkorpus

⁴ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 196; *Obergfell*, FS Büscher, 223, 225; *Dreier/Schulze*, UrhG, 6. Aufl. 2018, § 60 d Rn. 1; *Raue*, GRÜR 2017, 10, 14.

⁵ Zu den Lösungsansätzen 2 und 3 auch *Schöch et al*, Teil 3, C. in diesem Band.

⁶ Vgl. ErwGr. 10 DSM-RL 2019/790.

für die wissenschaftliche Forschung erlaubt, soweit die Forscher rechtmäßigen Zugang zu den Schutzgegenständen haben.⁷

Die neue Text und Data Mining-Schranke schafft Rechtssicherheit und Forschungsfreiheit, soweit Forscher eigene und öffentlich zugängliche Textbestände mit algorithmischer Unterstützung erforschen wollen. Wesentlicher Nachteil ist momentan aus Sicht der Wissenschaftler, dass die geltende Fassung des § 60 d III UrhG anordnet, dass Forschende das Korpus und sämtliche Vielfältigkeitsstücke des Ursprungsmaterials nach Abschluss der Forschungsarbeiten löschen müssen. Diese Löschungspflicht hat die Text und Data Mining-Schranke in Art. 3 II DSM-RL 2019/790 nun entschärft. Sie erlaubt Forschern nach Ablauf der Umsetzungsfrist im Juni 2021, die Textkorpora ohne zeitliche Beschränkung für die weitere wissenschaftliche Forschung aufzubewahren und zu verwenden.⁸

Die Text und Data Mining-Schranke hat einen weiteren Nachteil. Sie ermöglicht zwar den Aufbau, die Optimierung und die automatisierte Analyse des Textkorpus, nicht aber dessen offene Publikation und Weitergabe. Die Text und Data Mining-Schranke erlaubt lediglich, die Textkorpora in einer Forschungsgruppe bzw. zur Überprüfung der wissenschaftlichen Ergebnisse im Rahmen eines Peer Review-Prozesses zugänglich zu machen, demnächst auch für die spätere Überprüfung wissenschaftlicher Erkenntnisse (Art. 3 II DSM-RL 2019/790). Abgesehen davon dürfen Forscher ihre Datengrundlage der (wissenschaftlichen Fach-)Öffentlichkeit nicht zur Verfügung stellen und ihr so erlauben, die Analyseergebnisse auch außerhalb formaler Qualitätssicherungsprozesse transparent nachzuvollziehen. Auch die Nachnutzung teilweise sehr aufwendig aufbereiteter Textkorpora für die Anschlussforschung Dritter wird dadurch verhindert. Beide Aspekte sind jedoch für eine transparente und nachhaltige Forschung sowie den Zugang zu großen Referenzkorpora von erheblicher Bedeutung.

Darüber hinaus berechtigt die Schranke nur zum Minen von Texten, zu denen die Forschenden rechtmäßigen Zugang haben.⁹ Deswegen müssen sie in analoger Form vorliegende Textbestände selbst digitalisieren¹⁰ und ggf. für die weitere

⁷ Zu § 60 d UrhG etwa *Raue*, CR 2017, 656 ff.; *Spindler*, ZGE 2018, 273, 277 ff.; *Specht*, OdW 2018, 285; *Obergfell*, FS Büscher, 223, 228 ff.

⁸ *Spindler*, CR 2019, 277, 279 f.; *Raue*, ZUM 2019, 684, 688.

⁹ RegE, BT-Drucks. 18/12329, 41; *Raue*, CR 2017, 656, 658; *Dreier/Schulze*, UrhG, 6. Aufl. 2018, § 60 d Rn. 4.

¹⁰ Wozu sie § 60 d UrhG berechtigt, RegE, BT-Drucks. 18/12329, 41; *Raue*, CR 2017, 656, 659; *Dreier/Schulze*, UrhG, 6. Aufl. 2018, § 60d Rn. 4. Sie dürfen dafür auch Dritte einsetzen, *Dreier* aaO Rn. 5.

Verwendung aufbereiten. Das erschwert und verteuert das Zusammenstellen sehr großer Textkorpora oder die Etablierung moderner Referenzkorpora, mit denen die Leistungsfähigkeit neuer Analysetools bzw. -ansätze demonstriert werden kann.

Der Gesetzgeber steht aber vor dem Dilemma, dass er ohne eine Beschränkung der öffentlichen Zugänglichmachung kaum das berechtigte und in internationalen Konventionen¹¹ abgesicherte Interesse der Rechteinhaber sicherstellen kann, durch die Text und Data Mining-Schranke die weitere Verwertung ihrer Inhalte nicht unmöglich zu machen.

2. Lizenzierungslösungen

Soweit kein eigener Zugang zu (größeren) Textkorpora besteht, stellen etwa große Verlage Zugang zu ihren Textkorpora über Schnittstellen (API) zur Verfügung. Allerdings sind diese Textbestände im Regelfall auf das Repertoire des jeweiligen Verlags beschränkt; auch gewähren nicht alle Verlage Zugang. Um ihre weiteren Verwertungsinteressen nicht zu gefährden, untersagen Verlage im Regelfall die Weitergabe der Textkorpora.

Teilweise beschränken die Datenbankinhaber weitergehend sogar den direkten Zugriff auf die Textbestände. Den Forschern ist es in diesen Fällen unmöglich, die Texte über die API herunterzuladen und die Qualität des Forschungskorpus durch Annotationen, Normalisierung oder durch die Ergänzung mit anderen Datensätzen zu verbessern. Zudem können die Forschenden typischerweise auch nicht ihre selbstentwickelten Analyseprogramme verwenden, sondern sind auf die Analysetools des jeweiligen Datenbankankbieters beschränkt.

3. Closed Room-Zugang (§§ 60 e IV, 60 f I UrhG)

Beim sogenannten *closed room*-Zugang machen insbesondere öffentlich zugängliche Bibliotheken und Archive, gestützt auf §§ 60 e IV, 60 f I UrhG, ihre Bestände auf Workstations in den Räumen der Institution zugänglich. Den Zugriff von außerhalb, auch mittels VPN-Clients wollte der Gesetzgeber nicht freistellen.¹² Zwar wird hier den Forschern typischerweise ermöglicht, ihre eigenen

¹¹ Etwa durch den 3-Stufen-Test in Art. 13 TRIPs und Art. 10 I WCT, wonach Beschränkungen und Ausnahmen von ausschließlichen Rechten auf bestimmte Sonderfälle beschränkt werden müssen und diese weder die normale Auswertung des Werkes beeinträchtigen noch die berechtigten Interessen des Rechtsinhabers unzumutbar verletzen dürfen.

¹² BT-Drs. 16/1828, 26; Dreier/Schulze, UrhG, 6. Aufl. 2018, § 60 e Rn. 17.

Analysertools zu verwenden. Erheblicher Nachteil ist neben der Ortsgebundenheit und der fehlenden Möglichkeit, diesen Datensatz durch die Kombination mit anderen Datensätzen zu erweitern, dass die Forschenden das Textkorpus weder für die eigene Nachnutzung kopieren noch Dritten für die wissenschaftliche Nachprüfung zur Verfügung stellen dürfen.

4. Abgeleitete Textformate

Dementsprechend lösen die Lizenzierungs- und *closed room*-Lösungen im Vergleich zu der Freistellung durch die Text und Data Mining-Schranke lediglich das Problem des rechtmäßigen Zugangs, nicht aber die nur sehr eingeschränkte Möglichkeit, Dritten die verwendeten Textkorpora zur Verfügung zu stellen. Diese Schwächen können abgeleitete Datenformate lösen, bei denen durch eine Informationsreduktion der urheberrechtliche Schutz der gespeicherten Texte entfällt. Der große und entscheidende Vorteil dieser Lösung ist, dass auf diese Weise sehr große Textkorpora öffentlich zugänglich gemacht und weiterverwendet werden können, ohne dass die berechtigten Interessen der Urheber dadurch berührt werden. Der auf der Hand liegende Nachteil dieser Lösung liegt in ihrer Konstruktion begründet: Durch die Informationsreduktion können mit diesen abgeleiteten Textformaten nicht alle Analyseverfahren durchgeführt werden. Zudem kann in einzelnen Fällen die Qualität der Ergebnisse der dennoch möglichen Analyseverfahren sinken.¹³ Weil diese Lösung aber als einzige eine Veröffentlichung der verwendeten Textkorpora und eine Zusammenführung verschiedener Textkorpora ermöglicht, soll sie nun näher dargestellt werden.

III. Anforderungen an abgeleitete Textformate aus Sicht der Rechtswissenschaft und der Anwendungswissenschaften

1. Urheberrechtliche Vorgaben und Lösungsstrategien

a) Schutzfähigkeit

Aus urheberrechtlicher Sicht muss zunächst definiert werden, was die urheberrechtliche Schutzfähigkeit von Texten ausmacht.¹⁴ Bei Sachtexten sind die in

¹³ Hierzu besteht weiterer Forschungsbedarf, dazu: *Schöch* et al, Teil 3, C. in diesem Band.

¹⁴ Dazu ausführlich *Jotzo*, Teil 2, B. in diesem Band, 61, 64 ff.

ihnen enthaltenen Informationen urheberrechtlich nicht schutzfähig.¹⁵ Die nach § 2 II UrhG erforderliche Individualität können daher allein ihre Struktur und Gedankenführung sowie individuelle Formulierungen erreichen. Gerade bei Gebrauchstexten werden hieran relativ hohe Anforderungen gestellt, so dass kleinere Textausschnitte typischerweise urheberrechtlich schutzunfähig sind.¹⁶ Literarische Texte weisen dagegen mit hoher Wahrscheinlichkeit eigenschöpferische Inhalte und Formulierungen auf. Zudem können fiktionale Inhalte, insbesondere Erzählstränge, Charakteristika der Figuren und deren Beziehungsgefüge urheberrechtlich geschützt sein.¹⁷

b) Reduktion urheberschutzbegründender Merkmale

Dementsprechend müssen aus abgeleiteten Textformaten die schutzbegründenden individuellen Formulierungen, Strukturen und fiktionalen Elemente entfernt werden.¹⁸

aa) Nichterkennbarkeit schutzfähiger Elemente

Als maßgebliches Kriterium dafür wird man die Erkennbarkeit der schutzbegründenden Elemente heranziehen müssen.¹⁹ Vervielfältigungen und das öffentliche Zugänglichmachen eines Textkorpus greifen danach nicht mehr in das Urheberrecht der Ursprungstexte ein, wenn das Korpus keine schutzfähigen Teile enthält bzw. diese nicht mehr erkennbar sind.²⁰

Problematisch ist, dass kleinere Textausschnitte nur *typischerweise* urheberrechtlich schutzunfähig sind.²¹ Wird ein großes Textkorpus automatisiert aus einer Vielzahl von kurzen Textausschnitten zusammengestellt, kann daher nicht

¹⁵ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 196; *Obergfell*, FS Büscher, 223, 225; *Dreier/Schulze-Dreier*, UrhG, 6. Aufl. 2018, § 60 d Rn. 1; *Raue*, GRUR 2017, 10, 14.

¹⁶ Dazu *Jotzo*, Teil 2, B. in diesem Band, 61, 70 f mwN.

¹⁷ *Jotzo*, Teil 2, B. in diesem Band, 61, 69 mwN.

¹⁸ Dazu ausführlich *Grisse*, Teil 3, B. in diesem Band, 141, 145 ff.

¹⁹ *Grisse*, Teil 3, B. in diesem Band, 141, 145ff in Verallgemeinerung von EuGH, Urt. v. 29.07.2019 – C-476/17, ECLI:EU:C:2019:624 = GRUR 2019, 929 Tz. 26 ff., 39 – *Pelham*.

²⁰ *Grisse*, Teil 3, B. in diesem Band, 141, 144f, 148f.

²¹ Der EuGH hat entschieden, dass Textausschnitte mit elf Wörtern schutzfähig sein können, EuGH, Urt. v. 16.7.2009 – C-5/08, ECLI:EU:C:2009:465 = GRUR 2009, 1041 Tz. 48 – *Infopaq I*. Vgl. hierzu das Beispiel des OLG München, Urt. v. 17.9.2009 – 29 U 3271/09, ZUM 2009, 970 f. – *Typisch München!*, wonach die Formulierung „Vom Ernst des Lebens halb verschont / Ist der schon, der in München wohnt“ urheberschutzfähig ist. Dazu *Jotzo*, Teil 2, B. in diesem Band, 61, 70.

für alle Textarten mit Sicherheit ausgeschlossen werden, dass das Korpus vereinzelt schutzfähige Textausschnitte enthält.²² Diese Problematik lässt sich aus unserer Sicht aber mit der Schranke des § 57 UrhG lösen. Vereinzelt urheberrechtlich schutzfähige Elemente sind in einem sehr großen Textkorpus als „unwesentliches Beiwerk neben dem eigentlichen Gegenstand der Vervielfältigung, Verbreitung oder öffentlichen Wiedergabe anzusehen“. § 57 UrhG setzt Art. 5 Abs. 3 lit. i InfoSoc-RL um, wonach eine „beiläufige Einbeziehung eines Werks oder sonstigen Schutzgegenstands in anderes Material“ freigestellt wird. Diese Voraussetzungen erfüllen vereinzelt schutzfähige Textsequenzen, weil sie in das Textkorpus nicht wegen ihrer einprägsamen, individuellen Formulierung einbezogen worden sind, sondern allein wegen der darin enthaltenen Sequenzinformationen, die urheberrechtlich gerade nicht schutzfähig sind. Die Textteile werden also nicht wegen, sondern trotz ihres schutzfähigen Gehalts aus unvermeidbaren technischen Gründen²³ in das Textkorpus einbezogen. Nach dem BGH liegt die Voraussetzung der „beiläufigen Einbeziehung“ immer dann vor, wenn einem durchschnittlichen Betrachter ein urheberschutzfähiger Bestandteil in dem eigentlichen Gegenstand, also hier dem Gesamttextkorpus, nicht auffällt.²⁴ Ein großes Textkorpus wird von einem durchschnittlichen Betrachter oder durchschnittlichen Nutzer aber gerade nicht lesend, sondern nur in Form einer statistischen Analyse betrachtet. Dabei wird ein einzelnes schutzfähiges Textelement gerade nicht wahrgenommen, sondern lediglich bei der statistischen Auswertung berücksichtigt.

bb) Nicht-Rekonstruierbarkeit

Neben der Nichterkennbarkeit wird man als zweite Anforderung von einem urheberrechtsfreien, abgeleiteten Textformat verlangen müssen, dass die ursprünglichen Texte nicht aufgrund von Positionsangaben der Textsequenzen oder sonstiger Sequenzinformationen mit verhältnismäßigem Aufwand rekonstruierbar sind.²⁵ Denn wenn die ursprünglichen, urheberschutzfähigen Texte von Nutzern rekonstruiert werden könnten, wären die abgeleiteten Textformate geeignet, Verwertungsinteressen der Urheber zu beeinträchtigen.

²² *Jotzo*, Teil 2, B. in diesem Band, 61, 71 f; *Grisse*, Teil 3, B. in diesem Band, 141, 147-

²³ Vgl. dazu EuGH, Urt. v. 18.10.2012 – C-173/11, ECLI:EU:C:2012:642 Tz. 36 – *Football Dataco/Sportradar*; EuGH, Urt. v. 12.07.2011 – C-324/09, ECLI:EU:C:2011:474 Tz. 64 – *L'Oréal/eBay*; BGH, Urt. v. 9.11.2017 – I ZR 134/16 = GRUR 2018, 417 Tz. 46 ff. – *Resistograph* (zur Begrenzung des internationalen Anwendungsbereichs des Urheber- bzw. Markenrechts).

²⁴ BGH, Urt. v. 17.11.2014 – I ZR 177/13 = GRUR 2015, 667 Tz. 27 – *Möbelkatalog*.

²⁵ *Grisse*, Teil 3, B. in diesem Band, 141, 149 f.

c) Rechtmäßigkeit der Umwandlungshandlung

Um den Ursprungstext in ein abgeleitetes Textformat zu transformieren, muss dieser im Arbeitsspeicher eines Computers abgespeichert werden. Eine solche Vervielfältigung fällt in den Schutzbereich des Vervielfältigungsrechts nach § 16 UrhG. Daher muss sich der Umwandelnde für die Transformation auf eine urheberrechtliche Schranke berufen können.²⁶ Grundvoraussetzung für alle Schranken ist im Regelfall, dass der Umwandelnde rechtmäßigen Zugang zu den Ursprungstexten hat.²⁷ In diesem Fall kann die Umwandlung des Ursprungstextes in ein urheberrechtsfreies Format auf § 44 a UrhG gestützt werden, wenn der Ursprungstext nur vorübergehend im Arbeitsspeicher vervielfältigt und danach automatisch aus diesem wieder gelöscht wird.²⁸ Weil das ursprüngliche Werk in diesem Fall im abgewandelten Textformat nicht mehr erkennbar ist, sind urheberpersönlichkeitsrechtliche Interessen nicht betroffen.²⁹

2. Identifikation geeigneter Textformate

a) Verfahren

Vor diesem Hintergrund wurden im Workshop verbreitete Analyseverfahren der Digital Humanities daraufhin untersucht, ob sie auch auf Textkorpora angewendet werden können, deren Informationsgehalt erheblich reduziert wurde, um die urheberrechtlichen Vorgaben zu erfüllen. Aus Sicht der Anwendungswissenschaften lassen sich zwei große Gruppen von Analyseverfahren unterscheiden.³⁰ Der ersten Gruppe zuzuordnen sind Analyseverfahren, die auf eine präzise Sequenzinformation der einzelnen Wortformen im Textverlauf angewiesen sind (insb. die meisten Verfahren aus der Sentiment Analyse, einige Verfahren der Netzwerkanalyse, Verfahren des Text Re-Use sowie das Erstellen von Sprachmodellen).³¹ Zur zweiten Gruppe gehören Analyseverfahren, die eine solche präzise Sequenzinformation nicht erfordern. Diese Verfahren können auf Worthäufigkeiten operieren, profitieren aber stark davon, wenn sich diese Häufigkeiten auf kleinere Segmente innerhalb der Texte beziehen sowie

²⁶ Dazu ausführlich *Grisse*, Teil 3, B. in diesem Band, 141, 155 f.

²⁷ *Grisse*, Teil 3, B. in diesem Band, 141, 155.

²⁸ *Grisse*, Teil 3, B. in diesem Band, 141, 156 ff. sowie zur Anwendbarkeit von §§ 60 c und 60 d UrhG.

²⁹ *Grisse*, Teil 3, B. in diesem Band, 141, 159 f.

³⁰ Dazu ausführlich *Schöch et al*, RuZ 2020, 160, 163, 169 unter einer CC BY-SA Lizenz.

³¹ Die einzelnen Verfahren werden vorgestellt von *Schöch et al*, Teil 3, C. in diesem Band.

wenn die Information über die Reihenfolge der Segmente im Gesamttext erhalten bleibt (dazu gehören Verfahren der Autorschaftsattribuion, die Extraktion distinktiver Merkmale und das Topic Modeling sowie ggf. einige Verfahren der Netzwerkanalyse).

b) Geeignete abgeleitete Textformate

Vereinfachend lassen sich Texte als eine Zeichenfolge auffassen, die man in verschiedene Teilabschnitte gliedern kann: Token (bspw. einzelne Wörter), Sätze, Segmente und Gesamttext. Die Token wiederum können mit explizierenden Informationen angereichert werden (Wortform, Lemma, Wortart, semantischer Gehalt, Relationen, Sequenzinformation, Häufigkeit). Abgeleitete Textformate entstehen dadurch, dass eine gezielte Kombination von Informationsanreicherung und Informationsreduktion gewählt wird.³² Es lassen sich zwei Grundansätze unterscheiden:

aa) Token-basierte Textformate

Token-basierte, abgeleitete Textformate gehen von den einzelnen Wörtern als Grundeinheit aus und basieren darüber hinaus in der Regel auf dem vollständigen Einzeltext als Analyseeinheit. Im einfachsten Fall enthält ein solches Format lediglich die Information über die Häufigkeit der verschiedenen Wortformen in jedem einzelnen Text als Ganzes, was einer Entfernung jeglicher Sequenzinformation entspricht. Solche Formate sind im Regelfall urheberrechtlich unbedenklich.³³

Etwas reichhaltigere Textformate beruhen auf der beschriebenen Anreicherung der Tokens um sprachlich relevante Informationen (wie Lemma, Wortart, Syntax, Semantik); auf der Segmentierung der Texte in kleinere Einheiten, sodass die Häufigkeitsinformationen sich feiner aufgliedern; und auf der gezielten Entfernung von Informationen wie der Wortform oder des Lemmas unter Beibehaltung bspw. der Wortart-Information. Je nach spezifischem Textformat und den verwendeten Parametern ergeben sich hier unterschiedliche Einschätzungen bezüglich der Nützlichkeit für die Forschung in den Digital Humanities einerseits, bezüglich der urheberrechtlichen Beurteilung andererseits. Aus rechtlicher Perspektive kann das Löschen hinreichend vieler Stopp- oder Funktionswörter (z.B. Artikeln, Konjunktionen, Pronomen, Präpositionen, Modal- und

³² Dazu ausführlich *Schöch et al*, RuZ 2020, 160, 161 unter einer CC BY-SA Lizenz.

³³ *Grisse*, Teil 3, B. in diesem Band, 141, 151.

Hilfsverben)³⁴, ggf. kombiniert mit der Entfernung von Eigennamen ein gangbarer Weg sein.³⁵ Vielversprechender – sowohl aus Sicht der Anwender als auch des Urheberrechts – sind aber Formate mit gestörten Sequenzinformationen.³⁶ Hierzu werden die Wörter eines Textsegments, bildlich gesprochen, durcheinandergewirbelt und so die urheberrechtlich geschützte Struktur und ggf. ein fiktionaler Inhalt unkenntlich gemacht. Bei einer ausreichend großen Segmentlänge von etwa 50 bis 100 Wörtern ist der ursprüngliche Text kaum noch rekonstruierbar, aber für eine Reihe von Analysemethoden nutzbar.

bb) Textformate auf (Teil-)Korpusebene

Eine andere Gruppe von Textformaten verwendet nicht die vollständigen Einzeltexte als Bezugsgröße, sondern größere Bestände von Texten. Diese werden aber so in Teilbestände gegliedert (bspw. nach Jahr und/oder Textsorte), dass dennoch analytisch gestützte Aussagen über die stilistischen Eigenschaften oder inhaltlichen Muster der Textgruppen als Ganzes gemacht werden können. Hierzu gehören Textformate, die statt der Häufigkeit einzelner Tokens die Häufigkeiten von Wortsequenzen einer voreingestellten Länge von N-Wörtern (sog. N-Gramme)³⁷ verfügbar machen: auf Einzeltext-Ebene wäre hier die Rekonstruierbarkeit trivial; wenn die N-Gramm-Häufigkeiten sich aber auf eine größere Textmenge beziehen und seltene N-Gramme gelöscht werden, ist das vermutlich nicht mehr der Fall. Soweit die Ursprungstexte nicht rekonstruierbar sind, sind solche Formate urheberrechtlich unbedenklich, solange keine zu großen N-Gramme gewählt werden.³⁸ 5-Gramme etwa, also Textsegmente mit der Länge von 5 Wörtern, bilden bei den meisten Textgattungen keine urheberschutzfähigen Textteile. Sollten sehr ungewöhnliche Formulierungen ausnahmsweise doch schutzfähig sein, so würden diese wegen ihrer Seltenheit gelöscht werden. Sollte auch dieser Filter versagen, etwa weil das Fragment oft zitiert wird, darf es wegen § 57 UrhG trotzdem in große Textkorpora aufgenommen werden.³⁹

In diese Gruppe gehören zudem vektorbasierte Textformate, die auf der Repräsentation der Wörter in Form von Vektoren beruhen, die deren syntaktische

³⁴ Hierzu *Schöch et al*, RuZ 2020, 160, 167 unter einer CC BY-SA Lizenz.

³⁵ *Grisse*, Teil 3, B. in diesem Band, 141, 152.

³⁶ Dazu aus Anwendersicht *Schöch et al*, RuZ 2020, 160, 165 ff. unter einer CC BY-SA Lizenz; und aus urheberrechtlicher Perspektive *Grisse*, Teil 3, B. in diesem Band, 141, 152 f.

³⁷ Dazu mit Beispielen *Schöch et al*, RuZ 2020, 160, 169 ff. unter einer CC BY-SA Lizenz.

³⁸ *Grisse*, Teil 3, B. in diesem Band, 141, 153f.

³⁹ Dazu oben III. 1. b) aa).

und semantische Eigenschaften kodieren und damit für Analysen nutzbar machen.⁴⁰ Diese werden vor allem für die computergestützte Verarbeitung natürlicher Sprache verwendet (Natural Language Processing, NLP). Urheberrechtlich sind sie grundsätzlich unbedenklich, weil die Texte mangels Textgestalt für einen Menschen nicht mehr lesbar sind. Es muss aber sichergestellt sein, dass die ursprüngliche Textform nicht rekonstruiert werden kann.

IV. Fazit

Intelligent konstruierte, abgeleitete Textformate ermöglichen, große Korpora urheberrechtlich noch geschützter Texte zu erforschen und frei zugänglich zu machen. Damit gleichen sie die Schwächen anderer Lösungen aus, die versuchen, Interessen von Urhebern und Forschern in Einklang zu bringen. Die abgeleiteten Textformate sind allerdings nur deswegen urheberrechtsfrei, weil bei ihnen Wortgruppen, die Textstruktur bzw. andere Informationen gezielt entfernt wurden. Deswegen sind mit ihnen zwar viele, aber nicht alle wissenschaftlichen Analysemethoden möglich. Als strategische Ergänzung vorhandener Lösungsansätze können sie das literaturwissenschaftliche 20. und 21. Jahrhundert trotzdem digital auf.

⁴⁰ Dazu *Schöbch et al*, RuZ 2020, 160, 171 ff. unter einer CC BY-SA Lizenz.

B. Nutzbarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte*

Karina Grisse

Version (2020), CC BY-SA 4.0.

I. Einleitung

Beim Textmining werden große Mengen von Texten gesammelt, digital aufbereitet und zur Beantwortung von Forschungsfragen automatisiert ausgewertet. Große Textkorpora können außerdem dazu genutzt werden, intelligente Sprachsysteme zu trainieren. Tragen Forschende Texte für Textmining-Prozesse in digitaler Form zusammen und werten sie diese maschinell aus, entstehen zwangsläufig digitale Kopien. Sofern die Texte dem Schutz des Urheberrechtsgesetzes unterliegen, wird dadurch in das den Schutzrechtsinhabern gemäß § 16 UrhG zugewiesene Vervielfältigungsrecht eingegriffen. Vervielfältigungshandlungen sind nur mit Zustimmung der Rechteinhaber oder bei Eingreifen einer gesetzlichen Schrankenbestimmung zulässig. Damit das Urheberrecht der Nutzung des enormen Forschungspotentials von Text- und Data-Mining (TDM)¹ nicht im Wege steht, hat der deutsche Gesetzgeber § 60d UrhG als Schrankenregelung eingeführt,² mit dem Ziel, Rechtssicherheit für TDM zu Forschungszwecken zu schaffen. Die 2019 beschlossene Richtlinie 2019/790/EU³ (UrhR-RL) enthält in den Art. 3 und Art. 4 ebenfalls zwei Schrankenregelungen zu Gunsten

* Dieser Beitrag wurde zuerst veröffentlicht als *Grisse RuZ 2020*, 143-159 unter einer CC BY-SA Lizenz.

¹ Für Anwendungsbeispiele siehe: *Schöch et al.*, *RuZ 2020*, 160, 162 ff. unter einer CC BY-SA Lizenz; *Spindler*, *GRUR 2016*, 1112, 1113; *Raue*, *GRUR 2017*, 11, 12 f.

² Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft – *UrhWissG*, *BGBI 2017 Teil I Nr. 61*, 3346 ff.

³ Richtlinie 2019/790/EU des Europäischen Parlaments und des Rates vom 17.4.2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG.

von TDM.⁴ Diese Schrankenbestimmungen stellen insbesondere Vervielfältigungshandlungen zum Zwecke des TDM frei.⁵ Zulässig ist nach § 60d Abs. 1 Nr. 2 UrhG, der sich auf die allgemeine Wissenschaftsschranke in Art. 5 Abs. 3 lit. a InfoSocRL stützt, auch das Zugänglichmachen eines Korpus innerhalb einer abgegrenzten Forschungsgruppe und gegenüber einzelnen Reviewern (Abs. 1 Nr. 2). Dabei bleibt es auch nach der Umsetzung der UrhR-RL, die dazu keine eigene Regelung trifft, sondern auf die vorbestehende Wissenschaftsschranke verweist.⁶

Obwohl mit den genannten Regelungen für die Forschung einiges gewonnen ist, bleibt das Urheberrecht in vielen Situationen ein Hindernis. Zunächst setzen die Schranken voraus, dass bereits ein rechtmäßiger Zugang zu den verwendeten Werken besteht, verschaffen aber kein Zugangsrecht.⁷ Schwierigkeiten bereiten im gegebenen Rechtsrahmen vor allem ressourcensparende Anschlussnutzungen, bei denen bestehende Korpora für neue Forschungsfragen genutzt werden sollen. Hieran besteht aber ein großes Interesse, denn die Neuerstellung ist ressourcenaufwändig. Nach Abschluss eines Forschungsprojektes dürfen Korpora zwar unter bestimmten Bedingungen aufbewahrt werden,⁸ ob aber Anschlussnutzungen der archivierten Korpora für die Beantwortung neuer Forschungsfragen erlaubt sind, bleibt de lege lata unklar. Begründbar wäre dies jedenfalls nur für nicht kommerzielle Forschung;⁹ die Modalitäten der Verfügbarma-

⁴ Das Bundesministerium der Justiz und für Verbraucherschutz (BMJV) hat am 13.10.2020 einen Referentenentwurf für ein „Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes“ veröffentlicht, der auch einen Umsetzungsvorschlag für Art. 3 und Art. 4 der RL 2019/790/EU beinhaltet; zu dem Referentenentwurf vorangehenden Diskussionsentwurf: *Raue*, ZUM 2020, 172.

⁵ Ausführlich zum Inhalt von § 60d UrhG: *Raue*, CR 2017, 656; zu den Art. 3 und Art. 4 UrhR-RL: *Raue*, ZUM 2019, 684, 684 ff.; *Spindler*, CR 2019, 277, 278 f.

⁶ ErwGr 15 S. 5 UrhR-RL.

⁷ Dazu *Raue*, CR 2017, 656, 658.

⁸ De lege lata nach § 60d Abs. 3 S. 2 UrhG nur durch die in den §§ 60e und f UrhG genannten Organisationen. Art. 3 UrhR-RL enthält diese Beschränkung nicht, vgl. insofern § 60d UrhG-E. Zu den bestehenden Unsicherheiten hinsichtlich des Speicherortes siehe die Stellungnahme des GRUR Fachausschusses für Urheber- und Verlagsrecht zum Diskussionsentwurf BMJV, S. 3, abrufbar unter: https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/DE/Gesetz_I_Anpassung-Urheberrecht-dig-Binnenmarkt.html, zuletzt abgerufen am 19.10.2020.

⁹ So *Raue*, der eine „Annexkompetenz“ konstruiert: *Raue*, CR 2017, 656, 661; hierauf verweist: *Dreier/Schulze-Dreier*, Urheberrechtsgesetz, 6. Aufl. 2018, § 60d UrhG Rn. 13.

chung blieben selbst dann unklar. Die neuen Richtlinienbestimmungen dürften an diesem Ergebnis und den bestehenden Unsicherheiten¹⁰ nichts ändern. Jedenfalls außerhalb der speziellen Regelung des Art. 3 UrhR-RL für nicht kommerzielle Forschungsorganisationen ist die Aufbewahrung und weitere Nutzung der zum Zwecke eines TDM-Projektes erstellten Vervielfältigungsstücke auf die Nutzung zu eben diesem Projekt und der Überprüfung der Erkenntnisse beschränkt.¹¹

Darüber hinaus besteht großes Interesse an der Erstellung und dem Zugriff auf große Korpora, deren Textauswahl nicht von vornherein ausgerichtet an einer bestimmten Forschungsfrage zusammengestellt ist, die vielmehr derart offen angelegt ist, dass auch weitere Forschungsfragen damit bedient werden können. Vor allem für die Computerlinguistik und das Trainieren künstlicher Intelligenz sind sehr große, verschiedenste Textgattungen umfassende Textkorpora essentiell. Der Zugriff auf vorbestehende, z.B. ganze Bibliotheksbestände umfassende Korpora wäre für viele Forschungsprojekte eine Bereicherung.¹² Die bestehenden Schrankenbestimmungen erlauben es nicht, Korpora projektunabhängig zu erstellen und öffentlich an eine unbestimmte Zahl potenzieller Empfänger, etwa einem großen Forschungsnetzwerk, universitätsweit oder gar frei auf einer Internetseite zugänglich zu machen.

Oft ist es für die Forschung jedoch nicht erforderlich, dass ein Text in seiner (geschützten) Ursprungsform oder überhaupt als sinnvoll lesbarer Text verwendet werden kann. Wie *Christof Schöch et al* erläutern, können viele Analyseverfahren unter Verwendung von aus Primärtexten abgeleiteten Textformaten durchgeführt werden.¹³ Hierin liegt die Chance, auch geschützte Texte – jedoch in anderer Form, in abgeleiteten Formaten –, außerhalb der urheberrechtlichen Schrankenregelungen für die Forschung durch Dritte nutzbar zu machen, ohne mit den berechtigten Interessen der Rechteinhaber in Konflikt zu geraten.

¹⁰ Siehe exemplarisch: Stellungnahme der Deutschen Nationalbibliothek zum Entwurf eines Ersten Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes vom 31.01.2020, abrufbar unter: https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/DE/Gesetz_I_Anpassung-Urheberrecht-dig-Binnenmarkt.html, zuletzt abgerufen am 19.10.2020.

¹¹ ErwGr 18 Abs. 1 UrhR-RL.

¹² Man denke z.B. (aber natürlich nicht nur) an die Doktorandin, der bei fehlendem Anschluss an die erforderliche technische Infrastruktur die Erstellung und Annotation eines eigenen umfangreichen Textkorpus unmöglich sein könnte. Zur möglichen zentralen Rolle von Bibliotheken und Archiven bei der Etablierung abgeleiteter Formate: *Schöch et al*, Teil 3, C. in diesem Band.

¹³ *Schöch et al*, RuZ 2020, 160, 162 ff. unter einer CC BY-SA Lizenz.

Florian Jotzo hat in diesem Zusammenhang aufgezeigt, welche Schutzrechte das Urheberrechtsgesetz für Texte unter welchen Bedingungen vorsieht. Dieser Beitrag knüpft daran an. Nachdem zunächst (II.) kurz die Möglichkeit der Verwendung abgeleiteter Textformate als Forschungsgegenstand umrissen wird, zeigt der Beitrag auf, welche Anforderungen abgeleitete Formate aus urheberrechtlicher Sicht erfüllen müssen (III.1.), damit ihre Verwendung nicht in Schutzrechte eingreift. Es werden dann konkrete Formate betrachtet (III.2.). Schließlich werden die rechtlichen Voraussetzungen erläutert, die beim Erstellen abgeleiteter Formate zu beachten sind (IV.).

II. Abgeleitete Textformate als Forschungsgegenstand

Abhängig von der konkret mithilfe von Textmining zu beantwortenden Forschungsfrage benötigt man für die Beantwortung häufig nicht die Texte in ihrer geschützten Ursprungsform, sondern lediglich die in den Texten enthaltenen, ohnehin nicht geschützten Informationen oder ebenfalls nicht geschützte Informationen über die Texte.¹⁴ Für Autorschaftsattributionsverfahren, bei denen Texte unbekannter oder umstrittener Autorschaft einem Autor zugeordnet werden sollen, um nur ein Beispiel zu nennen, kommt es entscheidend auf den Wortschatz des Textes und die Worthäufigkeit pro Text an, während das Verständnis des konkreten Textinhaltes keine Rolle spielt.¹⁵ So können z.B. aus Primärtexten erzeugte Worthäufigkeitsstatistiken (sog. Term-Dokument Matrizen), lemmatisierte und annotierte Wortlisten oder auch Wortvektor-basierte Repräsentationen für bestimmte Einsatzfelder der Digital Humanities bereits wertvolle Analysequellen sein.¹⁶ Solche Formate lassen sich mithilfe von Algorithmen aus Primärtexten automatisiert generieren. Die Rede ist deshalb hier von abgeleiteten Formaten.

Bei entsprechender Gestaltung enthalten die Formate keine sinnvoll lesbaren Versionen der Primärtexte. Sie ermöglichen also nicht den Werkgenuss. Wiederrum bei entsprechender Gestaltung handelt es sich bei den abgeleiteten Formaten weder um Vervielfältigungen i.S.d. § 16 UrhG noch um Bearbeitungen der Primärtexte i.S.d. § 23 UrhG, sondern um Formate, die außerhalb des Schutz-

¹⁴ *Schöch et al.*, RuZ 2020, 160, 162 ff. unter einer CC BY-SA Lizenz, erläutern exemplarisch, welche Informationen für gängige TDM-Analyseverfahren der Digital Humanities benötigt werden; siehe auch *Schöch et al.*, Teil 3, C. in diesem Band.

¹⁵ *Schöch et al.*, RuZ 2020, 160, 165 unter einer CC BY-SA Lizenz..

¹⁶ Siehe für die Beschreibung solcher Formate und deren Verwendungsmöglichkeiten: *Schöch et al.*, RuZ 2020, 160, 163 ff. unter einer CC BY-SA Lizenz.

umfangs der Primärtexte liegen. Die abgeleiteten Formate dürfen die schutzbe-
gründenden Merkmale nicht enthalten. Dann werden die durch Urheberrecht
und Leistungsschutzrechte geschützten Verwertungsinteressen der Rechteinha-
ber bei der Nutzung solcher Formate zu Forschungszwecken und zur For-
schungstransparenz nicht beeinträchtigt. Die Verwendung und Zugänglichma-
chung der abgeleiteten Formate für automatisierte Analyseverfahren ist deshalb
auch außerhalb der urheberrechtlichen Schrankenbestimmungen möglich.
Die Herausforderung besteht darin, Formate zu finden, die sich technisch zu-
verlässig ohne zu hohen Rechenaufwand erstellen lassen.¹⁷ Die Formate dürfen
den Schutz der Primärtexte nicht tangieren, müssen aber dennoch die for-
schungsrelevanten Textinformationen oder -eigenschaften enthalten.

III. Urheberrechtliche Anforderungen an abgeleitete Formate

1. Kriterien für nicht schutzrechtsbeeinträchtigende abgeleitete Formate

Ein abgeleitetes Format berührt die Rechte am Primärtext nicht, wenn es sich
nicht um eine (relevante) (Teil)Vervielfältigung des Primärtextes handelt. In der
Pelham-Entscheidung hat der EuGH die Wiedererkennbarkeit als äußerste
Grenze des Vervielfältigungsrechts etabliert.¹⁸ Wird ein Schutzgegenstand oder
ein Teil davon¹⁹ technisch derart übernommen, dass er nicht mehr erkennbar
ist, ist die Festlegung in dieser Form keine Vervielfältigung des Schutzgegenstan-
des.²⁰ Diese aus dem allgemeinen Sprachgebrauch entnommene Erkenntnis²¹
des EuGH dürfte nicht auf den Vervielfältigungsbegriff im Bereich des Tonträ-
gerherstellerschutzes beschränkt sein, auf den sich die konkrete Entscheidung
bezog. Vielmehr sollte sie auch für andere Schutzgegenstände, für Werke und
Werkteile und insbesondere auch für die textbezogenen Leistungsschutzrechte,
gelten.²² Ergänzend zieht der EuGH grundrechtliche Ausgleichserwägungen

¹⁷ Vgl. *Schöch et al*, RuZ 2020, 160 unter einer CC BY-SA Lizenz; *Schöch et al*, Teil 3, C. in
diesem Band.

¹⁸ EuGH, Urt. v. 29.07.2019 – C-476/17, ECLI:EU:C:2019:624 = GRUR 2019, 929 Tz.
26 ff., 39 – *Pelham*; *Leistner*, GRUR 2019, 1008, 1010.

¹⁹ Auch Teilvervielfältigungen sind dem Rechteinhaber vorbehalten, wenn der Teil selbst
schon die Schutzvoraussetzungen erfüllt, Art. 2 RL 2001/29 („ganz oder teilweise“);
Dreier/Schulze-Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 16 UrhG Rn. 9.

²⁰ EuGH, Urt. v. 29.07.2019 – C-476/17, ECLI:EU:C:2019:624 = GRUR 2019, 929 Tz.
31 – *Pelham*.

²¹ EuGH, Urt. v. 29.07.2019 – C-476/17, ECLI:EU:C:2019:624 = GRUR 2019, 929 Tz.
37 – *Pelham*.

²² So im Ansatz wohl auch: *Grünberger*, ZUM 2020, 175, 184.

heran: Die Rechte des geistigen Eigentums werden nicht schrankenlos gewährt. Ändernde, nicht wiedererkennbare Nutzungen in Ausübung eines anderen Grundrechts²³ sind zur Herstellung eines angemessenen Gleichgewichts zwischen verschiedenen Grundrechtsinteressen hinzunehmen.²⁴ Auch diese Erwägung ist auf andere urheberrechtliche Schutzgegenstände übertragbar.

Nicht immer ist ein Text in Gänze urheberrechtlich geschützt. Häufig enthält ein Text lediglich Passagen, welche die Schutzkriterien des § 2 Abs. 2 UrhG erfüllen. Eine Vervielfältigung ist dann schon vom urheberrechtlichen Verbot des § 16 UrhG nicht mehr erfasst, wenn die neue Festlegung keine schutzfähigen Teile enthält bzw. diese nicht mehr erkennbar sind,²⁵ auch wenn der Text im Übrigen erkennbar bleibt; die Vervielfältigung ist dann urheberrechtlich irrelevant.²⁶ Das gilt in ähnlicher Weise für wissenschaftliche Ausgaben, sofern die auf der wissenschaftlich sichtenden Tätigkeit beruhenden Teile nicht erkennbar sind. Denn nur auf diese bezieht sich der Schutz.²⁷ Auch bei nach § 71 UrhG geschützten nachgelassenen Werken sind nur solche Vervielfältigungen dem Rechteinhaber vorbehalten, die Teile enthalten, die die Voraussetzungen von § 2 Abs. 2 UrhG erfüllen.²⁸ Der Schutz desjenigen, der ein Werk nach Ablauf des Urheberrechtsschutzes erscheinen lässt, kann nicht weiter gehen, als der Schutz, den der Urheber selbst erhalten hätte. Der Rechteinhaber wird durch § 71 UrhG hinsichtlich der Verwertungsrechte dem Urheber gleich,²⁹ jedoch nicht bessergestellt. Für Texte, die dem Presseverlegerleistungsschutzrecht unterfallen,³⁰ ist dagegen jede technische Kopie relevant, die sich nicht auf kleinste Teile beschränkt, da der Schutz nicht von weiteren den Text selbst betreffenden Kriterien abhängt. Mit den oben angeführten Argumenten (Sprachgebrauch

²³ Für die abgeleiteten Formate lässt sich die Forschungsfreiheit ins Feld führen.

²⁴ EuGH, Urt. v. 29.07.2019 – C-476/17, ECLI:EU:C:2019:624 = GRUR 2019, 929 Tz. 33 ff. – *Pelham*.

²⁵ Vgl. BGH Urt. v. 16.04.2015 – I ZR 225/12 = GRUR 2015, 1189 Tz. 109 – *Goldraper*.

²⁶ Schricker/Loewenheim-Loewenheim, Urheberrecht, 6. Aufl. 2020, § 16 UrhG Rn. 14; Dreier/Schulze-Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 16 UrhG Rn. 9.

²⁷ Schricker/Loewenheim-Loewenheim, Urheberrecht, 6. Aufl. 2020, § 70 UrhG Rn. 10.

²⁸ Dreier/Schulze-Dreier, Urheberrechtsgesetz, 6. Aufl. 2018, § 71 UrhG Rn. 10. Der Schutz bezieht sich nämlich nur auf Ausgaben mit Werkcharakter: Dreier/Schulze-Dreier, Urheberrechtsgesetz, 6. Aufl. 2018, § 71 UrhG Rn. 3.

²⁹ Schricker/Loewenheim-Loewenheim, Urheberrecht, 6. Aufl. 2020, § 71 UrhG Rn. 12; Wandtke/Bullinger-Thum, Urheberrecht, 5. Aufl. 2019, § 71 Rn 29.

³⁰ Die §§ 87f ff. UrhG sind zwar wegen unterbliebener Notifizierung nicht anwendbar, jedoch sind Nachfolgeregelungen bereits im Entstehen, siehe dazu: *Jotzo*, Teil 2, B. in diesem Band, 61, 73 und den Referentenentwurf des BMJV zur Umsetzung von Art. 15 UrhR-RL.

und grundrechtliches Gleichgewicht) muss aber die Wiedererkennbarkeit als äußerste Grenze bleiben.³¹

Obwohl also nicht jede technische Vervielfältigung, die Teile eines Textes noch erkennen lässt, in die Rechte am Primärtext eingreift, hilft dies für die Entwicklung abgeleiteter Formate wenig weiter. Denn die Problematik liegt in der Bestimmung der schutzfähigen Textteile und dem Schutzzumfang.³² Das gilt umso mehr, da die Erstellung großer Korpora und die Möglichkeit, verschiedene Korpora zusammenzufassen, einen Formatstandard erfordert, der auf unterschiedliche Textgattungen angewandt werden kann.³³ Da die automatisierte Bestimmung geschützter Teile, zumal bei einer Vielzahl unterschiedlicher Texte unterschiedlicher Textgattungen, kaum möglich ist, muss die automatisierte Erstellung ableitender Formate darauf abzielen, dass die Ausgangstexte im Endprodukt insgesamt nicht wiedererkennbar sind. Nur so ist sichergestellt, dass geschützte Interessen nicht beeinträchtigt werden und ein Konflikt mit dem Urheberrecht nicht entsteht.

Unter diesem Gesichtspunkt hilft auch § 24 UrhG und das durch die deutsche Rechtsprechung dazu entwickelte Kriterium des Verblassens der eigenschöpferischen Merkmale in der freien Benutzung gegenüber dem Originalwerk nicht weiter. Auch dieses erfordert nämlich eine individuelle Beurteilung.³⁴ Als Schrankenbestimmung kann § 24 UrhG nach der *Pelham*-Entscheidung des EuGH zudem keine Anwendung mehr finden, soweit er das Urheberrecht jenseits des Schrankenkatalogs von Art. 5 RL 2001/29/EG einschränkt.³⁵

³¹ Nach der Begründung des EuGH liegt nämlich dann keine Vervielfältigung vor.

³² Wie *Jotzo*, Teil 2, B. in diesem Band, 61, 68, 71f dargelegt, ist eine schematische Betrachtung nicht möglich, sondern es bedarf jeweils einer Gesamtbetrachtung unter Einbeziehung verschiedener Faktoren.

³³ Vgl. zum Standardisierungserfordernis: *Schöch et al.*, RuZ 2020, 160, 175 unter einer CC BY-SA Lizenz; sowie *Schöch et al.*, Teil 3, C. in diesem Band.

³⁴ Dreier/Schulze-Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 16 UrhG Rn. 8.

³⁵ *Leistner*, GRUR 2019, 1008; *Wagner*, MMR 2019, 727, 729; *Grünberger*, ZUM 2020, 175, 184; a.A.: *Schulze*, GRUR 2020, 128, 129, der die Ansicht vertritt, § 24 UrhG könne als Schutzbereichsbestimmung fortbestehen und auch herangezogen werden, um im Unionsrecht vorgesehene, jedoch nicht spezifisch umgesetzte Schrankenbestimmungen (z.B. die Parodie-schranke) anzuwenden. Der am 13.10.2020 veröffentlichte Referentenentwurf des BMJV für ein „Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes“ sieht (mit Verweis auf die *Pelham*-Entscheidung des EuGH) die Aufhebung von § 24 UrhG und eine neue explizite Schrankenregelung für Karikaturen, Parodien und Pastiches vor. Im Rahmen des neu formulierten § 23 UrhG soll die Schutzbereichsbeschränkung von § 24 UrhG fortbestehen (S. 82). Bei hinreichendem Abstand zum vorbestehenden Werk soll keine

Ganz vereinzelt stehenbleibende schutzfähige Textteile, die einem durchschnittlichen Betrachter nicht auffallen, können möglicherweise aufgrund ihrer untergeordneten Bedeutung im Korpus als unwesentliche Beiwerke i.S.v. § 57 UrhG angesehen werden.³⁶ Dennoch sollte das Ziel sein, dass schutzfähige Textteile möglichst nicht wiedererkennbar sind. § 57 UrhG und das unionsrechtliche Pendant des Art. 5 Abs. 3 lit. i InfoSocRL sind für Textwerke in der Rechtsprechung bisher nicht konturiert worden. Zudem stellen sich im Rahmen dieser Schrankenbestimmung schwierige Abgrenzungsfragen.³⁷ Die Schranke kann deshalb zwar im Einzelfall helfen, bietet aber kein sicheres Auffangnetz für vielfach in einem abgeleiteten Format noch enthaltene schutzfähige Textteile. Der EuGH hat zum Maßstab für die fehlende Wiedererkennbarkeit bisher keine Ausführungen gemacht, sodass die Konkretisierung zunächst weiteren Rechtsstreitigkeiten vorbehalten bleibt.³⁸ Im Detail wirft die Bestimmung des richtigen Maßstabes einige Schwierigkeiten auf.³⁹ Abzustellen ist in jedem Fall auf den Rezipienten. Darauf, ob der Rechteinhaber selbst seinen Schutzgegenstand noch als Ausgangspunkt des neuen Formats erkennt, kann es dagegen nicht ankommen. Allein dadurch werden nämlich seine geschützten Verwertungsinteressen und seine durch das Urheberrecht im Außenverhältnis⁴⁰ geschützte Beziehung zum Werk nicht beeinträchtigt. Der Werkgenuss wird nicht substituiert. Zudem ermöglicht nur das Abstellen auf die Rezipienten eine externe Über-

Bearbeitung vorliegen. Hinreichender Abstand sei gegeben, „wenn die aus einem vorbestehenden Werk entlehnten eigenpersönlichen Züge dem Gesamteindruck nach gegenüber der Eigenart des neuen Werkes so stark „verblassen“, dass das vorbestehende Werk nicht mehr oder nur noch rudimentär zu erkennen ist“.

³⁶ *Raue/Schöch*, Teil 3, A. in diesem Band, 129, 135 f. § 57 UrhG setzt Art. 5 Abs. 3 lit. i InfoSocRL um, der die beiläufige Einbeziehung eines Werks oder sonstigen Schutzgegenstands in anderes Material erlaubt. Entsprechend richtlinienkonform ist § 57 UrhG auszulegen.

³⁷ Z.B. die Bestimmung der Beiläufigkeit, nach deutschem Verständnis die Abgrenzung zwischen Haupt- und Beiwerk (dazu: BGH, Urt. v. 17.11.2014 – I ZR 177/13 = GRUR 2015, 667 – *Möbelkatalog*), und die Frage, wie die Schranke bei wiederholtem Auftreten einbezogener Werkteile anzuwenden ist.

³⁸ *Wagner*, MMR 2019, 727, 732; *Hieber*, ZUM 2019, 746, 747.

³⁹ Siehe mit ersten Maßstabsskizzen (jeweils im Zusammenhang mit Sampling) *Leistner*, GRUR 2019, 1008, 1010; *Apel*, MMR 2019, 601, 602; *Grünberger*, ZUM 2020, 175, 184. Man wird wahrscheinlich zwischen verschiedenen Werkarten und Leistungsschutzrechten differenzieren müssen.

⁴⁰ Das zeigt sich in den §§ 12 ff. UrhG, vgl. dazu: Schrickler/Loewenheim-Peukert, Urheberrecht, 6. Aufl. 20120, § 14 UrhG Rn. 25 (Interessenbeeinträchtigung durch Entstellung i.d.R. nur bei Öffentlichkeit); § 13 UrhG sichert die „nach außen dokumentierte Urheberschaft“, vgl.: Schrickler/Loewenheim-Peukert, Urheberrecht, 6. Aufl. 2020, § 13 UrhG Rn. 1; § 12 UrhG sichert dem Urheber das Veröffentlichungsrecht.

prüfbarkeit. Daher ist darauf abzustellen, ob ein durchschnittlicher, nicht vorinformierter⁴¹ Leser, bei direktem Vergleich, den geschützten Text(teil) im abgeleiteten Format noch erkennt.⁴²

Für bestimmte Forschungsfragen ist die Beziehung von Worten zueinander von Bedeutung. In einem abgeleiteten Format kann diese Information durch die Positionsangabe eines Wortes im Text oder schlicht durch den Erhalt der Wortsequenz innerhalb eines Textfragmentes erhalten werden.⁴³ Das Format erlaubt so unmittelbar nicht den Werkgenuss. Allerdings ermöglichen die Sequenzinformationen – abhängig von den weiteren vorhandenen Informationen – die mehr oder weniger vollständige Rekonstruktion des Primärtextes. Damit ist die Frage aufgeworfen, welche Bedeutung die Rekonstruierbarkeit für den Vervielfältigungsbegriff hat.

Die Rekonstruktion eines zerstörten oder fragmentierten⁴⁴ Werkes stellt eine relevante Vervielfältigungshandlung dar.⁴⁵ Die Rekonstruktion des geschützten Primärtextes aus einem abgeleiteten Format ist danach eine zustimmungsbedürftige Vervielfältigungshandlung. Aber wie ist ein Format zu bewerten, das die Rekonstruktion ermöglicht, also die bloße Rekonstruierbarkeit?

Zweifellos kann die Tatsache der Rekonstruierbarkeit des geschützten Ursprungstextes aus einem abgeleiteten Format nicht grundsätzlich bedeutungslos sein. So wird man ein Motivpuzzle – ungeachtet der Frage, ob Einzelteile Teilervielfältigungen sind – als Vervielfältigung des Motivs ansehen, auch wenn die Puzzleteile ohne Motivvorlage durcheinander gewürfelt sind, sich aber ohne weiteres wieder zu dem Motiv zusammensetzen lassen. Bei der technisch bedingten fragmentierten Übertragung eines Werkes (Routing, z.B. zwecks

⁴¹ Denn bei entsprechender Vorinformation stellt sich das Erkennen zwangsläufig ein: *Grünberger*, ZUM 2020, 175, 182 zum Sampling.

⁴² Vgl. *Leistner*, GRUR 2019, 1008, 1010; *Grünberger*, ZUM 2020, 175, 182 jeweils zum Sampling.

⁴³ Zur Bedeutung der Sequenzinformation sowie zu verschiedenen Formaten mit und ohne Sequenzinformation: *Schöch et al.*, RuZ 2020, 160 unter einer CC BY-SA Lizenz, passim.

⁴⁴ Werden Werke, bedingt durch Übertragungstechnik, in kleine Fragmente aufgeteilt, so gesendet und beim Empfänger wieder zusammengesetzt, kommt es nach h.M. und der Rechtsprechung des EuGH jedenfalls beim Empfänger zu einer relevanten Vervielfältigungshandlung, wenn die gleichzeitig zusammengesetzten Fragmente im Empfangsgerät einen Werkteil bilden, der die Schutzvoraussetzungen erfüllt: EuGH, Urt. v. 04.10.2011 – C-403/08 und C-429/08, ECLI:EU:C:2011:631 = MMR 2011, 817 – Rn. 153 ff., insb. Rn. 157 – *FAPL/Murphy*; Dreier/Schulze-Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 16 UrhG Rn. 12; *Busch*, GRUR 2011, 496, 498.

⁴⁵ Schrickler/Loewenheim-Loewenheim, Urheberrecht, 6. Aufl. 2020, § 16 UrhG Rn. 15 mwN; Dreier/Schulze-Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 16 UrhG Rn. 11, 12.

Streamings) ist umstritten, ob es während der Übertragung zu relevanten Vervielfältigungshandlungen kommt. Der Streit knüpft an die Frage, ob dabei die einzelnen Fragmente maßgeblich sind, die meist so klein sind, dass sie die Schutzvoraussetzungen nicht erfüllen,⁴⁶ oder ob die allein aus technischen Gründen einzeln versandten Fragmente gemeinsam betrachtet werden sollten.⁴⁷ Allerdings unterscheidet sich die Sach- und Interessenlage bei den abgeleiteten Formaten deutlich von den genannten Beispielen. Während bei Puzzleteilen und Streaming-Fragmenten die Rekonstruktion des Ganzen und der anschließende Werkgenuss beabsichtigt ist und deshalb die Verwertungsinteressen der Rechteinhaber durch die Nutzung berührt werden, dienen Sequenzinformationen in einem abgeleiteten Format der Ermöglichung und Verifizierbarkeit wissenschaftlicher Erkenntnisse, jedoch nicht der Rekonstruktion. Die Rekonstruktion dürfte für den durchschnittlichen Werkrezipienten zudem sehr aufwendig sein. Voraussetzung wäre hier natürlich, dass evtl. zur Rekonstruktion zur Verfügung stehende Mittel (Algorithmen, Word Embedding Modelle⁴⁸ o.Ä.) nicht zur Verfügung gestellt, insbesondere nicht veröffentlicht werden. Die Verwertungsinteressen des Rechteinhabers werden dann allenfalls marginal berührt. Dem stehen Forschungsinteresse der Nutzer und grundsätzlich auch das gesellschaftliche Allgemeininteresse an den Forschungsergebnissen gegenüber, die in diesem Fall in aller Regel überwiegen. Lässt man bei der Bestimmung der Verwertungshandlung eine Einzelfallbetrachtung⁴⁹ zu,⁵⁰ lässt es sich begründen, in einem abgeleiteten Format, trotz theoretischer Rekonstruierbarkeit, keine urheberrechtlich relevante Vervielfältigung zu sehen. Der Rechtsprechung des EuGH ist es nicht fremd, Interessenabwägungen in die Auslegung von Verwertungsrechten einfließen zu lassen.⁵¹

⁴⁶ Hoeren/Sieber/Holznapel-*Ernst*, Multimedia-Recht, 51. EL Stand: Februar 2020, Teil 7.1 Rn. 61.

⁴⁷ So Dreier/Schulze-*Dreier*, Urheberrechtsgesetz, 6. Aufl. 2018, § 16 UrhG Rn. 12; *Busch*, GRUR 2011, 496, 499.

⁴⁸ Siehe unten e).

⁴⁹ Eine Beurteilung anhand folgender Kriterien erscheint sinnvoll: Ist die Rekonstruktion beabsichtigt? Soll Werkgenuss ermöglicht werden? Beeinträchtigt die Rekonstruktionsmöglichkeit die geschützten Verwertungsinteressen am Werk? Wie hoch ist der Aufwand für den durchschnittlichen Werkrezipienten? Wie stehen die betroffenen Interessen zueinander (Interessenabwägung)?

⁵⁰ Mit einem ähnlichen Ansatz eines „subjektiven Vervielfältigungsrechts“, der auf die berechnete und sozialadäquate Nutzersicht bei Internetnutzungen abstellt: *Raue*, ZGE 2017, 514.

⁵¹ Z.B. in: EuGH, Urt. v. 29.07.2019 – C-476/17, ECLI:EU:C:2019:624 = GRUR 2019, 929 Tz. 32 ff. – *Pelham*; auch die Rechtsprechung zum Recht der öffentlichen Wiedergabe ist

Letztlich handelt es sich hierbei jedoch um eine bisher nicht entschiedene Rechts- und Wertungsfrage. Zu raten ist deshalb, abgeleitete Formate so zu gestalten, dass eine Rekonstruierbarkeit möglichst ausgeschlossen, jedenfalls aber die Chance einer qualitativ guten Rekonstruktion minimiert ist.

2. Beispiele⁵² und urheberrechtliche Bewertung abgeleiteter Formate

a) Term-Dokument-Matrizen⁵³

In einfachen und segmentbasierten Term-Dokument-Matrizen werden die in einem Text bzw. in einem nicht zu kleinen Textsegment vorkommenden Wörter aufgelistet und ihre Häufigkeit angegeben. Als reine Worthäufigkeitsstatistiken sind sie keine Vervielfältigungen der Primärtexte. Sie enthalten nur noch die nicht geschützte Information, welche Worte mit welcher Häufigkeit in einem Text oder Textsegment vorkommen. Der Primärtext, sofern er nicht sehr kurz ist⁵⁴ und vorausgesetzt, die Segmentlänge ist ebenfalls nicht zu kurz, wird im Regelfall in diesem Format insgesamt nicht mehr erkennbar sein.

b) Lemmatisierung und Wort(art)selektion⁵⁵

Ein Format, in dem lediglich alle Worte eines Textes unter Beibehaltung der Wortreihenfolge durch Lemmata⁵⁶ ersetzt werden, dürfte dagegen eine bearbeitende Vervielfältigung sein. Insbesondere geschützte Handlungsstränge können in einem solchen Format erkennbar bleiben. Es steht damit für eine erlaubnisfreie Nutzung nicht zur Verfügung.

von solchen Erwägungen geprägt, siehe z.B.: EuGH, Urt. v. 8.9.2016 – C-160/15 = GRUR 2016, 1152 – *GS Media*; *Raue*, ZGE 2017, 514, 518 ff.; *Leistner*, GRUR 2019, 1008, 1009 f. sieht in der Rechtsprechung jedoch einen Trend, die Verwertungsrechte – jenseits der öffentlichen Wiedergabe und nur mittelbarer Nutzungshandlungen – strikt und objektiv zu definieren, und die Schaffung eines angemessenen Interessenausgleichs in den Schrankenbestimmungen zu verorten.

⁵² Für eine Darstellung der hier diskutierten abgeleiteten Formate und deren praktische Verwendungsmöglichkeiten siehe: *Schöbch et al*, RuZ 2020, 160, 163 ff. unter einer CC BY-SA Lizenz.

⁵³ *Schöbch et al*, RuZ 2020, 160, 163 ff. unter einer CC BY-SA Lizenz.

⁵⁴ Bei sehr kurzen Texten könnte sich aufgrund zwingender grammatikalischer Regelungen aus den wenigen aufgeführten Worten der Texte rekonstruieren lassen.

⁵⁵ Vgl. *Schöbch et al*, RuZ 2020, 160, 164, 167 f. unter einer CC BY-SA Lizenz.

⁵⁶ Lemma ist die Grund- oder Nennform eines Wortes, „Lemma (Lexikographie)“, in: Wikipedia, Die freie Enzyklopädie, Stand 28. April 2019, 11:23 UTC, abrufbar unter: [https://de.wikipedia.org/w/index.php?title=Lemma_\(Lexikographie\)&oldid=187995927](https://de.wikipedia.org/w/index.php?title=Lemma_(Lexikographie)&oldid=187995927), zuletzt abgerufen am 19.10.2020.

Das Löschen von Stopp- oder Funktionswörtern kann, muss aber nicht dazu führen, dass ein geschützter Text nicht wiedererkennbar ist. Stopp- oder Funktionswörter sind solche Wörter bzw. Wortarten, die in einer Sprache häufig vorkommen, eher grammatikalische Funktionen haben und die daher für die Erfassung des Inhalts eines Textes wenig relevant sind, z.B. Artikel, Konjunktionen, Pronomen, Präpositionen, Modal- und Hilfsverben.⁵⁷ Ob ein Text nicht wiedererkennbar ist, wird wesentlich davon abhängen, welche Wortarten und wie viele Wörter herausselektiert werden. In vielen Fällen wird die urheberrechtlich konkret geschützte Ausdrucksform durch das Beseitigen mehrerer Wortarten hinreichend beseitigt werden. Eine pauschale Beurteilung ist jedoch nicht möglich, sodass diese Formatierung bei Beibehaltung der Wortsequenz im Einzelfall noch eine Vervielfältigung eines geschützten Textes darstellen könnte.

c) Formate mit gestörter Sequenzinformation⁵⁸

Werden in einem abgeleiteten Format die Wörter eines Textes willkürlich durcheinander gewürfelt, sodass Satzstrukturen und Inhalt aufgelöst werden, wird das Format in aller Regel keine Schutzrechte tangieren.⁵⁹ Reine Informationen, Daten und einzelne Wörter unterliegen nicht dem Urheberrechtsschutz.⁶⁰ Wohl aber können bereits wenige aufeinander folgende Worte, wie etwa Satzteile, ausnahmsweise geschützt sein, wenn dieses Textfragment bereits selbst die Schutzvoraussetzungen erfüllt.⁶¹ Vorsicht ist auch bei Presserzeugnissen mit

⁵⁷ Vgl. „Stoppwort“, in: Wikipedia, Die freie Enzyklopädie, Stand: 10. April 2019, 15:03 UTC, <https://de.wikipedia.org/w/index.php?title=Stoppwort&oldid=187442343>, zuletzt abgerufen am 19.10.2020; „Synsemantikum“, in: Wikipedia, Die freie Enzyklopädie, Stand: 25.02.2019, 09:48 UTC, abrufbar unter: <https://de.wikipedia.org/w/index.php?title=Synsemantikum&oldid=186016518>, zuletzt abgerufen am: 19.10.2020.

⁵⁸ *Schöch et al.*, RuZ 2020, 160, 165 f. unter einer CC BY-SA Lizenz.

⁵⁹ Das gilt jedenfalls wiederum dann, wenn es sich nicht um einen sehr kurzen Text handelt. Bei Texten bestehend aus wenigen Worten könnte die Möglichkeit bestehen, die Worte nach grammatikalischen Regeln wieder zu ordnen und damit den Text zu rekonstruieren. Allerdings sind solch kurze Texte nur in Ausnahmefällen geschützt, siehe: *Jotzo*, Teil 2, B. in diesem Band, 61, 66 f., 71 f.

⁶⁰ Siehe erläuternd: *Jotzo*, Teil 2, B. in diesem Band, 61, 66f.

⁶¹ EuGH, Urt. v. 16.07.2009 – C-5/08, ECLI:EU:C:2009:465 = GRUR 2009, 1041 Tz. 48 – *Infopaq*.

Blick auf das Presseverlegerleistungsschutzrecht⁶² geboten, wenn größere Satzfragmente erhalten bleiben.⁶³ Dass lediglich die Satzreihenfolge eines Textes durchmischt wird, führt somit zu keinem geeigneten Format. Ob das Durcheinanderwerfen von Worten oder Satzfragmenten zu einem geeigneten abgeleiteten Format führt, das keine Vervielfältigungen des Primärtextes enthält, hängt wesentlich von der Fragmentlänge ab. Zudem ist die Größe eines Segments, innerhalb dessen die Mischung vorgenommen wird (innerhalb eines Satzes, innerhalb je 100 Worten, innerhalb eines Absatzes, innerhalb eines Kapitels etc.), von Bedeutung. Je größer das Segment, desto eher werden auch geschützte Handlungen oder Figurenbeschreibungen⁶⁴ mit Sicherheit unkenntlich und desto schwieriger wird die Rekonstruktion bzw. sie wird gar unmöglich.

d) N-Gramm-Tabellen⁶⁵

Ein übliches Vorgehen in der Korpuslinguistik besteht darin, Texte in Fragmente zu zerlegen (sog. N-Gramme), z.B. drei aufeinander folgende Wörter (3-grams) und mit Metadaten anzureichern. Ob ein solches Format als Vervielfältigung des Primärtextes weiter den Ausschließlichkeitsrechten der Rechteinhaber unterliegt, hängt von weiteren Faktoren ab. Zunächst kann nur bei sehr kurzen N-Grammen ausgeschlossen werden, dass diese nicht als Werkteile bereits dem Schutz unterliegen. Insofern gilt das zuvor Gesagte. Das Auslassen von Stoppwörtern kann auch hier helfen. Werden N-Gramme allerdings in ihrer im Text enthaltenen Reihenfolge aufgeführt, bleibt das Ursprungswerk darin leicht erkennbar und bei Ausblendung der Annotationsinformationen auch lesbar. Zudem ermöglicht das Überlappen von N-Grammen die Rekonstruierbarkeit. Dieses Problem wird entschärft, wenn nicht die N-Gramme für einen einzelnen Text, sondern für eine ganze Textsammlung zusammengefasst werden.

⁶² Die §§ 87f ff. UrhG sind zwar wegen unterbliebener Notifizierung nicht anwendbar, jedoch sind Nachfolgeregelungen bereits im Entstehen, siehe dazu: *Jotzo*, Teil 2, B. in diesem Band, 61, 72 f. und den Referentenentwurf des BMJV vom 13.10.2020 zur Umsetzung von Art. 15 UrhR-RL.

⁶³ Aus dem Schutz ausgenommen sind nach Art. 15 Abs. 1 S. 4 UrhR-RL jedoch einzelne Wörter und sehr kurze Auszüge.

⁶⁴ Zum Schutz literarischer Figuren: *Jotzo*, Teil 2, B. in diesem Band, 61, 69 f., mwN.

⁶⁵ Zur Verwendung von N-Grammen: *Schöch et al.*, RuZ 2020, 160, 169 ff. unter einer CC BY-SA Lizenz.

e) Formate aus Wort-Vektoren auf Grundlage eines Word Embedding Modells⁶⁶

Eine weitere Möglichkeit besteht darin, die Worte der Primärtexte in Zahlensequenzen (Vektoren) zu übersetzen, welche die Worte und ihre Umgebung in einem multidimensionalen Raum beschreiben. Dazu muss zunächst textunabhängig ein Wortembedding-Modell erstellt werden. Nach der Vektorisierung kann man Worte nicht mehr direkt erkennen, wohl aber Ähnlichkeiten zwischen den Worten an der Anordnung im Vektorraum. Der Vorteil dieses Modells besteht darin, dass sehr viele Informationen in die Vektorberechnung einfließen und damit erhalten werden können, ohne dass das Produkt überhaupt noch ein lesbarer Text ist. Allerdings kann, sofern das genutzte Wortembedding-Modell zur Verfügung steht, der Text algorithmisch weitgehend wiederhergestellt werden. Insofern ist bei der Erstellung und Zugänglichmachung dieser Formate mit Blick auf die beschriebene Rechtsunsicherheit⁶⁷ Vorsicht geboten.

In Erweiterung dieses Modells kann in die Vektorberechnung der Kontext einbezogen werden, was dazu führt, dass jede Worteinheit einen individuellen Vektor erhält. Als Folge kann ein Wort nicht mehr einfach zurückübersetzt werden. Die Ursprungstexte sind in einem solchen Format nicht mehr erkennbar und aus dem Format auch praktisch nicht rekonstruierbar. Der Verwendung eines solchen Formates stünde das Urheberrecht nicht entgegen.

IV. Rechtliche Anforderungen und Bewertung der Erstellung abgeleiteter Formate

1. Vorbedingungen der Formaterstellung

Voraussetzung für die rechtmäßige Erstellung abgeleiteter Formate ist zunächst, dass rechtmäßiger Zugang⁶⁸ zu den Primärtexten besteht. Sofern geschützte Texte erst auf einem Server gesammelt und damit nicht nur vorübergehend⁶⁹ digital vervielfältigt werden, setzt dies die Zustimmung der

⁶⁶ *Schöch et al*, RuZ 2020, 160, 171 ff. unter einer CC BY-SA Lizenz.

⁶⁷ Siehe oben III.1. a.E.

⁶⁸ Siehe dazu: ErwGr 14 und ErwGr 18 Abs. 2 S. 1 UrhR-RL.

⁶⁹ Vorübergehend i.S.d. § 44a UrhG, siehe dazu sogleich 2. Vgl. auch ErwGr 9 UrhR-RL.

Rechteinhaber oder das Vorliegen einer gesetzlichen Schrankenbestimmung voraus. Als Schrankenbestimmungen kommen insbesondere § 60d⁷⁰ oder § 60c⁷¹ UrhG in Betracht. Zu beachten ist dabei, dass sich auf diese Vorschriften nur berufen kann, wer selbst Textmining bzw. nicht kommerzielle wissenschaftliche Forschung betreiben will.⁷² Die Vervielfältigungen dürfen allerdings durch Dritte im Auftrag und nach Anweisung des Berechtigten vorgenommen werden.⁷³ Weiter ist zu beachten, dass die Schrankenregelungen voraussetzen, dass rechtmäßiger Zugang zu den Schutzgegenständen gegeben ist, diese Schranken selbst aber keinen Anspruch auf Zugang begründen.⁷⁴ Wirksame technische Schutzmaßnahmen dürfen nicht eigenmächtig überwunden werden, § 95a Abs. 1 UrhG.⁷⁵ Rechtmäßiger Zugang besteht zu Schutzgegenständen, die sich bereits im Besitz des Nutzers befinden, die ihm zur Verfügung gestellt werden oder die im Internet frei abrufbar sind.⁷⁶

⁷⁰ Zukünftig wohl auch § 44b UrhG, vgl. den Referentenentwurf des BMJV zur Umsetzung von Art. 3 und Art. 4 UrhR-RL; zum vorangehenden Diskussionsentwurf: *Raue*, ZUM 2020, 172.

⁷¹ Dabei sind die quantitativen Einschränkungen der Regelung zu beachten.

⁷² Amtliche Begründung, BT-Drs. 18/12329, 41.

⁷³ Die Regierungsbegründung nennt beispielhaft Vervielfältigungen durch Mitarbeiter einer Bibliothek, BT-Drs. 18/12329, 41. Darüber hinaus können aber auch IT-Abteilungen und kommerziell handelnde Dritte mit der Vornahme der Vervielfältigungen beauftragt werden, vgl.: Spindler/Schuster-Anton, Recht der elektronischen Medien, 4. Aufl. 2019, § 60d UrhG Rn. 6; *Raue*, CR 2017, 656, 657. Denn nur derjenige, der die Forschung selbst vornimmt muss (de lege lata noch) zu nicht kommerziellen Zwecken handeln, siehe: Dreier/Schulze-Dreier, Urheberrechtsgesetz, 6. Aufl. 2018, § 60d UrhG Rn. 5. In der Gesetzesbegründung zu § 60c UrhG (Wissenschaftsschranke) heißt es, „(d)ie Nutzungshandlungen dürfen auch durch einen Dritten vorgenommen werden, der selbst keine Forschungszwecke verfolgt“, BT-Drs. 18/12329, 39. Sodann wird darauf verwiesen, dass das auch im Rahmen der Vorgängernorm § 53 Abs. 2 S. 1 Nr. 1 UrhG zulässig war. Insofern soll also an der bisherigen Rechtslage festgehalten werden. Während für § 53 Abs. 1 UrhG gilt, dass die Vervielfältigung durch Dritte grundsätzlich nur unentgeltlich erfolgen darf, galt diese Einschränkung für § 53 Abs. 2 UrhG nicht. Auch das legt nahe, dass ein Dritter, der für einen nach § 60d UrhG berechtigten Nutzer handelt, dies auch gewerblich tun darf.

⁷⁴ Siehe dazu *Raue*, CR 2017, 656, 658.

⁷⁵ Das gilt selbst dann, wenn eine gesetzliche Schrankenregelung die Nutzung erlaubt. Es besteht dann lediglich ein Anspruch auf Ermöglichung der erlaubten Nutzung, § 95b Abs. 2 UrhG.

⁷⁶ *Raue*, CR 2017, 656, 658.

Sofern Texte in Datenbanken eingegliedert sind, ist gegebenenfalls der sui generis Schutz des Datenbankherstellers (§ 87b Abs. 1 UrhG) zu beachten.⁷⁷ Das Vervielfältigen wesentlicher Teile einer Datenbank – und dem gleichgestellt die wiederholte und systematische Vervielfältigung unwesentlicher Teile – ist dem Datenbankhersteller vorbehalten, es bedarf deshalb dessen oder einer gesetzlichen Erlaubnis. § 87c Abs. 1 Nr. 2 UrhG verweist insofern auf die §§ 60c und 60d UrhG.⁷⁸

Verfügt der Formatersteller bereits rechtmäßig über digitale Kopien (z.B. eine Bibliothek gestützt auf § 60e UrhG), können diese zur Formaterstellung verwendet werden.

2. Freistellung der Erstellung abgeleiteter Formate durch § 44a UrhG

Wenn ein abgeleitetes Format keine Vervielfältigung des Primärtextes darstellt, also den Schutz des Textes nicht berührt, ist seine Nutzung und Verwertung unbeschränkt zulässig. Im Rahmen der automatisierten Generierung der Formate, im Ableitungsprozess, wird es jedoch in der Regel noch zu urheberrechtlich relevanten Vervielfältigungen der Texte oder geschützter Textteile im maschinellen Arbeitsspeicher kommen. Diese Vervielfältigungen bedürfen wiederum einer (gesetzlichen) Erlaubnis.⁷⁹ § 44a UrhG⁸⁰ stellt (1) vorübergehende Vervielfältigungen frei, die (2) „flüchtig oder begleitend sind und (3) einen integralen und wesentlichen Teil eines technischen Verfahrens darstellen und deren alleiniger Zweck es ist, [...] (4) eine rechtmäßige Nutzung eines Werkes oder sonstigen Schutzgegenstands zu ermöglichen, und (5) die keine eigenständige wirtschaftliche Bedeutung haben.“⁸¹

Um die Voraussetzungen dieser Schrankenregelung zu erfüllen, muss der Ableitungsprozess, im Rahmen dessen die Vervielfältigungen aus rein technischen Gründen zwangsweise anfallen, so ausgestaltet sein, dass die Vervielfältigungen

⁷⁷ Dreier/Schulze-Dreier, Urheberrechtsgesetz, 6. Aufl. 2018, § 87b UrhG Rn. 4; Jotzo, Teil 2, B. in diesem Band, 61, 75 f.; Raue, ZUM 2019, 684, 685. In ein mögliches Urheberrecht an Datenbank- und Sammelwerken (§ 4 UrhG) wird dagegen durch die Entnahme von Texten aus der Datenbank in der Regel nicht eingegriffen, siehe: Jotzo, Teil 2, B. in diesem Band, 61, 75; Spindler GRUR 2016, 1112, 1114.

⁷⁸ Dazu: Raue, CR 2017, 656, 660.

⁷⁹ Da die Formaterstellung nicht unmittelbar dem TDM für ein bestimmtes Projekt dient, sondern gerade die erweiterte Nutzung auch außerhalb der TDM-Schranke ermöglichen soll, ist zweifelhaft, ob sich diese Vervielfältigungen innerhalb des Erstellungsprozesses auf § 60d der § 60c UrhG stützen lassen.

⁸⁰ § 44a UrhG setzt die zwingende Schranke des Art. 5 Abs. 1 RL 2001/29/EG um.

⁸¹ § 44a UrhG.

nach einer nicht ins Gewicht fallenden Zeit aus dem Arbeitsspeicher automatisch – ohne menschliches Eingreifen – wieder gelöscht werden.⁸² Sie dürfen nur im Rahmen des technischen Verfahrens anfallen und keinen darüberhinausgehenden eigenständigen Zweck erfüllen.⁸³

Die so vorkommenden vorübergehenden Vervielfältigungen dienen im hier behandelten Sachverhalt dem alleinigen Zweck, eine rechtmäßige Nutzung zu ermöglichen.⁸⁴ Rechtmäßige Nutzungen sind nicht nur solche, die durch den Rechteinhaber oder das Gesetz explizit erlaubt werden, sondern insbesondere auch solche, die durch Gesetz schon nicht beschränkt sind.⁸⁵

Die vorübergehenden Vervielfältigungen im Prozess der Formatgenerierung dürfen keine eigenständige wirtschaftliche Bedeutung haben. Nach der Rechtsprechung des EuGH fehlt eine eigene wirtschaftliche Bedeutung, wenn der sich aus der vorübergehenden Vervielfältigung ergebende wirtschaftliche Vorteil nicht von dem wirtschaftlichen Vorteil aus der rechtmäßigen Nutzung des Schutzgegenstandes zu unterscheiden oder zu trennen ist und kein zusätzlicher wirtschaftlicher Vorteil erzeugt wird, der über denjenigen hinausgeht, der sich aus dieser Nutzung des geschützten Werkes ergibt.⁸⁶ Eine eigenständige wirtschaftliche Bedeutung besteht also, wenn „aus der wirtschaftlichen Verwertung der vorübergehenden Vervielfältigungen selbst Gewinne erzielt“ werden können.⁸⁷ Auf den wirtschaftlichen Wert eines möglichen, aus dem technischen

⁸² Dreier/Schulze-Dreier, Urheberrechtsgesetz, 6. Aufl. 2018, § 44a UrhG Rn. 4 f.; vgl. auch EuGH, Urt. v. 05.06.2014 – C-360/13, ECLI:EU:C:2014:1195 = GRUR 2014, 654 Rn. 43 – *PRCA/NLA*.

⁸³ Dreier/Schulze-Dreier, Urheberrechtsgesetz, 6. Aufl. 2018, § 44a UrhG Rn. 6; siehe insb. auch EuGH, Urt. v. 05.06.2014 – C-360/13, ECLI:EU:C:2014:1195 = GRUR 2014, 654 Rn. 28 – *PRCA/NLA*; EuGH, Beschl. v. 17.01.2012 – C-302/10, ECLI:EU:C:2012:16 = MMR 2013, 45 Tz. 30 – *Infopaq II*; EUGH, Urt. v. 16.07.2009 – C-5/08, ECLI:EU:C:2009:465 = GRUR 2009, 1041 Tz. 61 – *Infopaq*.

⁸⁴ Vgl. den Sachverhalt: EuGH, Beschl. v. 17.01.2012 – C-302/10, ECLI:EU:C:2012:16 = MMR 2013, 45 Tz. 40 ff. – *Infopaq II*. Vorübergehende Vervielfältigungshandlungen im Rahmen der Erstellung von Zusammenfassungen geschützter Texte sind nach § 44a UrhG zulässig. Die Zusammenfassungen greifen nicht in den Schutzbereich des Textes ein.

⁸⁵ ErwGr 33 RL 2001/29/EG: „Eine Nutzung sollte als rechtmäßig gelten, soweit sie vom Rechtsinhaber zugelassen bzw. nicht durch Gesetze beschränkt ist.“ S.a. EuGH, Beschl. v. 17.01.2012 – C-302/10, ECLI:EU:C:2012:16 = MMR 2013, 45 Tz. 40 ff. – *Infopaq II*; EuGH, Urt. v. 04.10.2011 – C-403/08 und C-429/08, ECLI:EU:C:2011:631 = MMR 2011, 817 – Rn. 153 ff., insb. Rn. 171 – *FAPL/Murphy*, m.Anm. Stieper; Dreier/Schulze-Dreier, Urheberrechtsgesetz, 6. Aufl. 2018, § 44a UrhG Rn. 8.

⁸⁶ EuGH, Beschl. v. 17.01.2012 – C-302/10, ECLI:EU:C:2012:16 = MMR 2013, 45 Tz. 50 – *Infopaq II*.

⁸⁷ EuGH, Beschl. v. 17.01.2012 – C-302/10, ECLI:EU:C:2012:16 = MMR 2013, 45 Tz. 52 – *Infopaq II*.

Prozess hervorgehenden Produkts kommt es dagegen nicht an. Eine eigenständige wirtschaftliche Bedeutung sollen vorübergehende Vervielfältigungen allerdings auch haben, wenn sie „zu einer Änderung des vervielfältigten Objekts, wie es zum Zeitpunkt der Einleitung des betreffenden technischen Verfahrens besteht, führen, da diese Handlungen dann nicht mehr darauf abzielen, seine Nutzung zu vereinfachen, sondern die Nutzung eines anderen Objekts.“⁸⁸ Dieses Kriterium des EuGH bleibt vage. Jedenfalls hat aber der EuGH in den *Infopaq*-Entscheidungen das Erstellen und kommerzielle Verwenden aus dem Schutzbereich herausfallender Textzusammenfassungen nicht wegen Änderung des Ursprungstextes und dadurch begründeter eigenständiger wirtschaftlicher Bedeutung für grundsätzlich unzulässig erklärt. Auch in der Sache *Pelham* wurde die verändernde Übernahme einer Tonsequenz nicht als problematisch angesehen.⁸⁹ Im Übrigen muss gelten, dass das Urheberrecht nicht über seinen Schutzbereich hinaus ausgedehnt werden darf und die Nutzung eines den Schutzbereich nicht berührenden Textformates, sowie dessen Erstellung, zulässig sein muss. Eine Änderung im urheberrechtlichen Sinne ist die hier gegenständliche Formatgenerierung nicht.⁹⁰

Bei entsprechender Verfahrensgestaltung dürfte die Erstellung abgeleiteter Formate nach § 44a UrhG freigestellt sein. Die Schranke des § 44a UrhG gilt auch für nach den §§ 70, 71, 87f UrhG geschützte Texte.⁹¹

3. Urheberpersönlichkeitsrechtliche Aspekte

Man mag die Frage aufwerfen, ob die abgeleiteten Formate geschützte Werke entstellen. Diese Frage stellt sich zunächst für den Fall, dass das Werk in dem abgeleiteten Format gar nicht mehr erkennbar ist, dann aber besonders auch für den Fall, dass das Gesamtwerk noch erkennbar ist und nur die eigenschöpferischen Elemente ausgespart oder nicht wiedererkennbar sind. Entstellungen und andere Beeinträchtigungen kann der Urheber nach § 14 UrhG verbieten, wenn sie geeignet sind, „seine berechtigten geistigen oder persönlichen Interessen am

⁸⁸ EuGH, Beschl. v. 17.01.2012 – C-302/10, ECLI:EU:C:2012:16 = MMR 2013, 45 Tz. 53 – *Infopaq II*.

⁸⁹ Das mag allerdings auch an der konkreten Fragestellung der Vorlagefrage gelegen haben, sowie daran, dass es um das Tonträgerherstellungsrecht und nicht das Urheberrecht ging.

⁹⁰ Wenn der EuGH von einer Änderung des „Objekts“ spricht, kann er nicht den Schutzgegenstand selbst meinen, vgl.: Wandtke/Bullinger-*Leenen*, Urheberrecht, 5. Aufl. 2019, Art. 5 InfoSocRL, Rn. 25.

⁹¹ Siehe §§ 70 Abs. 1, 71 Abs. 1 S. 3, 87g Abs. 4 S. 2 UrhG.

Werk zu gefährden“. § 14 UrhG schützt das geistige und persönliche Interesse des Urhebers, dass sein Werk dem Publikum unverfälscht dargeboten wird.

Ist das Werk im abgeleiteten Format nicht erkennbar, tangiert das abgeleitete Format das Urheberrecht am Primärtext nicht. Das Werk ist dann im Ergebnis des Ableitungsprozesses gerade nicht mehr vorhanden und das Ergebnis damit auch nicht nur eine (entstellende) Änderung. Das Werk kommt vielmehr im Produkt gar nicht mehr zum Ausdruck. Das Bestands- und Integritätsinteresse des Urhebers an seinem Primärtext wird nicht verletzt.⁹² Aber auch wenn nicht schutzfähige Teile des Originaltextes noch erkennbar sind, kann das keine die geschützten Interessen des Urhebers gefährdende Entstellung begründen. Der Schutz des Urheberpersönlichkeitsrechts ist durch den besonderen Schutzgegenstand gerechtfertigt, nämlich die persönliche geistige Schöpfung, die im Werk ihren Ausdruck findet, in der damit auch die Persönlichkeit des Autors und seine kreativen Entscheidungen zum Ausdruck kommen.⁹³ Wo dieser Schutzgegenstand gar nicht betroffen ist, kommt eine Verletzung des Urheberpersönlichkeitsrechts nicht in Betracht. Das Werk wird, in dem es für die wissenschaftliche Forschung in einem abgeleiteten Format aufbereitet wird, auch in keinen negativen Kontext gestellt.

Die Angabe, auf welches Werk welchen Autors eine Formatdatei zurückgeht, gefährdet geschützte Interessen der Urheber ebenfalls nicht. Diese Angabe teilt lediglich mit, aus welchem Text die enthaltenen Informationen, Textfragmente etc. stammen. Sie erweckt aber nicht den Eindruck, der Text in der vorliegenden Form entspringe der Feder des Originalautors. Die Beziehung des Urhebers zu seinem Werk⁹⁴ wird dadurch – auch gegenüber der Öffentlichkeit – nicht gestört. Die Nennung des Urhebers in diesem Kontext rückt weder ihn noch sein Werk in einen irgendwie gearteten negativen Zusammenhang.

V. Fazit

Die Erstellung und Verwendung abgeleiteter Formate ist bei Einhaltung der aufgezeigten Anforderungen und Bedingungen ein rechtlich gangbarer Weg,

⁹² Durch die Erstellung abgeleiteter Formate werden die Originalwerke in ihrer Existenz nicht berührt. Eine Entstellung in Form der Zerstörung vgl.: BGH, Urt. v. 21.02.2019 – I ZR 98/17, GRUR 2019, 609 – *HHole*; BGH, Urt. v. 21.02.2019 – I ZR 99/17, ZUM 2019, 521; BGH, Urt. v. 21.02.2019 – I ZR 15/18, GRUR 2019, 619 – *Minigolfanlage*) kommt insofern nicht in Betracht.

⁹³ Schrickler/Loewenheim-Peukert, Urheberrecht, 6. Aufl. 2020, Vor §§ 12 ff. UrhG Rn. 4.

⁹⁴ § 11 S. 1 UrhG.

um Forschung durch Textmining mehr Freiheit zu verschaffen und gleichzeitig die geschützten Interessen der Rechteinhaber an den Texten zu wahren. Entscheidend ist es, ein Format zu verwenden, in dem die schutzfähigen Elemente aller Texte weitestgehend ausgespart werden. Geeignet sind deshalb Formate, in denen der Ursprungstext gar nicht mehr erkennbar ist oder zumindest die semantischen Strukturen derart kleinteilig aufgebrochen sind, dass sichergestellt ist, dass die schutzfähigen Elemente beseitigt sind.

Der bisher noch bestehende Konflikt zwischen Schutzrechten und allgemeingellschaftlich gestützten Forschungsinteressen kann durch die Verwendung abgeleiteter Formate schonend aufgelöst werden. Abgeleitete Formate sind nicht für jede Forschungsfrage nutzbar, dort wo mit ihnen gearbeitet werden kann, räumen sie den Forschenden aber Freiheit und weitgehende Rechtssicherheit ein.

C. Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen^{*}

Christof Schöch/Evelyn Gius/Peer Trilcke/Peter Leinen/Fotis Jannidis/Maria Hinzmann/Jörg Röpke

Version (2020), CC BY 4.0.

I. Einleitung

Es ist ein offenes Geheimnis in den *Digital Humanities* (DH), dass es für die *Computational Literary Studies* (CLS) bezüglich der verfügbaren Textbestände ein *window of opportunity* gibt, das sich um 1800 öffnet und um 1920 wieder schließt. Es öffnet sich um 1800, weil für Materialien vor dieser Zeit die technischen Herausforderungen im Bereich *Optical Character Recognition* (OCR) und Normalisierung von orthographischer Varianz immer noch so groß sind, dass deutlich weniger umfangreiche beziehungsweise qualitativ weniger hochwertige Textsammlungen zur Verfügung stehen als für die Zeit nach 1800. Und es schließt sich um 1920, weil für Texte, die später erschienen sind, in sehr vielen Fällen (abhängig vom Todesdatum der Autor*innen) das Urheberrecht nach wie vor greift und sowohl das Erstellen als auch das Teilen von Textsammlungen mit Dritten damit deutlich erschwert sind. Dieser Umstand hat bedauerlicherweise zur Folge, dass die Setzung von Forschungsschwerpunkten häufig nicht primär von den Erkenntnisinteressen und Zielen der Forschung selbst, sondern wesentlich von technischen und rechtlichen, also dieser Forschung externen Faktoren, bestimmt wird. Als Konsequenz daraus ist eine Forschung auf dem methodischen *state of the art* mit neueren Textbeständen nur begrenzt, teilweise sogar überhaupt nicht möglich. Die Forschung in den CLS verwendet zwar ak-

^{*} Dieser Beitrag wurde zuerst veröffentlicht in Zeitschrift für digitale Geisteswissenschaften unter einer CC BY Lizenz. Wolfenbüttel 2020. text/html Format. DOI: 10.17175/2020_006.

tuelle, oft aus Informatik, Computerlinguistik und Statistik adaptierte Verfahren, kann sie aber in den meisten Fällen nicht auch auf diejenigen Textbestände anwenden, die unsere zeitgenössische literarische Kultur ausmachen.¹ Allerdings verbessert sich die Lage seit einigen Jahren deutlich, sodass es Anlass zu Optimismus gibt: Auf der einen Seite wird derzeit verstärkt in neue Verfahren für OCR investiert, bei denen die Texterkennung auf neuronalen Netzen beruht und deutliche Verbesserungen auch für Materialien vor 1800, bis hin zu frühen Dru*cken und Handschriften, erreichen konnte². Auf der anderen Seite haben sich die rechtlichen Rahmenbedingungen in den letzten Jahren stark zum Positiven verändert, in Deutschland insbesondere mit dem März 2018 in Kraft getretenen Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft (UrhWissG)³, welches das Gesetz über Urheberrecht und verwandte Schutzrechte (UrhG) modifizierte. So enthält das derzeit geltende nationale Recht in § 60d UrhG eine Schranke⁴ von Urheber- und Leistungsschutzrechten zugunsten des Text und Data Mining.⁵ Diese Schranke gestattet nun Vervielfältigungen zum Zwecke des Text und Data Mining (TDM) in der Forschung, erlaubt allerdings nur unter engen Voraussetzungen die langfristige Speicherung und die Weitergabe der hierfür erstellten Korpora an Dritte. So darf beispielsweise die langfristige Speicherung (und fallweise Weitergabe) der Korpora nur durch Institutionen wie Bibliotheken und Archive erfolgen. Und es ist zwar gestattet, ein Korpus für die wissenschaftliche Qualitätssicherung (peer review) an einzelne Dritte weiterzugeben, nicht aber für die ebenso wichtige Anschlussforschung. Auch auf der europäischen Ebene bringt

¹ Der vorliegende Beitrag ist im Kontext der Workshopreihe »Strategien für die Nutzbarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte« am 28.11.2019 und 17.01.2020 in Trier, organisiert von Benjamin Raue (IRDT) und Christof Schöch (TCDH) entstanden. Wir danken den Teilnehmer*innen des Workshops für die engagierten und produktiven Diskussionen und der DFG für die Förderung der Veranstaltung im LIS-Programm. Ein wesentlicher Impuls für die Workshopreihe war die Trierer Tagung Text und Data Mining – in Recht, Wissenschaft und Gesellschaft 2018: <https://text-und-data-mining.de/>, zuletzt abgerufen am 15.03.24.

² Vgl. zu großen Initiativen wie OCR-D oder Werkzeugen wie OCR4all u. a. *Neudecker et al.*, OCR-D: An end-to-end open source OCR framework for historical documents, in *European Tech* (2019), H. 13; *Reul et al.*, OCR4all—An Open-Source Tool Providing a (Semi-) Automatic OCR Workflow for Historical Printings, in *Applied Sciences* 9 (2019), H. 22, Nr. 4853.

³ Hierbei handelt es sich um die Ergänzungen in §§ 60a–60h UrhG.

⁴ Eine »Schranke« bezeichnet im rechtswissenschaftlichen Kontext eine gezielte Einschränkung oder eine Ausnahme von einer allgemeineren rechtlichen Regelung.

⁵ Einführend dazu *Raue*, Rechtssicherheit für datengestützte Forschung, in *Zeitschrift für Urheber- und Medienrecht* 63 (2019), H. 8–9, S. 684–693.

die Europäische Richtlinie 2019/790 zum Urheberrecht im digitalen Binnenmarkt (DSM-RL) von 2019 erneut eine Öffnung für das TDM, die auch in deutsches Recht umgesetzt werden wird; doch auch diese Regelungen werden nicht alle Bedarfe des TDM erfüllen.

Der vorliegende Beitrag macht Vorschläge für einen pragmatischen Umgang mit der derzeitigen rechtlichen Situation bei der Nutzung von Methoden des TDM⁶ in den DH und speziell in den CLS.⁷ Es sollen Perspektiven und Möglichkeiten für die Erstellung, Analyse und Anschlussforschung an solchen Textsammlungen, die auf der Grundlage von urheberrechtlich geschützten Textbeständen entstanden sind, eröffnet werden. Ziel ist es, die offene Publikation und freie Nachnutzbarkeit von abgeleiteten Textformaten – zur Nachvollziehbarkeit von Analyseergebnissen für Dritte auch außerhalb formaler Qualitätssicherungsprozesse wie auch zur Anschlussforschung ohne rechtliche Einschränkungen – zu ermöglichen. Dies ist ein Bereich, den weder das geltende noch das anstehende neue Recht gesondert behandeln, wodurch die rechtsuchenden Forschenden weitgehend auf allgemeine Regeln und Grundsätze des Urheberrechts verwiesen bleiben.

Grundsätzlich kann man vier Ansätze zur Lösung der beschriebenen Problematik unterscheiden:

- den Zugang zu lizenzierten Inhalten über eine API
- die Nutzung von Analyseplattformen
- die Forschung im *closed room*
- und die Arbeit mit abgeleiteten Textformaten

Erstens kann man also auf die lizenzierten Angebote beispielsweise von großen Verlagen setzen, die es über Schnittstellen, sogenannte Application Programming Interfaces (APIs), ermöglichen, gezielt ausgewählte, große Mengen an Text-/Datenbeständen herunterzuladen, um sie dann selbst mit Methoden des

⁶ Vgl. zu TDM u. a. *Hotbo et al.*, A brief survey of text mining, in LDV Forum – GLDV Journal for Computational Linguistics and Language Technology 20 (2005), H. 1, 19–62; *Al-labyari et al.*, A brief survey of text mining: Classification, clustering and extraction techniques, in Computing Research Repository (CoRR) arXiv.org. Version 1 arXiv:1707.02919 vom 10.07.2017.

⁷ Vgl. zu CLS u. a. *Jannidis/Koble/Rehbein* (Hrsg.), Digital Humanities: Eine Einführung, Stuttgart 2017; *Jockers*, Macroanalysis – Digital Methods and Literary History, Urbana, IL u. a. 2013; *Schöch*, Quantitative Analyse, in Jannidis/Kohle/Rehbein (Hrsg.) Digital Humanities, 280–299.

TDM zu analysieren. Nachteilig hieran ist, dass man auf die Bestände der Verlage beschränkt ist und in der Regel die auf diese Weise erstellten Korpora nicht weitergeben darf. Zweitens kann man auf Plattformen von Datenanbietern setzen, die bestimmte Textbestände und Analyseverfahren zur Verfügung stellen und so Forschung gewissermaßen vom Browser aus ermöglichen. Nachteilig an diesem Modell ist, dass die Forschenden dabei die Daten selbst nicht einsehen, herunterladen, modifizieren oder ergänzen oder Datensätze aus verschiedenen Quellen kombinieren können. Außerdem können die Forschenden die Analyseverfahren selbst nicht implementieren oder grundlegend modifizieren. Drittens kann das closed room-Modell zum Einsatz kommen, bei dem Forschende in einem technisch abgeschotteten Raum in einer bestimmten Institution Zugang zu Textbeständen der Institution erlangen. Dies können sie dann auf dem vor Ort verfügbaren Arbeitsplatzrechner auch mit individuell entwickelten Analyseprogrammen untersuchen. Hier liegen die Nachteile in erster Linie in der Ortsgebundenheit der Forschungstätigkeit, die einen entsprechenden praktischen Aufwand mit sich bringt und in Zeiten der Digitalisierung wenig zeitgemäß erscheint; zudem skaliert dieses Modell nicht gut, können Datensätze aus mehreren Institutionen nicht kombiniert und an Dritte weitergegeben werden.

Die hier entwickelten Perspektiven beziehen sich in erster Linie auf den vierten Ansatz, die abgeleiteten Textformate. Abgeleitete Textformate sind auf der Grundlage eines Ausgangstextes systematisch generierte Repräsentationen des Textes, welche die Anwendung bestimmter Verfahren des Text und Data Mining erlauben, wobei urheberrechtlich geschützte Bestandteile des Ausgangstextes im abgeleiteten Format aber nicht mehr repräsentiert sind.⁸ Hieraus ergeben sich neue Möglichkeiten, aber auch neue Herausforderungen: Die Forschenden können dadurch ihre Forschungsfragen anhand größerer und vor allem aktuellerer Textbestände bearbeiten und damit auch neue Forschungsfragen erschließen; allerdings müssen die eingesetzten Verfahren gegebenenfalls auf die abgeleiteten Textformate angepasst werden. Ebenso werden durch die freie Verfügbarkeit der abgeleiteten Textformate Transparenz und Reproduzierbarkeit von

⁸ Die rechtlichen Kriterien, die ein Urteil hinsichtlich der Frage erlauben, ob ein abgeleitetes Textformat dem Urheberrecht unterliegt oder nicht, sind Gegenstand eines Beitrags von Karina Grisse im vorliegenden Band (*Grisse*, Teil 3, B. in diesem Band); zu den relevanten Faktoren gehören u. a. die Möglichkeit des Werkgenusses, die Wiedererkennbarkeit und die Rekonstruierbarkeit des Ausgangstextes. Die durch das Urheberrechtsgesetz vermittelten Schutzmöglichkeiten spezifisch für Texte sowie die Schutzvoraussetzungen erläutert Florian Jotzo (*Jotzo*, Teil 2, B. in diesem Band). Beide Beiträge sind als rechtswissenschaftliche *companion papers* zum vorliegenden Beitrag konzipiert.

Forschung und die unbeschränkte Zirkulation und Nachnutzbarkeit von Forschungsdaten bei urheberrechtlich geschützten Ursprungsmaterialien überhaupt erst umfassend ermöglicht; allerdings ergibt sich für alle Forschenden eine hohe Abhängigkeit von der Transparenz, Plausibilität und Korrektheit der Erstellungsprozesse der abgeleiteten Textformate.

Zunächst und noch grundlegender stellt sich aber die Frage, welche abgeleiteten Textformate überhaupt für einen solchen Ansatz geeignet sind. Mit dem Ziel, erste Antworten auf diese Frage zu entwickeln und eine Diskussion in der Fachgemeinde der DH anzuregen, wird im Folgenden zunächst die aktuelle rechtliche Situation zusammengefasst. Anschließend werden das Prinzip der abgeleiteten Textformate sowie die Anforderungen an abgeleitete Textformate für verschiedene relevante Verfahren des TDM diskutiert. Es folgt die Beschreibung und kritische Begutachtung mehrerer konkreter abgeleiteter Formate, und zwar sowohl bezüglich ihrer Erstellung, Publikation und Anwendung als auch hinsichtlich ihrer urheberrechtlichen Einordnung. Abschließend kommen im Sinne eines Ausblicks auf die nächsten erforderlichen Schritte einige weiterführende Punkte zur Sprache: Diese beleuchten insbesondere die neue Rolle von Bibliotheken und Archiven als Datenanbieter, skizzieren aber auch eine Agenda für Informatik, Computerlinguistik und DH. Ziel des Beitrags ist es in der Summe, eine Diskussion nicht nur innerhalb der DH, sondern auch mit den Bibliotheken und Archiven einerseits sowie den Rechtswissenschaften andererseits anzuregen. Am Ende einer solchen Diskussion könnte ein Konsens über ein Inventar geeigneter abgeleiteter Textformate stehen, für die standardisierte Lösungen zur Erstellung und Bereitstellung entwickelt und angeboten werden. Dies wird auch Grundlage für weitergehende Diskussionen z. B. mit Vertreter*innen der Rechteinhaber (u. a. der Verlage) sein können.

II. Die aktuelle Rechtslage zum Text und Data Mining mit urheberrechtlich relevanten Textbeständen

Bevor auf die im Beitrag zentrale Strategie der abgeleiteten Textformate eingegangen wird, soll an dieser Stelle zunächst die aktuelle Rechtslage zum Thema Text und Data Mining (TDM) mit urheberrechtlich relevanten Textbeständen zusammenfassend dargestellt werden. Unter TDM versteht das Urheberrecht

das automatisierte Auswerten einer Vielzahl von Werken⁹ beziehungsweise, wie es nun in Art. 2 DSM-RL EU-weit korrespondierend, aber etwas ausführlicher legaldefiniert ist, die »Technik für die automatisierte Analyse von Texten und Daten in digitaler Form, mit deren Hilfe Informationen unter anderem – aber nicht ausschließlich – über Muster, Trends und Korrelationen gewonnen werden können.«¹⁰

Das TDM selbst ist in der Regel keine urheberrechtsrelevante Handlung, sehr wohl aber all die Vervielfältigungs- und Bearbeitungsschritte, die mit der Korpusbildung einhergehen, und gegebenenfalls auch das, was nach der Analyse in der Ergebnispräsentation von der Verfasstheit und ästhetischen Identität der analysierten Werke wieder sichtbar wird. Hier entstehen Konflikte zwischen Forschungsinteressen einerseits und einem in seinem Kern vordigitalen, seit 1966 in Kraft befindlichen UrhG andererseits. Die Sache liegt zusätzlich schief, weil es beim TDM nicht um Werkintegrität und Werkgenuss im klassischen Sinne geht, also das, was das Urheberrecht zu regulieren sucht. Vielmehr handelt es sich um Informationen, die sich quasi in geschützten Werkhüllen finden, auch wenn sie oftmals für sich genommen selbst gar nicht schutzfähig sind. Zusätzlich verkompliziert wird die Lage dadurch, dass es im TDM vielfach um Massenanalysen geht, was Einzellizenzierungen der analysierten Werke und damit diese Art Forschung an geschützten Inhalten praktisch unmöglich macht. TDM war daher vor dem UrhWissG heikel und nur mit aufwendigen argumentativen Konstruktionen als erlaubnisfrei rechtlich herzuleiten. Die Sache war also mindestens rechtsunsicher, das Ausweichen auf klar rechtsfreie Analysegegenstände die Regel. Aus wissenschaftlicher Sicht war daher durch die technische und forschungspolitische Entwicklung ein Bedarf für eine Beschränkung der beim TDM berührten Urheber- und Leistungsschutzrechte entstanden, die einen Spielraum für rechtssicheres erlaubnisfreies TDM schafft. Dieser Einschätzung folgte auch der Gesetzgeber, zumal man darauf verweisen konnte, dass die grundrechtlich abgesicherten berechtigten persönlichkeitsrechtlichen und vor allem ökonomischen Interessen der Rechteinhaber durch TDM in der Regel überhaupt nicht nennenswert tangiert werden. Das derzeit geltende nationale Recht wurde vor diesem Hintergrund 2018 in Deutschland unter Bezugnahme auf die allgemeinen Ausnahmen und Beschränkungen zugunsten von Wissenschaft¹¹ und die 2001 in der sogenannten InfoSoc-RL 2001/29/EG EU-

⁹ Vgl. § 60d Abs. 1 UrhG.

¹⁰ Art. 2 Nr. 2 DSM-RL.

¹¹ Vgl. Art. 5 Abs. 3 Buchst. a InfoSoc-RL 2001/29/EG.

weit gewährten und vollharmonisierten Vervielfältigungsrechte von Urhebern und Leistungsschutzberechtigten¹² eingeführt. Es sieht seit nunmehr zwei Jahren in § 60d UrhG eine Schranke von Urheber- und Leistungsschutzrechten zugunsten des Text und Data Mining vor.¹³ Nach diesem seit März 2018 geltenden Recht ist nun insbesondere Folgendes erlaubnisfrei zulässig:¹⁴ das Erstellen von Korpora aus Werken jedweder Art¹⁵ für das TDM einschließlich aller für Korpusbildung und das anschließende TDM erforderlichen Bearbeitungs- und Vervielfältigungshandlungen etwa des Digitalisierens, Normalisierens, Strukturierens, Kategorisierens, Sortierens, Annotierens, Kombinierens aus verschiedenen Quellen,¹⁶ vorausgesetzt, man verfolgt nicht-kommerzielle, wissenschaftliche Forschung und hat rechtmäßigen Zugang zum Ursprungsmaterial. Für alle Arbeitsschritte der Vorbereitung und Durchführung dürfen sich die Forschenden dabei auch von Dritten, etwa Gedächtnisinstitutionen, unterstützen lassen; des Weiteren dürfen Forschende das Korpus bis zur Grenze eines fest bestimmten, aber doch schon rechtlich als öffentlich geltenden Personenkreises zur gemeinsamen Arbeit daran zugänglich machen; und sie dürfen nach Projektabschluss zur Sicherung von Referenzierbarkeit und Qualitätsprüfung das Korpus privilegierten Gedächtnisinstitutionen zur dauerhaften Aufbewahrung übergeben.

¹² Vgl. Art. 2 InfoSoc-RL 2001/29/EG.

¹³ Vgl. zu UrhWissG und DH/TDM *Berger*, Urheberrecht in der Wissensgesellschaft, in *Gewerblicher Rechtsschutz und Urheberrecht* 119 (2017), H. 10, 953–964; *Durantaye*, Neues Urheberrecht für Bildung und Wissenschaft – eine kritische Würdigung des Gesetzentwurfs, in *Gewerblicher Rechtsschutz und Urheberrecht* 119 (2017), H. 6, 558–567; *Pflüger/Hinte*, Das Urheber-Wissengesellschafts-Gesetz aus Sicht von Hochschulen und Bibliotheken, in *Zeitschrift für Urheber- und Medienrecht* 62 (2018), H. 3, 153–161; *Raue*, Das Urheberrecht der digitalen Wissen(schaft)s-gesellschaft, in *Gewerblicher Rechtsschutz und Urheberrecht* 119 (2017), H. 1, 11–19; *Ders.*, Text und Data Mining, in *Computer und Recht* 33 (2017), H. 10, 656–662; *Schack*, Das neue UrhWissG – Schranken für Unterricht, Wissenschaft und Institutionen, in *Zeitschrift für Urheber- und Medienrecht* 61 (2017), H. 11, 802–808; *Specht*, Die neue Schrankenregelung für Text und Data Mining und ihre Bedeutung für die Wissenschaft, in *Ordnung der Wissenschaft* (2018), H. 4, 285–289; *Spindler*, Text und Data Mining – urheber- und datenschutzrechtliche Fragen, in *Gewerblicher Rechtsschutz und Urheberrecht* 118 (2016), H. 11, 1112–1120; *Ders.*, Text und Datamining im neuen Urheberrecht und in der europäischen Diskussion, in *Zeitschrift für Geistiges Eigentum* 10 (2018), H. 3, 273–300.

¹⁴ Vgl. insoweit einführend die Standardcommentierungen zu § 60d UrhG in *Dreier*, in *Ders./Schulze*, *UrhG*, 6. Auflage 2018, § 60d Rn 116; *Nordemann*, in *Nordemann et al.*, *UrhG*, 12. Auflage 2018, § 60d Rn 1–13; *Anton*, in *Spindler/Schuster*, *UrhG*, 4. Auflage 2019, § 60d Rn 1–23; *Bullinger*, in *Wandtke/Bullinger*, *UrhG*, 5. Auflage 2019, § 60d Rn 1–21; *Hagemeyer*, in *Ahlberg/Götting*, *UrhG*, 28. Edition 2020, § 60d Rn 1–21.

¹⁵ §§ 2 UrhG: Texte, Töne, Bilder, Filme usw., sogar Datenbankwerke, § 60d Abs. 2 UrhG.

¹⁶ Vgl. § 23, 3; § 44a; § 60d, Abs. 1, S. 1, Nr. 1 UrhG.

Das ist viel, verglichen mit der Rechtslage zuvor. Doch so positiv diese mit dem UrhWissG vorgenommenen Änderungen für die wissenschaftliche Arbeit im Bereich des TDM in mancherlei Hinsicht auch sind, sie stellen weiterhin entscheidende Hürden für die Forschung dar, die den Mehrwert und die Praktikabilität der Regelung für die Forschung spürbar einschränken:

- Rechtmäßiger Zugang zu den verwendeten Materialien wird z. B. vorausgesetzt, nicht unter Verweis auf TDM zum Zwecke wissenschaftlicher Forschung automatisch gewährt.
- Es besteht kein Recht auf Umgehung etwaiger technischer Schutzmaßnahmen, nur das Recht, vom Beschränkenden Mittel zur Aufhebung des Schutzes zu erhalten,¹⁷ aber auch nur, wenn dies nicht bei Online-Inhalten durch vertragliche Vereinbarung, auch als AGB, ausgeschlossen ist.¹⁸
- Die TDM-Schranke gilt nur im Rahmen von nichtkommerzieller Forschung.
- Das Korpus ist nach Ende des Forschungsprojektes an eine Gedächtnisinstitution abzugeben oder zu löschen.
- Das Korpus kann der Wissenschaftscommunity jenseits des engen Bereichs der Qualitätsprüfung nicht für Anschlussforschung zugänglich gemacht werden.
- Änderungsverbot, Vergütungspflicht und Pflicht zu Quellenangaben sind zu beachten.
- Die Arbeit mit rechtswidrigen Werken (Plagiaten, geleaktem Material, jugendgefährdendem Material usw.) ist nicht geklärt.

All dies sind nur Beispiele für nach wie vor bestehende Hürden und Unsicherheiten. Und hier sind nur Urheber- und Leistungsschutzrecht angesprochen, das Datenschutzrecht kommt erst noch,¹⁹ ist aber ebenfalls an vielen Stellen zentral für die Arbeit mit Gegenwartskultur. So kann Forschung, die prinzipiell auf verfügbaren Vorgängerarbeiten aufbaut und so fortschreitet, nicht effektiv und effizient funktionieren.

Dieses Recht wird nun durch die Europäische Richtlinie 2019/790 zum Urheberrecht im digitalen Binnenmarkt (DSM-RL) von 2019 nochmals geändert

¹⁷ Vgl. § 95b Abs. 1 S. 1 Nr. 11 UrhG.

¹⁸ Vgl. § 95 Abs. 3 UrhG.

¹⁹ Vgl. Art. 5, 6, 89 DSGVO.

werden, die derzeit bis Juni 2021 in nationales Recht umgesetzt wird.²⁰ Die Umsetzung der DSM-RL wird das deutsche TDM-Recht in signifikanten Punkten modifizieren und dabei zumindest einige der vorgenannten Hürden und Unsicherheiten adressieren. So gilt insbesondere:

- Die Befristung des UrhWissG und damit des TDM auf lediglich 5 Jahre wird überflüssig.
- Die Pflichten zur Vergütung und Quellenangabe entfallen.
- Das TDM selbst an Einzelwerken wird zulässig.
- Die Option zur Beschränkung der Nutzung online verfügbarer Inhalte wird aufgehoben.
- Und insbesondere werden die Korpora nach Projektende für Anschlussforschung nachnutzbar.

Gerade Letzteres ist ein großer, potentiell folgenreicher Schritt. Doch so begrüßenswert all dies aus Sicht der Forschung ist, die Korpora können auch nach Inkrafttreten des neuen Rechts voraussichtlich weiterhin nicht allgemein öffentlich zugänglich gemacht werden. Dies gilt selbst, wenn dieser Zugang auf die dann vom neuen Recht privilegierten Akteursgruppen (Forschungsorganisationen, Einrichtungen des Kulturerbes und Einzelforschende) und ausschließlich zum Zwecke nichtkommerzieller, wissenschaftlicher Forschung durch diese be-

²⁰ Vgl. zu DSM-RL und DH/TDM *Dreier*, Die Schlacht ist geschlagen – ein Überblick. Zum Ergebnis des Copyright Package der EU-Kommission, in *Gewerblicher Rechtsschutz und Urheberrecht* 121 (2019), H. 8, 771–779; *Ducato/Strowel*, Limitations to Text and Data Mining and Consumer Empowerment, in *International Review of Intellectual Property and Competition Law* 50 (2019), H. 6, 649–684; *Flehsig*, Europäisches Urheberrecht in der Digitalität, in *Jur-PC. Internet-Zeitschrift für Rechtsinformatik und Informationsrecht* 145 (2019), Abs. 1–320; *Geiger et al.*, Text and Data Mining Articles 3 and 4 of the Directive 2019/790/EU, in *García/Llorca* (Hrsg.) *Propiedad intelectual y mercado único digital europeo*, Valencia 2019; *Raue*, Rechtssicherheit für datengestützte Forschung, in *Zeitschrift für Urheber- und Medienrecht* 63 (2019), H. 8–9, 684–693; *Ders.*, Die geplanten Text- und Data Mining-Schranken (§§ 44b und 60d UrhG-E), in *Zeitschrift für Urheber- und Medienrecht* 64 (2020), H. 3, 172–175; *Schaper/Verweyen*, Die Europäische Urheberrechtsrichtlinie (EU) 2019/790, in *Kommunikation und Recht* 22 (2019), H. 7/8, 433–441; *Spindler*, Die neue Urheberrechts-Richtlinie der EU, insbesondere >Upload-Filter< – Bittersweet? in *Computer und Recht* 35 (2019), H. 5, 277–291; *Steinbrecher*, Die EU-Urheberrechtsrichtlinie aus Sicht der Digitalwirtschaft. Zeit für Augenmaß und faktenbasierte Gesetzgebung, in *Multimedia und Recht* 22 (2019), H. 10, 639–643; *Stieper*, Das Verhältnis der verpflichtenden Schranken der DSM-RL zu den optionalen Schranken der InfoSoc-RL, in *Gewerblicher Rechtsschutz und Urheberrecht* 122 (2019), H. 1, 1–7.

schränkt würde. Sie sind stattdessen unter angemessenen Sicherheitsvorkehrungen nicht öffentlich aufzubewahren. Und es ist zwar wenig konturiert im anstehenden neuen Recht, aber die nun sogar ausdrücklich vorgesehene Voraussetzung des rechtmäßigen Zugangs zum Ursprungsmaterial lässt erwarten, dass die Korpora jedenfalls (jenseits gemeinsamer Projektarbeit mit den Zugangsberechtigten an den Korpora) nicht von Dritten ohne deren eigenen Zugang genutzt werden dürfen. Sonst könnte diese zentrale Bedingung im Interessenausgleich zwischen Rechteinhaber*innen und Forschenden leicht umgangen werden. Die Korpora werden demnach bedingt nachnutzbar, aber nicht publizierbar.

III. Das Prinzip der abgeleiteten Textformate

Vor dem skizzierten rechtlichen Hintergrund ist die Grundidee der abgeleiteten Textformate im Kern folgende: Es wird von Beständen urheberrechtlich geschützter Volltexte ausgegangen (die Ausgangstexte, im Urheberrecht als ›Ursprungsmaterial‹ bezeichnet; gegebenenfalls auch bereits als Korpus vorliegend), zu denen eine Institution legalen Zugang hat. Diese Textbestände werden durch die Anwendung von Verarbeitungsroutinen, die im Wesentlichen sowohl eine gezielte Informationsanreicherung (beispielsweise durch linguistische Annotation) als auch eine Informationsreduktion (beispielsweise durch Löschung der Wortformen oder Aufhebung der Sequenzinformation) darstellen, in sogenannte abgeleitete Textformate verwandelt.²¹ Diese Verarbeitungsroutinen können gegebenenfalls in Verbindung mit einem konkreten Forschungsvorhaben der eigenen Institution oder Dritter angewendet werden. Das einfachste Beispiel für ein solches abgeleitetes Textformat wäre eine Tabelle, die für einen Textbestand die Häufigkeiten jedes Wortes in jedem Text festhält. Solche abgeleiteten Textformate können für einen unstrukturierten Gesamtbestand, einen größeren Teilbestand oder aber für einen gezielt zusammengestellten, für die Bearbeitung einer bestimmten Forschungsfrage geeigneten Teilbestand von Texten erstellt werden.²² Die abgeleiteten Textformate sind dabei so

²¹ Das Verfahren hat eine gewisse Beziehung zum Prinzip der *differential privacy* bei sozialwissenschaftlichen Datenerhebungen. Dort wird durch gezielte Randomisierung von Antworten die Privatsphäre der einzelnen Teilnehmenden geschützt, ohne dass die Verlässlichkeit von Schlussfolgerungen auf der Ebene aller Befragten eingeschränkt wird. Vgl. u. a. *Dwork/Roth*, *The Algorithmic Foundations of Differential Privacy*, in *Foundations and Trends in Theoretical Computer Science* 9 (2014), H. 3–4, 5–27.

²² Erst wenn ein Datenbestand für eine Analyse zusammengestellt und gegebenenfalls systematisiert und aufbereitet wurde, spricht das Urheberrecht von einem ›Korpus‹.

gestaltet, dass die Texte in der dann vorliegenden Form einerseits nicht mehr in den Geltungsbereich des Urheberrechts fallen, andererseits dennoch die Anwendung möglichst vielfältiger quantitativer Analysen der Texte erlauben. Zwar enthalten die abgeleiteten Textformate abhängig von der jeweiligen Umsetzung verglichen mit den Ausgangstexten manche Informationen nicht mehr, sodass nicht alle relevanten Analyseverfahren uneingeschränkt umsetzbar sind (siehe hierzu ausführlich Abschnitt 5). Dafür können solche Datenbestände ohne Einschränkungen gespeichert, in der Forschung genutzt, veröffentlicht und von Dritten nachgenutzt werden. Über die Nachnutzung und Veröffentlichung hinaus ist – bei rechtmäßigem Zugang zum Ursprungsmaterial – zudem auch die Erstellung der abgeleiteten Textformate erlaubnisfrei möglich.

Die Idee der abgeleiteten Textformate ist nicht neu, vielmehr gibt es bereits mehrere Beispiele für die erfolgreiche Umsetzung dieses Prinzips. Zu den prominentesten Beispielen für den Einsatz abgeleiteter Textformate zählen das Google Ngram Dataset sowie das HTRC Extracted Features Dataset der Hathi Trust Digital Library. Das Google Ngram Dataset basiert auf einem Korpus von rund 8 Millionen Büchern in mehreren Sprachen, das auch eine starke diachrone Dimension aufweist. Der Datensatz enthält für jedes Wort (und für jedes N-Gramm der Länge 2–5) in jedem Jahr die Angabe darüber, wie häufig das Wort vorkommt und in wie vielen verschiedenen Dokumenten es vorkommt.²³ Seit der Version 2 von 2012 sind die Texte auch nach Wortarten annotiert. Das HTRC Extracted Features Dataset enthält für jede einzelne Seite im umfangreichen Dokumentenbestand der Hathi Trust Digital Library das Inventar an Tokens mit Angabe der Wortart und der Häufigkeit auf der Seite.²⁴ Derzeit enthält der Datensatz diese Informationen für rund 15 Millionen Dokumente oder fast 6 Milliarden Seiten. Knapp zwei Drittel des Datenbestands sind dabei urheberrechtlich geschützt, der Rest ist gemeinfrei. Der Datensatz kann ganz oder in

²³ Vgl. *Lin et al.*, Syntactic Annotations for the Google Books NGram Corpus, in Zhang (Hrsg.) System demonstrations. 50th annual meeting of the Association for Computational Linguistics 2012 (ACL 2012: 50, Jeju Island, 08.–14.07.2012), Red Hook 2012, NY, 169–174.

²⁴ Vgl. *Bhattacharyya et al.*, A Fragmentizing Interface to a Large Corpus of Digitized Text: (Post)humanism and Non-consumptive Reading via Features, in *Interdisciplinary Science Reviews* 40 (2015), H. 1, 61–77; *Jett et al.*, The HathiTrust Research Center Extracted Feature Dataset (2.0), in HathiTrust Research Center (Hrsg.) wiki.htrc.illinois.edu, Blogbeitrag v.2.0 vom 16.06.2020.

Teilen heruntergeladen oder mit dem HTRC Feature Reader ausgelesen werden. Die Dokumente sind nach Seiten segmentiert und mit detaillierten bibliografischen Metadaten angereichert.

Weitere Beispiele für den Einsatz abgeleiteter Textformate sind der CrossAsia N-Gram Service der Staatsbibliothek zu Berlin, die Datensätze im Paket *stylo* für R, die Leipzig Corpora Collection und das Open Super-large Crawled ALMA-naCH coRpus (OSCAR). Im Fall des CrossAsia N-Gram Service sind die lizenzierten Volltextbestände registrierten Nutzer*innen vorbehalten, frei verfügbar sind aber die davon abgeleiteten Häufigkeitsinformationen, die für verschiedene N-Gramm-Größen angeboten werden. Für jedes Dokument und jede N-Gramm-Größe liegt eine separate Datei mit den Häufigkeitsinformationen vor sowie eine separate Metadatentabelle. Mit insgesamt rund 13.000 Dokumenten handelt es sich um einen substantiellen Datenbestand historischer Texte. Im Fall von *stylo* (einem Werkzeug für stilometrische Analysen in der Statistikumgebung R), werden mehrere Sammlungen urheberrechtlich geschützter Texte in Form einer einfachen Term-Dokument-Matrix bei der Installation mitgeliefert und stehen direkt für Beispielanalysen zur Verfügung. Es handelt sich um vergleichsweise kleine Datensätze: Der Galbraith-Datensatz enthält beispielsweise die Häufigkeiten der 3.000 häufigsten Wörter für 28 Romane, ohne weitere Segmentierung.²⁵ Die Leipzig Corpora Collection bietet für rund 250 Sprachen teils sehr umfangreiche Korpora an, die auf überwiegend urheberrechtlich geschützten Texten aus dem Internet basieren, von denen ein zufälliges Sample einzelner Sätze verfügbar gemacht wird.²⁶ OSCAR ist ein äußerst umfangreiches Korpus, in dem 166 Sprachen vertreten sind, wobei hier pro Sprache die Reihenfolge der Zeilen gegenüber den Ausgangsdokumenten randomisiert wurde.²⁷

Diese Beispiele zeigen, dass die Vorteile der Idee abgeleiteter Textformate durchaus bereits erkannt worden sind. Es wird aber auch deutlich, dass erstens die Umsetzung bisher nur wenig programmatisch erfolgt, denn es gibt kaum Forschungsliteratur, die sich spezifisch diesem Thema widmet, und dass es zweitens

²⁵ Vgl. *Eder et al.*, *Stylometry with R: A package for computational text analysis*, in *The R Journal* 8 (2016), H. 1, 107–121.

²⁶ Vgl. *Goldbahn et al.*, *Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages*, in Nicoletta (Hrsg.) *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12, Istanbul, 21.–27.05.2012)*, Paris 2012.

²⁷ Vgl. *Suárez et al.*, *Asynchronous Pipeline for Processing Huge Corpora on Medium to Low Resource Infrastructures*, in Bánski et al. (Hrsg.) *Proceedings of the Workshop on Challenges in the Management of Large Corpora*, Mannheim 2019.

bisher kaum Bemühungen um eine Standardisierung von Formaten und Strategien über Einzelprojekte oder einzelne Institutionen hinweg gibt.

Wie kann man sich dem Konzept der abgeleiteten Textformate also grundsätzlicher nähern? Zunächst ist zu konstatieren, dass ein Text nur scheinbar aus einer schlichten Abfolge von Wortformen oder gar Zeichen besteht. Denn für die verstehende Lektüre eines Textes ist die Kenntnis nicht nur der Wortformen und ihrer genauen Reihenfolge notwendig, sondern auch die Kenntnis der Bedeutung und grammatikalischen Funktion der Wortformen im Satz sowie der semantischen und syntaktischen Beziehungen zwischen den Wortformen. Hinzu kommen noch Kontext und Pragmatik des Textes über die Satzgrenzen hinaus.

Um eine systematische Modellierung der abgeleiteten Textformate vorzunehmen, wird hier abstrahierend davon ausgegangen, dass ein Text lediglich in die folgenden Teile zu gliedern ist:

- Token (vereinfacht gesagt: ein einzelnes Wort)
- Satz
- Segment (Abschnitte fester, aber willkürlicher Länge)
- Gesamttext²⁸

Ausgehend von dieser abstrakten Modellierung kann man damit zusammenfassend von den folgenden Teilinformationen ausgehen, die sich jeweils auf ein Token im Text beziehen:

- Die Information über die Wortform des Tokens, also die Abfolge der Zeichen; mit oder ohne Berücksichtigung der Groß-/Kleinschreibung.
- Das Lemma, also die unflektierte Grundform, wie man sie als Wörterbucheintrag finden würde.
- Die Wortart, also die grammatikalische Klasse (Substantiv, Verb, Adjektiv, Pronomen, etc.), gegebenenfalls und sofern relevant auch weitere morpho-syntaktische Informationen (Genus, Numerus, Casus)
- Die Bedeutung, also der semantische Gehalt des Wortes; repräsentiert beispielsweise über Zuordnung eines Wortvektors aus einem *Word Embedding Model* oder eines Synsets in WordNet.

²⁸ Die Frage der texteigenen Untergliederung in Kapitel und Absätze (in Erzähltexten und Essays), Szenen, Akte und Reden (in Dramen) oder in Strophen (in der Lyrik) bleibt hier zunächst unberücksichtigt, wäre aber grundsätzlich in das Modell integrierbar.

- Die Relationen, also insbesondere die syntaktische Rolle des Tokens und seine Beziehung zu anderen Tokens im Satz, wie etwa die Bestimmung als Prädikat oder die Auflösung der Referenz eines Pronomens.
- Die Sequenzinformation, also die syntagmatische Position des Tokens relativ zu anderen Tokens im Satz; die Position des Satzes relativ zu anderen Sätzen im Segment; und die Position des Segments im Text.
- Die Häufigkeit des Tokens, wobei die Häufigkeit im Satz, im jeweiligen Segment oder im Gesamttext gemeint sein kann; zudem kann sie als binäre, absolute oder relative Häufigkeit ausgedrückt werden.

Die meisten der hier beschriebenen abgeleiteten Textformate beruhen grundsätzlich auf einem tokenisierten und linguistisch annotierten Text. Ein solches annotiertes Textformat erlaubt die grammatikalische Disambiguierung bestimmter Wörter, die Zusammenfassung mehrerer unterschiedlicher Wortformen auf der Ebene ihres gemeinsamen Lemmas oder ihrer gemeinsamen Wortart und die Suche und Filterung auf Wortart-Ebene. Weitere Annotationsebenen, wie bezüglich der syntaktischen Funktion im Satz oder als semantische und morpho-syntaktische Beschreibung von Wörtern durch Wort-Vektoren, sind ebenso denkbar, aber deutlich aufwändiger und daher weniger verbreitet.

Abgeleitete Textformate können auf der Grundlage des skizzierten Verständnisses von Text solchermaßen definiert werden, dass man jeweils beschreibt, welche der genannten, unterschiedlichen Teilinformationen durch Annotation expliziert werden, welche vereinfacht oder entfernt werden, und welche erhalten bleiben. Außerdem ergeben sich aus diesem Verständnis der Textformate eine Reihe von Parametern eines Textformats. Diese Parameter können sich insbesondere auf die folgenden Punkte beziehen:

- Welche Strategie der Tokenisierung wurde angewandt, wobei insbesondere der Umgang mit zusammengesetzten Begriffen geklärt sein muss. Besteht ›New York Times‹ aus einem, zwei oder drei Tokens?
- Welche Informationen werden für jedes Token bereitgestellt: die Wortform, das Lemma, die Wortart, die syntaktische Rolle, eine Repräsentation der Semantik, oder eine Kombination dieser Informationen?
- Wird der Text in einzelne Tokens gegliedert oder bilden Folgen mehrerer Tokens (sogenannte N-Gramme) die Darstellungseinheit?
- In welchem Umfang bleibt die Sequenzinformation erhalten?

- Gibt es eine Aufteilung jedes Gesamtdokumentes in Segmente und/oder Sätze, und wenn ja, welche Länge (in Worten oder Sätzen) haben die Segmente?
- Schließlich: Handelt es sich um ein Textformat, das grundsätzlich das einzelne Dokument innerhalb eines Korpus als Einheit beibehält, oder werden die Dokumentengrenzen aufgelöst?

Durch die Festlegung der jeweils explizierten, reduzierten oder entfernten und erhaltenen Informationen, spezifiziert über die jeweils gewählten Parameter, ergeben sich eine Vielzahl möglicher Transformationen der Ausgangstexte in verschiedene abgeleitete Textformate. Unterschiedliche Textformate eignen sich dabei je unterschiedlich gut für bestimmte Analyseverfahren. Der folgende Abschnitt nimmt dieses Verhältnis genauer in den Blick.

IV. Verbreitete Analyseverfahren der Digital Humanities und ihr Informationsbedarf

Aus der Perspektive der Anwendung von TDM ist es wünschenswert, dass die abgeleiteten Formate einen möglichst geringen Informationsverlust gegenüber den Ausgangstexten erfahren. Dies soll es ermöglichen, dass auf Grundlage der abgeleiteten Formate möglichst weitreichende und vielfältige Analysemethoden eingesetzt werden können. Dabei ist je nach Analyseverfahren auch entscheidend, welche Arten von Informationen jeweils beibehalten oder reduziert wurden. Im vorliegenden Abschnitt werden daher mehrere in den DH übliche Verfahren in ihren Grundprinzipien knapp erläutert und speziell mit Blick auf ihre jeweiligen Anforderungen an den Informationsgehalt der abgeleiteten Textformate hin charakterisiert. Auf dieser Grundlage können dann verschiedene Textformate auf ihr Einsatzspektrum hin geprüft werden. Da unterschiedliche Analyseverfahren unterschiedliche Anteile der gesamten Textinformation nutzen, ist es wahrscheinlich, dass mehrere unterschiedliche Textformate notwendig sind, um eine breite Menge an Analyseverfahren zu unterstützen. Dies können jedoch auch nicht beliebig viele Formate sein, da sonst der Kuratierungsaufwand massiv ansteigt und zudem die Möglichkeit besteht, dass Ausgangstexte aus der Kombination der Formate rekonstruiert werden können.

Die hier diskutierten Analyseverfahren sind in den DH und insbesondere in den Computational Literary Studies (CLS) einschlägig, wobei die Darstellung keinesfalls Anspruch auf Vollständigkeit erhebt, sondern exemplarisch angelegt ist. Die folgenden Verfahren werden diskutiert:

- Die Klassifikation und Clustering von Texten u. a. für die *Autorschaftsattribuion*
- Die Extraktion distinktiver Merkmale
- Die semantische Analyse mit *Topic Modeling*
- Die Analyse von Polarität mit *Sentimentanalyse*
- Der Blick auf Figurenbeziehungen mit der *Netzwerkanalyse*
- Die Analyse von Beziehungen zwischen Texten beispielsweise beim *Text Re-Use*
- Sowie allgemein der Einsatz von Sprachmodellen für verschiedenste Aufgaben

1. Klassifikation und Clustering von Texten inkl. Autorschaftsattribuion

Überwachtes und nicht-überwachtes maschinelles Lernen (also Klassifikation und Clustering) werden im Bereich des Text Mining mit guten Resultaten durchgeführt, um Texte mit bestimmten Inhalten oder anderen semantischen Gemeinsamkeiten in größeren Textsammlungen zu finden. Dies erfolgt regelmäßig auf der Grundlage von Term-Dokument-Matrizen.²⁹ Eine besondere Anwendungsdomäne, bei der es um die Zuordnung von Texten zu ihren Autor*innen aufgrund von (lexikalischer, stilistischer etc.) Textähnlichkeit geht, ist die stilometrische Autorschaftsattribuion.³⁰ Grundprinzip ist hier die Erfassung des lexikalischen, insbesondere stilistischen Grades der Ähnlichkeit zwischen mehreren Texten. Die Ähnlichkeit kann auf der Ebene ganzer Texte, Teilsegmente oder Sätze ermittelt werden sowie auf der Grundlage der Häufigkeit von Wortformen oder anderen Merkmalen erfolgen. Da diese Verfahren in den allermeisten Fällen ohnehin auf einer Term-Dokument-Matrize operieren und keine Sequenzinformation berücksichtigen, können sie gut auch mit einfachen abgeleiteten Formaten unterstützt werden.

²⁹ Vgl. *Feldman/Sanger*, *The Text Mining Handbook. Advanced approaches in analyzing unstructured data*, Cambridge u. a. 2007.

³⁰ Vgl. *Stamatatos*, *A survey of modern authorship attribution methods*, in *Journal of the Association for Information Science and Technology* 60 (2009), H. 3, 538–556.

2. Extraktion distinktiver Merkmale

Ein weiteres Verfahren aus dem Methodeninventar der CLS ist die Extraktion distinktiver Merkmale. Hier geht es um die Identifikation von Wortformen oder anderen Merkmalen, die für einen Text oder eine Textgruppe im Vergleich mit einer anderen Textgruppe charakteristisch sind.³¹ Dies erlaubt es beispielsweise, die stilistischen und inhaltlichen Eigenheiten einer Autorin oder eines Autors, einer bestimmten Textsorte oder einer Epoche zu ermitteln und für weitere Analysen zu nutzen. Literaturwissenschaftliche Anwendungsbeispiele gibt es u. a. zu Shakespeare, dem britischen Roman oder dem französischen Drama.³² Sehr einfache Varianten dieses Verfahrens erfordern keine Sequenzinformation, sondern beruhen lediglich auf einem Vergleich der relativen Häufigkeit der Merkmale in den beiden Textgruppen. Etwas avanciertere Verfahren vergleichen lediglich die Verteilung der Häufigkeiten in den Texten der beiden Textgruppen und sind daher ebenfalls nicht auf Sequenzinformation angewiesen. Präzisere Verfahren benötigen allerdings Informationen zur Verteilung der Merkmale auch innerhalb der Texte (die sogenannte Dispersion). Diese Verfahren sind auf Sequenzinformation angewiesen, wobei eine geringe Segmentgröße zwar grundsätzlich wünschenswert ist, die Verfahren aber von jeder Segmentierung der vollständigen Texte in kleinere Teile profitieren. Für manche Varianten des Verfahrens ist die Information über die Reihenfolge der Segmente im Text nützlich, für andere ist sie nicht ausschlaggebend.

3. Topic Modeling

Topic Modeling ist ein Verfahren aus dem Bereich des unüberwachten Machine Learning mit dem latente, im weitesten Sinne semantische Strukturen in größeren Textsammlungen entdeckt werden können. Dabei werden aufgrund des wiederholten, gemeinsamen Vorkommens von Wörtern beziehungsweise

³¹ Vgl. Kilgarriff, Comparing Corpora, in *International Journal of Corpus Linguistics* 6 (2001), H. 1, 97–133.

³² Vgl. *Craig/Kinney*, Shakespeare, Computers, and the Mystery of Authorship, Cambridge u. a. 2009; *Schöch*, Zeta für die kontrastive Analyse literarischer Texte. Theorie, Implementierung, Fallstudie in Bernhart et al. (Hrsg.) *Quantitative Ansätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven (Scientia Quantitatis*, Hannover, 30.09.–02.10.2014), Berlin u. a. 2018, 77–94; *Hoover*, Teasing out Authorship and Style with t-tests and Zeta 2010, in *Digital Humanities 2010. Conference abstracts (DH 2010, London, 07–10.07.2010)*, London 2010.

dem wiederholten Vorkommen von Wörtern in ähnlichen Kontexten Gruppen von Wörtern gebildet, zwischen denen eine wie auch immer geartete semantische Beziehung besteht, wobei dann die Verteilung dieser Wortgruppen in der Textsammlung ermittelt werden kann.³³ Die semantische Beziehung der Wörter kann sich bei fiktionalen Texten u. a. auf ein abstraktes Thema (Gerechtigkeit, Fortschritt), ein wiederkehrendes erzählerisches Motiv (Eisenbahnfahrt, Konzertbesuch) oder das Vokabular für die Beschreibung von Handlungsorten (Innenräume, Landschaften) beziehen. Es gibt in den CLS zahlreiche Anwendungsbeispiele für diese Methode, sei es zu Tagebüchern, Romanen, Dramen oder Lyrik.³⁴

Topic Modeling profitiert auf jeden Fall von der Verfügbarkeit der Information zu Lemma und Wortart, weil mit dieser Information semantisch eng verwandte Wörter zusammengeführt und Funktionswörter sowie Namen besser herausgefiltert werden können, als wenn lediglich die Häufigkeitsinformation verfügbar ist. Zudem ist für Topic Modeling, zumindest bei umfangreicheren Texten wie Theaterstücken und insbesondere Romanen eine Segmentierung der Texte in kleinere Segmente notwendig, um möglichst präzise und semantisch kohärente Topics zu erhalten. Auch hier gilt allerdings, dass eine geringe Segmentgröße grundsätzlich wünschenswert ist, die Verfahren aber von jeder Segmentierung der vollständigen Texte in kleinere Teile profitieren. Ideal wäre für das Topic Modeling vermutlich eine Segmentbildung, bei der Sätze oder sogar Absätze nicht getrennt werden, aber auch eine Segmentierung in Textabschnitte willkürlicher Länge in Wörtern (beispielsweise Segmente einer Länge von 500 Wörtern) ist nützlich. Information zur genauen Position eines Wortes im Text ist nicht notwendig, da das Verfahren in der Regel ohnehin dem Bag-of-Words-Modell folgt. Allerdings ist die Information über die Position eines Segmentes im Gesamttext durchaus nützlich, weil dadurch Muster in der Topic-Prävalenz nach Position im Text (beispielsweise Textanfang vs. Textende) möglich werden.

³³ Vgl. *Blei*, Introduction to probabilistic topic models, in *Communications of the ACM* 55 (2011), H. 4, 77–84.

³⁴ Vgl. *Blevins*, Topic Modeling Martha Ballard's Diary, in *History.org*. Blogbeitrag vom 01.04.2010; *Jockers*, Macroanalysis – Digital Methods and Literary History, Urbana, IL u. a. 2013; *Rbody*, Topic Modeling and Figurative Language in *Journal of Digital Humanities* 2 (2012), H. 1; *Schöch*, Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama, in *Digital Humanities Quarterly* 11 (2017), H. 2.

4. Netzwerkanalyse

Die Netzwerkanalyse beruht in der Regel auf der Erkennung der Entitäten, die als Knoten des Netzwerks dienen sollen, etwa Figuren, sowie auf einem Kriterium, aus dem sich eine Relation zwischen den Entitäten ergibt, beispielsweise die Interaktion in einem Dialog oder die Erwähnung in einem bestimmten Segment innerhalb des Textes.³⁵ Die Durchführung von Netzwerkanalysen setzt dabei in der Regel ein hohes Maß an automatisierter oder händischer Vorverarbeitung eines Textes voraus. In Hinblick auf die Extraktion von Entitäten müssen z. B. Figuren mittels Verfahren der *Named Entity Recognition* identifiziert werden; ist eine höhere Präzision angestrebt, müssen zudem pronominale Bezüge mittels Verfahren der *Coreference Resolution* aufgelöst werden. Alternativ können hier – wie etwa bei dramatischen Texten in Gestalt der Sprecherangaben – vorgegebene Strukturinformationen aus den Texten übernommen oder aber eine händische Auszeichnung von Figuren vorgenommen werden. Die Extraktion von Relationen setzt bei avancierteren Verfahren die computerlinguistische Identifikation von Formen der Rede- und Gedankenwiedergabe, einschließlich Adressant- und Adressatenzuweisung voraus (etwa: Wer spricht mit wem? Wer denkt an wen?). Weniger avancierte Verfahren greifen auf vorgegebene Segmentierungen in den Texten zurück, etwa Szenenstrukturen in dramatischen Texten oder Kapitelstrukturen in narrativen Texten (welche Figuren werden in einer Szene/einem Kapitel genannt?). Alternativ werden stärker formale Segmente (etwa Absätze) oder willkürliche Segmentierungen (etwa eine bestimmte Anzahl von Wörtern) als Definitionsgrundlage für die Extraktion von Relationen verwendet.

Obwohl aktuelle Studien in der Netzwerkanalyse literarischer Texte vermehrt komplexe semantische und sprechaktpragmatische Informationen für die Extraktion und Spezifizierung von Relationen verwenden, basiert ein Großteil der bisherigen Forschung vor allem a) auf einer zuverlässigen Identifikation von Entitäten und b) auf einer basalen, recht großzügigen Segmentierung. Auf sehr großen Korpora lassen sich insofern bereits mittels sehr weniger Informationen abs-

³⁵ Vgl. Jannidis, Netzwerke, in Ders. (Hrsg.) *Digital Humanities: Eine Einführung*, Stuttgart 2017, 147–161; Trilcke, *Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft*, in Ajouri et al. (Hrsg.) *Empirie in der Literaturwissenschaft*, Münster 2013, 223–226 sowie 236–246.

trakte Strukturmodelle vergleichend untersuchen. Eine in diesem Sinne abstrakte Netzwerkanalyse erfordert lediglich die Information darüber, welche Named Entities in welchen großzügig gefassten Textsegmenten (von z. B. 500 Wörtern) vorkommen, sowie die Information über die Position der Segmente im Text, kann aber auf alle übrigen semantischen, syntaktischen oder positionalen Informationen verzichten. Als problematisch erweist sich jedoch das Verfahren der Named Entity Recognition, dessen Ergebnisse in der Regel insbesondere in Hinblick auf die Zuordnung unterschiedlicher Named Entities zu Figuren (z. B. Vater und Ehemann als eine Entität) keine befriedigenden Ergebnisse liefern und insofern regelmäßig die Überprüfung und Nachkorrektur anhand des Volltextes voraussetzen.

5. Sentimentanalyse

Ziel der Sentimentanalyse ist die Ermittlung der Polarität (positiv, neutral, negativ) eines Textabschnittes, beispielsweise eines Satzes oder Segmentes.³⁶ In Erweiterung des Paradigmas kann es auch um die Ermittlung der Prävalenz verschiedener Basisemotionen (Freude, Furcht, Trauer, Glück, etc.) in einem Textabschnitt gehen. Neben dem Fokus – Analyse von Sentiment oder von Emotionen – unterscheiden sich die verschiedenen Ansätze darin, ob für die Analyse Wörterbücher, manuelle Annotationen oder Verfahren des Machine Learning genutzt werden.³⁷

Eine Herausforderung für die Sentimentanalyse sind Phänomene wie Verneinung, deren Skopus wörterbuchbasierte Ansätze ermitteln und berücksichtigen müssen, um die korrekte Polarität oder Basisemotion zu ermitteln. Diese Phänomene sind in der Regel nur mit einer syntaktischen Analyse, in jedem Fall aber mit einer gewissen lokalen Sequenzinformation, in den Griff zu bekommen. Gleichzeitig sind sie in literarischen Texten vergleichsweise häufig, da diese ein hohes Maß an so genanntem uneigentlichem Sprechen aufweisen, das u. a. Metaphern, Sarkasmus und Ironie umfasst. Im Falle der manuellen Annotationen müssen hingegen die entsprechend annotierten Textabschnitte – also der Text, aus dem Quellenmaterial zusammen mit einer ihm zugewiesenen Annotation

³⁶ Vgl. *Liu*, *Sentiment analysis and opinion mining* (Synthesis lectures on human language technologies), San Rafael, CA 2012, insbesondere 5–101.

³⁷ Für eine Übersicht von Sentimentanalyse-Ansätzen in den CLS vgl. *Kim/Klinger*, *A Survey on Sentiment and Emotion Analysis for Computational Literary Studies*, in: *Zeitschrift für digitale Geisteswissenschaften* 4 (2019).

eines Sentiments beziehungsweise einer Emotion – zugänglich sein. Die Machine-Learning-Verfahren werden schließlich üblicherweise mindestens auf vollständigen Sätzen, die z. T. ebenfalls annotiert werden, entwickelt und eingesetzt.³⁸ Einfache abgeleitete Formate ohne lokale Sequenzinformation sind daher für dieses Verfahren vermutlich nicht geeignet.

Für alle Verfahren gilt außerdem, dass für die Weiter- oder Neuentwicklung eines Verfahrens mehr Textinformationen nötig sind, da sowohl die Wörterbucherstellung als auch das maschinelle Lernen zumeist über Annotationen des Textmaterials in Bezug auf seine Sentimentwerte erfolgt. Außerdem benötigen Verfahren der Sentimentanalyse Sequenzinformationen (Position von Satz oder Segment im Gesamttext), sobald sie Analysen über den Textverlauf vornehmen, etwa die Sentiment-Ausprägung eines Themas oder einer Figur im Verlauf des gesamten Textes oder als Kontraste zwischen Anfang und Ende eines Textes.

6. Text Re-Use

Bei Verfahren, die unter Text Re-Use zusammengefasst werden, geht es um die Identifikation von identischen oder sehr ähnlichen Passagen in mehreren Texten oder in einer umfangreichen Textsammlung, wie z. B. den Nachweis von Shakespeare-Zitaten in anderen Texten.³⁹ Dabei kann der Text Re-Use auch in verschiedenen Sprachen und über verschiedene Medien hinweg analysiert werden.⁴⁰

Aufgrund des zum Teil sehr unterschiedlichen Begriffs von Text Re-Use und durch die unterschiedlichen Textsorten, die in das Untersuchungskorpus integriert werden, unterscheiden sich die Verfahren stark voneinander.⁴¹ In den Grundzügen funktionieren die Verfahren aber so, dass sie nach Ähnlichkeiten in Texten suchen, indem sie diese paarweise vergleichen und Passagen ausfindig machen, die in beiden Texten identisch sind oder aber durch nur wenige Lösch-

³⁸ Vgl. *Pang/Lee*, *Opinion Mining and Sentiment Analysis*, Boston, MA u. a. 2008, 1–135.

³⁹ Vgl. *Hobl-Trillini/Quassdorf*, A >key to all quotations<? A corpus-based parameter model of intertextuality, in *Literary and Linguistic Computing* 25 (2010), H. 3, 269–286.

⁴⁰ Vgl. *Burghardt et al.*, >The Bard meets the Doctor< – Computergestützte Identifikation intertextueller Shakespearebezüge in der Science Fiction-Serie Dr. Who, in Sahle (Hrsg.) *DHd 2019 Digital Humanities: multimedial und multimodal. Konferenzabstracts*, 2019.

⁴¹ Für eine Übersicht, vgl. *Büchler et al.*, *Towards a Historical Text Re-use Detection*, in Biemann/Mehler (Hrsg.) *Text Mining: From Ontology Learning to Automated Text Processing Applications*, Heidelberg u. a. 2014, 221–238.

oder Ergänzungsoperationen ineinander überführt werden können. Das Konzept von Identität oder Ähnlichkeit kann dabei syntaktisch, lexikalisch und/oder semantisch bestimmt sein. Entsprechend reicht die Bandbreite der von Text Re-Use-Anwendungen aufgefundenen Passagen von wörtlichen Zitaten über die Nutzung bestimmter morphosyntaktischer oder syntaktischer Muster bis hin zu freien Paraphrasen oder gar der lexikalisch und syntaktisch kaum sichtbaren Übernahme von Gedankengängen.

Mit Blick auf geeignete abgeleitete Textformate ist allen Zugängen gemeinsam, dass die Sequenzinformation unabdingbar ist. Darüber hinaus basieren Text Re-Use-Zugänge je nach ihrer Modellierung auf erweiterten N-Gramm-Analysen, syntaktischen Analysen oder anderen hier beschriebene TDM-Verfahren wie Topic Modelling und Word Embeddings. Entsprechend sind weitere der oben aufgelisteten Parameter abgeleiteter Textformate einzeln oder kombiniert nötig.

7. Sprachmodelle/Word Embeddings

Zahlreiche der Verfahren, die in den Computational Literary Studies verwendet werden, entstammen der Computerlinguistik beziehungsweise dem *Natural Language Processing* (NLP). Seit 2017 hat sich in der Verarbeitung von natürlicher Sprache ein Ansatz als besonders erfolgreich erwiesen: die Erstellung großer Sprachmodelle mit tiefen neuronalen Netzen, die dann in einem letzten Schritt auf die spezifische Aufgabe abgestimmt werden.⁴² Diese Ansätze sind so erfolgreich, dass sie alle anderen weitgehend verdrängt haben, da diese Modelle in sehr hohem Maße Informationen über semantische, syntaktische und andere Aspekte von Sprache enthalten. Allerdings braucht man für ihre Erzeugung die ganzen Texte beziehungsweise zumindest vollständige Sätze. Im Gegenzug ist die Information, aus welchem Einzeltext ein Satz nun gerade kommt (oder zumindest die Information, in welcher Reihenfolge die Sätze in einem Text vorkommen), im Grunde unerheblich. Wichtiger ist, dass die Zusammensetzung der gesamten Textsammlung in ihren wesentlichen Parametern (u. a. Anteile der enthaltenen Textsorten, Anteile verschiedener zeitlicher Abschnitte) bekannt ist. Selbst wenn man so vorgehen würde, dass die großen Sprachmodelle von den Gedächtnisinstitutionen trainiert werden (wofür auch spricht, dass dieser

⁴² Beispielsweise BERT, vgl. *Devlin et al.*, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in Computing Research Repository (CoRR) in arXiv.org. Version 1 arXiv:1810.04805 vom 11.10.2018.

Schritt sehr zeit- und rechenintensiv ist), so würde der letzte Schritt, das Abstimmen auf das spezifische Problem beziehungsweise die spezifische Anwendungsdomäne, immer noch voraussetzen, dass man zumindest eine größere Menge ganzer Sätze für diesen Schritt zur Verfügung hat.

8. Zwischenfazit Analyseverfahren

In der Summe zeigt sich hier, dass man wohl zwei große Gruppen von Analyseverfahren unterscheiden kann, wenn man ihre Anforderungen an den Informationsgehalt der abgeleiteten Textformate zu Grunde legt: einerseits diejenigen Formate, die auf eine präzise, insbesondere auch lokale, Sequenzinformation der einzelnen Tokens im Textverlauf angewiesen sind (dazu gehören wohl die meisten Verfahren aus der Sentimentanalyse, einige Verfahren der Netzwerkanalyse, sicherlich fast alle Verfahren des Text Re-Use sowie das Erstellen von Sprachmodellen); und andererseits diejenigen Formate, die diese Art von Sequenzinformation nicht erfordern, aber von einer nicht zu großen Segmentlänge sowie der Information über die Reihenfolge der Segmente im Gesamttext profitieren (dazu gehören sicherlich die Autorschaftsattributionsverfahren, die Extraktion distinktiver Merkmale und das Topic Modeling, gegebenenfalls auch einige Verfahren der Netzwerkanalyse).

Über die Nützlichkeit für bestimmte Analyseverfahren in den Digital Humanities hinaus müssen abgeleitete Textformate weitere Kriterien erfüllen, damit ihre Erstellung und Publikation in der Praxis tatsächlich umgesetzt werden kann. Diese kommen insbesondere aus den Rechtswissenschaften und betreffen die urheberrechtliche Unbedenklichkeit der abgeleiteten Formate. Wann die abgeleiteten Textformate so gestaltet sind, dass die verbliebenen Informationen klar nicht mehr urheberrechtlich relevant sind, hängt von mehreren Faktoren ab, die von Karina Grisse und Florian Jotzo genauer diskutiert werden,⁴³ zu denen aber sicherlich Folgendes gehört:

- Es handelt sich weder um Vervielfältigungen noch um Bearbeitungen der Primärtexte im Sinne des Urheberrechtsgesetzes.
- Die Menge an zusammenhängendem Text liegt unter einer bestimmten Schwelle.

⁴³ Vgl. Grisse, Teil 3, B. in diesem Band; Jotzo, Teil 2, B. in diesem Band.

- Der Werkgenuss (im Sinne der ›normalen‹ Lektüre durch einen Menschen) ist ausgeschlossen.
- Der Grad der Wiedererkennbarkeit des Textes, insbesondere seiner individuellen ästhetischen Qualitäten, ist für Normalbürger*innen gering.
- Die Rekonstruktion des Textes (oder auch kleinerer, aber urheberrechtlich relevanter Passagen des Textes) ist nicht trivial und/oder mit Unsicherheiten behaftet.

Aus der Perspektive der Gedächtnisinstitutionen (also Bibliotheken, Archive, Museen, die digitale, urheberrechtlich geschützte Textbestände vorhalten) und damit aus Anbietersicht, sind abgeleitete Formate zudem dann geeignet, wenn sie leicht zu erstellen, zu speichern und vorzuhalten sind. Dies kann u. a. bedeuten, dass sie für jedes Textdokument unabhängig von anderen Dokumenten erstellbar sein sollten, damit ein inkrementeller Bestandsaufbau abgeleiteter Formate möglich ist. Dies ist bei den token-basierten Formaten der Fall, nicht aber insbesondere bei den auf Wort-Embeddings basierenden Formaten, die zur Erstellung umfangreiche Textbestände als Gesamtheit erfordern.⁴⁴

Aus Anwendungsperspektive ist sicherlich unabhängig von der gewählten Analyse­methode und dem spezifischen Format wünschenswert, dass abgeleitete Textformate in einem einfach zu verarbeitenden Datenformat vorliegen, was insbesondere bei einfach strukturierten und weit verbreiteten, standardisierten Formaten wie XML, JSON oder CSV der Fall sein dürfte. Außerdem sollten abgeleitete Textformate mit reichhaltigen, auch fachwissenschaftlich relevanten Metadaten, publiziert werden, sodass die Dokumentation der Provenienz der Texte und ihre fachwissenschaftliche Einordnung gewährleistet sind.

Ob diese Kriterien bei den hier vorgeschlagenen Textformaten jeweils gegeben sind, wird im folgenden Abschnitt kurz zusammengefasst dargestellt.

V. Vorschläge für abgeleitete Formate

In den folgenden Abschnitten wird eine Auswahl konkreter, abgeleiteter Textformate beschrieben und diskutiert. Diese Auswahl beruht auf einer vorgängigen Beurteilung einer größeren Anzahl von Formaten und beinhaltet nur solche Formate, die grundsätzlich aus den oben genannten Perspektiven zumindest vielversprechend erscheinen.

⁴⁴ Karina Grisse (*Grisse*, Teil 3, B. in diesem Band) geht in ihrem Beitrag ausführlicher auf die Frage der rechtlichen Einschätzung verschiedener Textformate ein.

Zur Veranschaulichung der Textformate werden im Text (sofern möglich beziehungsweise sinnvoll) jeweils Beispiele oder Ausschnitte der entstehenden Dateiformate abgebildet.⁴⁵ Die verwendeten Texte sind gemeinfrei, sodass die Ausgangstexte im Sinne der Transparenz der Transformationsverfahren mit publiziert werden können. Zur Illustration der Formate soll das erste Kapitel aus dem Roman *Effi Briest* (1894–95) von Theodor Fontane dienen; um einen direkten Vergleich mit dem Ausgangstext und eine Einschätzung bezüglich Werkzeugenuss und Wiedererkennbarkeit zu ermöglichen, sei der Anfang des Kapitels hier zitiert:

»In Front des schon seit Kurfürst Georg Wilhelm von der Familie von Briest bewohnten Herrenhauses zu Hohen-Cremmen fiel heller Sonnenschein auf die mittagsstille Dorfstraße, während nach der Park- und Gartenseite hin ein rechtwinklig angebauter Seitenflügel einen breiten Schatten erst auf einen weiß und grün quadrierten Fliesengang und dann über diesen hinaus auf ein großes, in seiner Mitte mit einer Sonnenuhr und an seinem Rande mit *Canna indica* und Rhabarberstauden besetztes Rondell warf. Einige zwanzig Schritte weiter, in Richtung und Lage genau dem Seitenflügel entsprechend, lief eine ganz in kleinblättrigem Efeu stehende, nur an einer Stelle von einer kleinen weißgestrichenen Eisentür unterbrochene Kirchhofsmauer, hinter der der Hohen-Cremmener Schindelturm mit seinem blitzenden, weil neuerdings erst wieder vergoldeten Wetterhahn auffragte.«⁴⁶

1. Tokenbasierte Textformate

Zunächst gehen wir auf tokenbasierte Textformate ein, die zugleich solche Formate sind, bei denen die Grundeinheit der Erstellung und Publikation in der Regel einzelne, vollständige Texte sind. Dies gilt für die anschließend vorgestellten Textformate, die auf N-Grammen oder Vektoren beruhen, nicht in gleicher Weise.

⁴⁵ Weitere Beispiele für die hier beschriebenen abgeleiteten Textformate sowie der sie erzeugende Programmcode liegen in einem Github-Repository (<https://github.com/dh-trier/tmr>, zuletzt abgerufen am 15.03.24) vor.

⁴⁶ *Fontane*, *Effi Briest*, in *Text Grid Repository*. Digitale Bibliothek, Göttingen 2012, Handle: 11858/00-1734-0000-0002-AF56-2, 7 vor.

a) Einfache Term-Dokument-Matrix

Das erste, sehr einfache abgeleitete Textformat ist die einfache Term-Dokument-Matrix. Sie besteht für jeden Einzeltext aus einer Liste der vorkommenden Tokens und ihrer absoluten Häufigkeit im Ausgangstext. Dabei kann zunächst für jeden Ausgangstext eine Datei erhalten bleiben (Tabelle 1). In der Praxis kann durch Zusammenführen mehrerer solcher Häufigkeitslisten eine ganze Textsammlung in Form einer Term-Dokument-Matrix repräsentiert werden, deren Größe von der Anzahl der enthaltenen Texte und der Anzahl der Types (d. h. der unterschiedlichen Wörter beziehungsweise des Gesamtvokabulars) bestimmt wird.

Aufgrund der prinzipiellen Einfachheit des Formats enthält es nur eine kleine Anzahl von Parametern, die es genauer spezifizieren:

- Welche Tokenisierung angesetzt wird, d. h. welche Definition von Token für die Segmentierung des Textes in einzelne Tokens, beispielsweise Wörter, verwendet wird.
- Welche Informationen auf Token-Ebene jeweils mitgeführt werden (und dafür auch erhoben werden müssen), d. h. ob lediglich die Wortform des Tokens, oder aber weitere Informationen über das Token – wie beispielsweise das Lemma, die Wortart, morphologische Information, die syntaktische Rolle im Satz oder eine Repräsentation der Wortbedeutung, beispielsweise als Wortvektor (siehe hierzu auch Abschnitt 5.2.2) –, angeboten werden.

Rang	Token (Wortform_Wortart_Lemma)	fontane_effi-briest
1	,_PUN_	10307
2	._PUN_.	5204
3	und_KON_und	4087
4	«_PUN_«	1937
5	»_PUN_»	1937

6	die_ART_die	1927
7	ich_PPEN_ich	1715
8	sie_PPEN_sie	1703
9	das_ART_der/die/das	1618
10	der_ART_der/die/das	1510
11	es_PPEN_es	1410
12	nicht_PTKNEG_nicht	1364
13	in_PRP_in	1102
14	so_ADV_so	1020
15	ist_VAFIN_sein	998
16	zu_KONJ_zu	983
...

Tab. 1: Ausschnitt aus der Term-Dokument-Matrix für Fontanes Effi Briest. Hier mit Wortform, Lemma und Wortart-Information, absteigend sortiert nach absoluter Häufigkeit. [Schöch et al. 2020]

Dieses abgeleitete Textformat kann folgendermaßen eingeschätzt werden:

- Für einige Analyseverfahren, insbesondere für einfache Varianten der Klassifikation und des Clustering beispielsweise für Fragen der Autorschaftsattribuion, und für einfache Distinktivitätsmaße ist das Format geeignet. Für viele andere Verfahren, darunter für Topic Modeling, Sentimentanalyse, Netzwerkanalyse oder Text Re-Use ist

dieses Textformat hingegen nicht ausreichend informationsreich: Insbesondere die vollständige Abwesenheit von Sequenzinformation auf allen Ebenen führt dazu, dass keine Verfahren eingesetzt werden können, die die (im Falle der Belletristik oft sehr umfangreichen) Texte nicht nur als Ganzes betrachten.

- Aus technischer Sicht ist sicherlich ein Vorteil dieses Textformats, dass es mit relativ trivialen Mitteln erstellt werden kann. Wie einfach das ist, hängt allerdings insbesondere vom oben genannten, zweiten Parameter ab. Denn die dafür jeweils notwendige linguistische Annotation ist nicht in allen Fällen trivial und in so gut wie keinem Fall gibt es nur eine einzige, standardisierte Vorgehensweise. Aus Anwendersicht ist zudem die einfache Nutzbarkeit eines solchen Formats ein Vorteil. Viele relevante Tools (u. a. Excel, Calc, R und Python) können eine solche Repräsentation in Form einer CSV-Datei direkt importieren und weiter verarbeiten.
- Aus rechtlicher Sicht ist die einfache Term-Dokument-Matrix ein ganz klar unbedenkliches Format. Eine Rekonstruktion des Ausgangstextes ist ebenso klar ausgeschlossen wie der Werkgenuss durch die Leser*innen oder auch nur die intuitive Wiedererkennbarkeit des Ausgangstextes. Dass die stilometrische Autorschaftsattribuion in der Lage ist, das individuelle stilistische Profil eines Autors aus einer solchen Matrix abzuleiten, bedeutet nicht, dass die individuellen Eigenschaften des Autors ohne technische Unterstützung erkennbar wären.
- Aus Anbietersicht schließlich ist das Format ebenfalls vergleichsweise unproblematisch, da es einfach erstellt werden kann, nach und nach Texte transformiert werden können und keine besonders umfangreichen Datenbestände entstehen.

In der Summe kann die Term-Dokument-Matrix demnach als rechtlich unbedenkliches, technisch eher unproblematisches, in der Anwendung aber eingeschränkt nützlich Format beschrieben werden. Es stellt damit in gewisser Weise die Baseline der abgeleiteten Formate dar.

b) Segmentweise Aufhebung der Sequenzinformation

Die Grundidee dieses abgeleiteten Formats ist es, die Reihenfolge der Wörter im Textverlauf durcheinanderzuwirbeln. Entscheidend ist hier allerdings, dass dies nicht für einen Einzeltext als Ganzes vorgenommen wird (dann wäre das Format bezüglich des Informationsgehalts mit der einfachen Term-Dokument-

Matrix identisch), sondern jeweils nur innerhalb kleinerer Segmente, wobei die ursprüngliche Reihenfolge dieser Segmente im Text aber beibehalten wird (Auszug 1). Es erfolgt also eine selektive Reduktion der Sequenzinformation. Die wesentlichen Parameter dieses Textformats sind wie bei den meisten Textformaten die Tokenisierung und die über das Token verfügbare Information. Wesentlich sowohl aus Anwendungs- als auch aus rechtlicher Perspektive ist hier allerdings der zusätzliche Parameter der Länge der Segmente in Tokens.

von_APPR_von Hohen-Cremmen_NN_Hohen-Cremmen Georg_NE_Georg
zu_APPR_zu
heller_ADJA_hell des_ART_die fiel_VVFIN_fallen schon_ADV_schon
bewohnten_ADJA_bewohnt In_APPR_in der_ART_die <SEG>
Mittagsstille_ADJA_Mittagsstille Gartenseite_NN_Gartenseite und_KON_und
erst_ADV_erst Park-TRUNC_Park- Dorfstraße_NN_Dorfstraße ,_PUN_
Seitenflügel_NN_Seitenflügel breiten_ADJA_breit die_ART_die hin_ADV_hin
während_KOUS_während angebauter_ADJA_angebaut der_ART_die
nach_APPR_nach
ein_ART_eine Schatten_NN_Schatten auf_APPR_auf einen_ART_eine
rechtwinklig_ADJD_rechtwinklig <SEG> großes_ADJA_groß ,_PUN_
auf_APPR_auf mit_APPR_mit in_APPR_in ein_ART_eine weiß_ADJD_weiß
und_KON_und
über_APPR_über quadrierten_ADJA_quadrierten und_KON_und diesen_PDAT_dies
auf_APPR_auf Mitte_NN_Mitte seiner_PPOSAT_sein dann_ADV_dann
Fliesengang_NN_Fliesengang hinaus_ADV_hinaus einen_ART_eine grün_ADJD_grün
<SEG>

Ausz. 1: Ausschnitt aus der Liste der Tokens mit Annotation bei segmentweiser Aufhebung der Sequenzinformation für den Beginn von Fontanes Effi Briest. Hier auf Unigramm-Basis und mit Wortform, Lemma und Wortart-Information sowie einer Segmentlänge von 20 Tokens. Man beachte die Markierung der Segmentgrenzen mit <SEG> nach jeweils 20 Tokens. [Schöch et al. 2020]

Bei diesem abgeleiteten Textformat gibt es keine Abhängigkeit zwischen den Texten in einer Textsammlung, sodass die Texte frei rekombiniert werden können. Die Größe der Segmente hat keine signifikante Auswirkung auf die Größe der resultierenden Dateien, weil nur die Reihenfolge der Merkmale verändert wird. Der Eingriff in die segmentübergreifende Textstruktur ist minimal, das lesende Verständnis des Textes erscheint aber schon bei sehr kleinen Segmentgrößen so gut wie ausgeschlossen.

Dieses abgeleitete Textformat kann folgendermaßen eingeschätzt werden:

- Aus der Anwendungsperspektive erscheint dieses Textformat für eine vergleichsweise große Anzahl von Analysemethoden nützlich, vorausgesetzt, die Segmentlänge wird nicht zu groß angesetzt (<50 Tokens wären sicherlich in einigen Szenarien ausreichend klein): für einfache stilometrische Verfahren auf jeden Fall, zudem auch für avanciertere Verfahren und das Ermitteln distinktiver Merkmale, wofür ein segmentierter Text erforderlich ist, um zu sampeln oder die Dispersion der Merkmale zu berücksichtigen.⁴⁷ Für Topic Modeling ist das Format ebenfalls gut geeignet. Nur bei einer sehr geringen Segmentlänge oder bei einer Segmentierung in Sätze erscheint eine einfache Sentimentanalyse denkbar. Verfahren der Netzwerkanalyse sind denkbar, wären aber auf eine vorgängige, hochwertige Named Entity Recognition und Coreference Resolution angewiesen. Mit diesem Format nicht durchführbar erscheinen avanciertere Verfahren des Text Re-Use, die stark auf einer feingranularen Sequenzinformation beruhen, oder nicht-triviale Verfahren aus dem Bereich der Sentimentanalyse.
- Aus technisch-informatischer Perspektive ist dieses Format unproblematisch, weil es einfach zu erstellen ist und keine besonderen Anforderungen an Speicherkapazitäten oder Datenstruktur erfordert. Es kann eine Datei pro Gesamttext erstellt werden, wodurch ein progressiver Bestandsaufbau ermöglicht wird; zudem erlaubt dies die einfache, nachträgliche Kombination von Texten zu einem je nach Forschungsfrage zusammengestellten Korpus.
- Aus rechtlicher Perspektive ist eine Rekonstruktion des Ausgangstextes mit einer so hohen Zuverlässigkeit, dass der Ausgangstext tatsächlich gelesen und verstanden werden könnte, schon bei einer Segmentlänge von >50 aufgrund der exponentiell steigenden Anzahl der möglichen Kombinationen kaum noch denkbar. Mit höherer Segmentlänge sinkt die Rekonstruierbarkeit weiter ab. Bei einer geringen Segmentlänge (beispielsweise <10 Tokens) oder bei einer satzweisen Segmentierung steigt sie hingegen; dann wäre eine Rekonstruierbarkeit des Ursprungstextes in einzelnen Fällen (also nicht für den Gesamttext, aber

⁴⁷ Nicht geeignet ist das Textformat für einen Spezialfall der stilometrischen Autorschafts-attribution, das sogenannte *rolling Delta*.

doch für mehrere längere Abschnitte des Textes) denkbar. Die genauen Verhältnisse wären allerdings erst empirisch nachzuweisen.

Damit handelt es sich hier um ein sehr empfehlenswertes Format, das bei entsprechend geeigneter Wahl des Parameters Segmentlänge (im Bereich von um die 50 Tokens) sowohl aus der Anwendungsperspektive für eine Reihe von Verfahren nützlich ist als auch aus rechtlicher Perspektive als unbedenklich eingeschätzt werden kann. Zu beachten ist zudem, dass das erste abgeleitete Textformat, die einfache Term-Dokument-Matrix, aus diesem Format ebenfalls generiert werden kann (nicht aber umgekehrt).

c) Selektiv reduzierte Information über einzelne Tokens

Die Grundidee dieses Formats ist es, die vollständige Sequenzinformation im Text beizubehalten, um bestimmte Verfahren zu ermöglichen, die auf diese Information angewiesen sind, dabei aber so viel Information über die einzelnen Tokens zu entfernen, dass dennoch von einer urheberrechtlichen Unbedenklichkeit ausgegangen werden kann. Zahlreiche Varianten sind denkbar, aber eine aus Anwendungssicht nützliche Implementierung dieses Textformats könnte folgendermaßen gestaltet sein: Ausgangspunkt wäre erneut ein tokenisierter und annotierter Text, sodass für jedes Token mindestens Wortform, Lemma und Wortart verfügbar sind. Dann wird beim Erstellen des Textformats aber beispielsweise für alle Funktionswörter (also u. a. Präpositionen, Pronomina und Artikel) die Information über die Wortform und das Lemma entfernt und lediglich die Information über die Wortart beibehalten (Auszug 2). Dadurch bleibt die Sequenzinformation vollständig erhalten, nicht nur in Bezug auf die Abfolge der Inhaltswörter, sondern auch in Bezug auf den exakten Abstand der Wörter zueinander im Ausgangstext. Wichtigster Parameter dieses Textformats ist sicherlich, für welche Wortarten die Information über Wortform und Lemma entfernt wird und für welche nicht.

APPR Front_NN_Front ART schon_ADV_schon APPR Kurfürst_NN_Kurfürst
Georg_NE_Georg Wilhelm_NE_Wilhelm APPR ART Familie_NN_Familie APPR
Briest_NN_Briest bewohnten_ADJA_bewohnt Herrenhauses_NN_Herrenhaus APPR
Hohen-Cremmen_NN_Hohen-Cremmen fiel_VVFIN_fallen heller_ADJA_hell
Sonnenschein_NN_Sonnenschein APPR ART Mittagsstille_NN_Mittagsstille
Dorfstraße_NN_Dorfstraße PUN KOUS APPR ART TRUNC KON
Gartenseite_NN_Gartenseite

hin_ADV_hin ART rechtwinklig_ADJD_rechtwinklig angebauter_ADJA_angebaut
 Seitenflügel_NN_Seitenflügel ART breiten_ADJA_breit Schatten_NN_Schatten
 erst_ADV_erst APPR ART weiß_ADJD_weiß KON grün_ADJD_grün
 quadrierten_ADJA_quadrierten Fliesengang_NN_Fliesengang KON dann_ADV_dann
 APPR
 PDAT hinaus_ADV_hinaus APPR ART großes_ADJA_groß PUN APPR PPOSAT
 Mitte_NN_Mitte
 APPR ART Sonnenuhr_NN_Sonnenuhr KON APPR PPOSAT Rande_NN_Rand APPR
 Canna_NN_Canna indica_NE_indica KON Rhabarberstauden_NN_Rhabarberstaude
 besetztes_ADJA_besetzt Rondell_NN_Rondell warf_VVFIN_werfen PUN

Ausz. 2: Abfolge der Tokens mit Annotation bei selektiver Entfernung der Wortform- und Lemma-Information für den Beginn von Fontanes Effi Briest. [Schöch et al. 2020]

Dieses abgeleitete Textformat kann folgendermaßen eingeschätzt werden:

- Dieses Textformat ist (in der beschriebenen Form) für die stilometrische Autorschaftsattributionskaum geeignet, weil für die Stilometrie gerade die feinen Unterschiede in den Häufigkeiten der einzelnen Funktionswörter entscheidend sind. Topic Modeling würde durch ein solches Format aber gut unterstützt, da hier meist ohnehin die Funktionswörter entfernt werden. Für die Ermittlung distinktiver Merkmale wäre das Verfahren nur geeignet, wenn es um die Ermittlung distinktiver Inhaltswörter oder distinktiver Wortarten geht. Für die Netzwerkanalyse ist auch dieses Format nur eingeschränkt nützlich, da zwar die Eigennamen von Personen und ihr Abstand im Text ersichtlich bleiben könnten, Informationen wie Koreferenz jedoch nicht rekonstruierbar sind. Avanciertere Verfahren der Netzwerkanalyse, die etwa die Zuordnung von Rede- oder Gedankenwiedergabe für die Extraktion und Spezifizierung von Relationen verwenden, sind nicht möglich.
- Für Verfahren wie den Text Re-Use hat das Format großes Anwendungspotential, denn Text Re-Use operiert ohnehin häufig mit N-Grammen, die auf den Lemmata der Inhaltswörter reduziert sind, um den *noise* zu reduzieren, der von kleineren stilistischen Varianzen produziert wird, und/oder auf die Inhaltswörter fokussiert ist. Einzig für die Sentimentanalyse wird auch dieses Format wenig gewinnbringend sein, weil vermutlich zu wenig syntaktische Information für die Berücksichtigung von Verneinungen u. ä. erhalten bleibt. Dies wäre allerdings je nach Parameter des Formats auch empirisch zu prüfen.

- Aus urheberrechtlicher Sicht erscheint hier problematisch, dass die Wiedererkennbarkeit des Textes aufgrund der Substantive und Eigennamen, die in der ursprünglichen Reihenfolge erhalten bleiben, vergleichsweise hoch ist, auch wenn von einem Werkgenuss wohl nicht die Rede sein kann. Dieser Effekt könnte durch das zusätzliche Entfernen der Eigennamen deutlich reduziert werden. Die Rekonstruierbarkeit erscheint für den korrekten Gesamttext kaum möglich, für kleinere Werkteile aber eventuell denkbar.

2. Textformate auf Korpus- oder Subkorpusebene

Die im vorigen Abschnitt verhandelten Textformate zeichnen sich alle dadurch aus, dass sie für jeden Einzeltext für sich genommen generiert werden können. Dies ist bei den folgenden Formaten anders, die den Einzeltext überschreiten können (bei den N-Grammen) beziehungsweise grundsätzlich unter Rückgriff auf ein umfangreicheres Korpus ermittelt werden (Wort-Embeddings).

a) N-Gramme

N-Gramme sind Sequenzen von mehreren aufeinander folgenden Tokens, ohne dass diese einer lexikalischen Einheit oder einer *Multi-Word Expression* entsprechen müssen. Im einfachsten Falle werden bei einem abgeleiteten Textformat, das auf N-Grammen beruht, die Häufigkeiten der in einem Text enthaltenen N-Gramme erhoben, ähnlich wie bei der einfachen Term-Dokument-Matrix (Abschnitt 5.1.1). Weil sie lokale Sequenzinformation beinhalten, sind N-Gramme als Hinweise auf Phänomene wie Kollokationen, Phraseme und andere lexikalisch-stilistische Muster für viele Analyseverfahren relevant. Aus diesem Grund wäre dieses Format besser als die bisher vorgestellten Formate für Text Re-Use geeignet.

Solange die Einheit des jeweiligen Einzeltextes nicht aufgelöst wird, dürfte allerdings aus urheberrechtlicher Perspektive selbst eine einfache Aufstellung der Häufigkeiten von N-Grammen der Größe 2–5 problematisch sein, weil durch die schindelartige Überlagerung mehrerer N-Gramme längere Textsequenzen rekonstruiert werden könnten. Dies gilt selbst dann als problematisch, wenn nicht der vollständige Text rekonstruiert werden kann, sondern nur eine grö-

ßere Menge von Fragmenten. Wenn die N-Gramm-Häufigkeiten sich auf kleinere Segmente innerhalb eines Textes beziehen, potenziert sich das Problem noch, weil die Rekonstruierbarkeit erleichtert wird. Im Falle des Formats, das auf der selektiv reduzierten Information über einzelne Tokens beruht (Abschnitt 5.1.3), sind allerdings verschiedenste N-Gramme indirekt enthalten, denn aus der ja vollständig vorhandenen, wenn auch nur lückenhaft mit Wortformen versehenen Tokensequenz lassen sich beliebig lange (allerdings wiederum nur teilweise mit Wortformen versehene) N-Gramme bilden.

Es ist allerdings auch möglich, sich vom Einzeltext als Bezugsgröße zu lösen und die N-Gramm-Häufigkeiten über mehrere beziehungsweise sehr viele Einzeltexte hinweg zu berechnen. Die Parameter eines solchen Formats sind (neben der Tokenisierung und Annotation) die N-Gramm-Länge und die Bezugsgröße für die N-Gramm-Häufigkeiten, beispielsweise jeweils alle Texte einer Textsorte und/oder eines Jahres (Tabelle 2).

Rang	N-Gramm	Häufigkeit
1	gott sei dank	43
2	ja gnädigste frau	17
3	auch heute wieder	13
4	doch auch wieder	11
5	ist doch auch	11
6	ist immer so	10
7	gnädigste frau ist	10
8	war so war	10
9	nein gnädigste frau	9

10	wird ja wohl	9
11	ist doch recht	9
12	doch immer noch	9
...

Tab. 2: Häufigkeiten von 3-Grammen über mehrere Texte hinweg, bei einer Mindesthäufigkeit von 5. Beispieldaten auf der Grundlage von fünf Erzähltexten von Theodor Fontane. [Schöch et al. 2020]

Ein solches Format kann folgendermaßen eingeschätzt werden:

- Aus Anwendungsperspektive ist das Einsatzspektrum eines solchen Textformats sicherlich geringer als bei den einzeltextbasierten N-Grammen. Immerhin sind solche Formate aber für bestimmte Fragestellungen und Anwendungen, die sich nicht auf den Einzeltext beziehen, immer noch informativ genug, da N-Gramme auch Informationen zum Sprachgebrauch enthalten. Solche Korpora wären bereits nützlich, um Einsichten in die sprachlichen Regeln bestimmter Felder zu gewinnen, z. B. welche Worte mit einer gewissen Wahrscheinlichkeit auf andere Worte folgen. Sie würden aber auch die Entwicklung und Verbesserung ganz praktischer Anwendungen, z. B. die Verbesserung von themenspezifischer Spracherkennung, unterstützen können. Sind die zugrundeliegenden Teilkorpora ausreichend spezifisch, ist auch die Extraktion distinktiver N-Gramme im Vergleich mehrerer Teilkorpora möglich.
- Die Rekonstruierbarkeit dürfte im Gegenzug deutlich eingeschränkt sein. Wenn nun Bibliotheken oder Archive sehr große Bestände etwa als 5-Gramme anbieten und dabei (wie Google) die N-Gramme des Korpus zählen, die in allen Büchern eines Jahres vorkommen, ist es sehr viel schwieriger, wenn nicht unmöglich, einen bestimmten Text oder auch nur längere Passagen beliebiger Texte aus den N-Grammen wiederherzustellen. Dies gilt insbesondere dann, wenn man dem Modell Googles auch in dem Punkt folgt, dass alle N-Gramme, die im

Gesamtkorpus eine bestimmte Mindesthäufigkeit nicht haben, auch nicht im Format enthalten sind.

- Eine Herausforderung aus Anbietersicht stellt hierbei die Frage dar, welche Aggregation von Einzeltexten innerhalb eines Gesamtkorpus (also z. B. alle digitalen Texte einer Bibliothek) für die Forschung relevant sind und an welchem Punkt eine rechtlich relevante Grenze überschritten wird: Neben der chronologischen Ordnung (jeweils die N-Gramm-Häufigkeiten aller Texte aus einem Jahr), die für Begriffs- und Ideengeschichte, aber auch Sprachgeschichte und andere historische Interessen brauchbar ist, könnte man sich auch andere Aggregationen vorstellen, die eher an Themen beziehungsweise Sachgruppen oder Textsorten orientiert sind (z. B. alle medizinischen oder auf die Wirtschaft bezogenen Texte). Es stellt sich dabei die Frage, wie klein die Gruppe sein kann und ob man sich eine Metrik vorstellen kann, die es einer Bibliothek leicht macht, zu entscheiden, ab welchem Punkt der Herstellbarkeit von längeren N-Gramm-Ketten die Bibliothek davon Abstand nehmen sollte. Löscht man die N-Gramme, die seltener vorkommen als ein bestimmter Schwellenwert besagt, dann kann man nicht alle, aber immer noch manche längeren N-Gramm-Ketten zusammensetzen, nämlich gerade da, wo häufig verwendete sprachliche Muster genutzt werden; eine entsprechende Metrik müsste also probabilistisch vorgehen.

b) Wort-Embeddings

Neben den tokenbasierten Textformaten und den N-Grammen spielen auch vektorbasierte Formate eine zunehmend wichtige Rolle. Die technische Entwicklung im Bereich der computergestützten Verarbeitung natürlicher Sprache (Natural Language Processing) hat aufgrund von vektorbasierten Textformaten seit etwa 2013 enorme Fortschritte zu verzeichnen. Obwohl die Ziele in der NLP-Forschung – hier geht es primär um die Mensch-Maschine-Interaktion – von den zuvor genannten Analyseverfahren der DH teilweise abweichen, werden die entwickelten Verfahren später oft für die DH angepasst oder weiterentwickelt. Wie bei Topic Modeling und Sentimentanalyse ist davon auszugehen, dass viele der vektorbasierten NLP-Verfahren, die derzeit noch wenig in den DH Anwendung gefunden haben, in Zukunft auch dort vermehrt eine Rolle spielen werden.

Die Grundidee vektorbasierter Textformate ist, Sprache nicht als symbolisches Zeichensystem zu betrachten, sondern Wörter und größere Einheiten wie Sätze oder Dokumente in einem algebraischen Vektorraum abzubilden. Man erhält so eine Informationsanreicherung der Wörter über das reine Symbol hinaus, da auch semantische und syntaktische Informationen im zum Wort gehörenden Vektor repräsentiert werden. Allerdings haben vektorbasierte Textformate auch einen offensichtlichen Nachteil, der gerade in den DH entscheidend sein kann: Durch die Umwandlung von Text in Vektoren gehen explizite, für qualitative Analysen oft entscheidende, Informationen verloren. Das Potential dieser Methoden für die Analyse von Textbeständen sowie als urheberrechtlich unbedenkliches Textformat ist aber naheliegend und soll im Folgenden skizziert werden.

Die erste Generation der vektorbasierten Textformate basiert auf dem Zählen des Auftretens von Wörtern im direkten Umfeld eines Wortes. Ein Wort wird somit als die Häufigkeit der anderen Wörter im Korpus repräsentiert und hat damit Ähnlichkeit zu den schon erwähnten Term-Dokument-Matrizen. Mit dem Word2vec-Verfahren wurden ab 2013 die Wort-Embeddings populär, die mit Verfahren des maschinellen Lernens Parametervektoren aus großen Textsammlungen schätzen.⁴⁸ Jedes Token im Vokabular wird demnach als ein Vektor von reellen Zahlen (üblicherweise wenige Hunderte) dargestellt und nicht mehr als Vektor von natürlichen Zahlen mit der Länge der Anzahl der Types im Korpus (mehrere Tausend). Allerdings haben nun die Werte eines Word Embedding Vectors keine explizit interpretierbare Bedeutung mehr. Wo bei den zählbasierten Wortvektoren jeder Eintrag die Häufigkeit des Auftretens eines Wortes im Umfeld repräsentierte, enthält ein Wort-Embedding latente Informationen über die Wahrscheinlichkeit des Auftretens von Wörtern im Umfeld. Diese Art von Wort-Embeddings ist somit komplementär zu den bereits erwähnten abgeleiteten Textformaten zu verstehen. Jeder unterschiedlichen Wortform, alternativ auch jedem unterschiedlichen Lemma, im Korpus wird exakt ein eindeutiger Wortvektor zugeordnet. Dieser repräsentiert die distributionale Semantik dieses Wortes in Bezug auf das gesamte Korpus. Es besteht somit eine global eindeutige Beziehung zwischen Vektor und Token.

⁴⁸ Vgl. *Mikolov et al.*, Distributed representations of words and phrases and their compositionality, in Burges et al. (Hrsg.) *Advances in Neural Information Processing Systems 25*. 26th Annual Conference on Neural Information Processing Systems, 4 Bände (NIPS'13: 26, Lake Tahoe, NV, 03.–06.12.2013), Red Hook, NY 2013, Bd. 2: 3111–3119.

In diesem Zusammenhang ergeben sich eine Reihe von denkbaren Szenarien, je nachdem, welche Art von Informationen im Rahmen eines abgeleiteten Textformats angeboten werden.

- Erstens könnte man alle Wortformen in den Ausgangstexten durch ihre Vektoren ersetzen und auch sämtliche Sequenzinformation beibehalten, allerdings um den Preis, dass jegliche Interpretierbarkeit des Textes unmöglich wird. Da bei einem solchen Format dennoch jede Wortform durch einen eindeutigen Vektor repräsentiert ist, kann beispielsweise stilometrische Autorschaftsattribuion damit weiterhin bewerkstelligt werden, mit der Einschränkung allerdings, dass die Merkmale nicht interpretierbar sind, weil die jeweils dazugehörige Wortform nicht vorliegt. Aus demselben Grunde wäre ein Verfahren wie Topic Modeling mit einem solchen Textformat zwar technisch möglich, aber wenig aufschlussreich. Urheberrechtlich dürfte das Format völlig unbedenklich sein, insbesondere wenn das Vokabular der so repräsentierten Texte nicht bekannt ist.
- Zweitens könnte man das Word Embedding Model als solches publizieren, also die Gesamtheit des Vokabulars einer Textsammlung mit ihren jeweiligen Wortvektoren. Dies ist urheberrechtlich ebenfalls unproblematisch, weil es keinerlei Bezug zu bestimmten Einzeltexten gibt. Allerdings handelt es sich hier dann in erster Linie um eine Ressource zur syntaktisch-semantischen Annotation von Texten, die auf einen geeigneten Textbestand im Sinne einer weiteren Annotationsschicht neben Lemmata und Wortarten angewandt werden könnte. Erstellt man solche Modelle (ähnlich wie für die N-Gramme vorgeschlagen) für verschiedene Subkorpora, kann der Vergleich der Modelle Einblicke in die Sprachentwicklung oder in die konzeptuelle Struktur bestimmter Textsorten bieten.⁴⁹
- Schließlich könnte man die oben beschriebenen Textformate über die Annotation nach Lemma und Wortart hinaus mit einer solchen syntaktisch-semantischen Annotationsschicht ausstatten. Eine gewisse Passung zwischen Word Embedding Model und zu annotierenden Texten ist dafür allerdings Voraussetzung. Urheberrechtlich würde dies

⁴⁹ Siehe z.B. *Hamilton et al.*, Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change, in Association for Computational Linguistics (Hrsg.) Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2 Bände (ACL: 54, Berlin, 07.–12.08.2016), Stroudsburg, PA 2016, Bd. 1: Long Papers, 1489–1501.

keinen entscheidenden Unterschied in der Beurteilung des jeweils in Frage stehenden tokenbasierten Textformats bedeuten; vorteilhaft wäre dies aber für verschiedenste Analyseverfahren, die so die Information über die semantischen und syntaktischen Ähnlichkeiten oder Unterschiede der Tokens nutzen könnten.

c) Kontextualisierte Embeddings

Der nächste essentielle Schritt zur Verbesserung bestehender NLP-Verfahren wurde durch das Kontextualisieren von Wort-Embeddings erreicht. Dabei wird Satz für Satz und Wort für Wort erst eine Ersetzung durch Wort-Embeddings durchgeführt, die danach jeweils individuell transformiert werden in Abhängigkeit der Worte die davor und danach in dem konkreten Satz auftreten. Die ersten Verfahren, die erfolgreich dafür eingesetzt wurden, sind rekurrente neuronale Netze, konkret die *Long-Short-Term-Memories*, und seit 2017 transformerbasierte Modelle, speziell das BERT-Modell.⁵⁰ Das Grundprinzip besteht darin, statt eines global statischen Vektors pro Type im Korpus einen individuellen Vektor für jedes Token in jedem bestimmten Satzkontext zu generieren. Wäre zuvor in zwei unterschiedlichen Sätzen, die beide ein Wort gemeinsam haben, dieses Wort durch denselben Vektor repräsentiert worden, ist bei kontextualisierten Embeddings jeder Wortvektor unterschiedlich, weil die umgebenden Wörter im Satz unterschiedlich sind. Damit hat jedes im Korpus auftretende Token prinzipiell eine individuelle Vektorrepräsentation und der Rückschluss von Vektor auf Wort ist nicht mehr trivial möglich.

Die kontextualisierten Embeddings haben damit zwei entscheidende Vorteile gegenüber den bisher dargelegten Textformaten:

- Die Rekonstruktion des ursprünglichen Textes, in dem alle Tokens durch ein kontextualisiertes Embedding ersetzt wurden, ist vermutlich nicht möglich, wenn das Mapping nicht für jedes einzelne Token explizit mit vorliegt. Um eine belastbare Aussage hierzu zu treffen, muss noch theoretische und empirische Forschung betrieben werden. Es lässt

⁵⁰ Siehe respektive *Devlin et al.*, BERT, in CoRR. Version 1 arXiv:1810.04805 vom 11.10.2018; *Hochreiter/Schmidhuber*, Long short-term memory in Neural computation 9 (1997), H. 8, 1735–1780 und *Vaswani et al.*, Attention is all you need, in Luxburg et al. (Hrsg.) *Advances in neural information processing systems* 30. 31st Annual Conference on Neural Information Processing Systems, 10 Bände (NIPS'17, Long Beach, CA, 04.–09.12.2017), Red Hook, NY 2017, Bd. 9: 6000–6010.

sich allerdings vermuten, dass Sätze ab einer gewissen Wortlänge (ca. >3) nicht rekonstruierbar sind. Dies gilt, ohne die Ursprungstexte in irgendeiner Art und Weise zu vereinfachen, also mit vollständigem Erhalt der Wortreihenfolge und Interpunktion. Dies wiederum hat erhebliches Potential für die Forschung auf Satzebene, u. a. zur Satzähnlichkeit und damit auch für Text Re-Use.

- Die in einem kontextualisierten Embedding beinhaltete syntaktische und semantische Information ist der anderer Repräsentationsformate deutlich überlegen. Die mit solchen Verfahren gewonnenen Ergebnisse auf Analyse-Benchmarks wie Sentimentanalyse, semantische Textähnlichkeit oder Paraphrasierung erreichen weit bessere Ergebnisse als bisherige Verfahren; oft übertreffen diese sogar menschliche Fähigkeiten von Nicht-Expert*innen.⁵¹

Vor diesem Hintergrund sind kontextualisierte Embeddings ein vielversprechender Kandidat für informationsreiche abgeleitete Textformate, die urheberrechtlich unbedenklich sind. Die Einschränkungen für die qualitative Forschung durch den Verzicht auf eine explizite Interpretierbarkeit der Worte bleiben aber auch hier bestehen. Damit befindet man sich mitten in der aktuellen Debatte über die Erklärbarkeit und Verlässlichkeit moderner Verfahren der künstlichen Intelligenz. Ebenso relevant wären kontextualisierte Embedding-Modelle beziehungsweise Transformer. Diese erlauben es, die kontextfreien und die kontextsensitiven Vektoren für Texte zu gewinnen, was zahlreiche Anwendungen in allen modernen digitalen textanalytischen Verfahren erlaubt.

⁵¹ Vgl. GLUE Benchmark (<https://gluebenchmark.com/>, zuletzt abgerufen am 15.03.24) und Wang et al., Superglue: A stickier benchmark for general-purpose language understanding systems, in Wallach et al. (Hrsg.) *Advances in neural information processing systems* 32. (NeurIPS 2019, Vancouver, 08.–14.12.2018), Red Hook, NY 2019, 3261–3275.

3. Zwischenfazit zu decn abgeleiteten Textformaten

Die Übersicht über einige denkbare abgeleitete Textformate zeigt, dass es durchaus mehrere vielversprechende Formate gibt, die in der Forschung nutzbringend eingesetzt werden können und die auch aus rechtlicher Sicht umsetzbar erscheinen. Eine kompakte Übersicht bietet Abbildung 1.⁵²

abgeleitetes Textformat	Nützlichkeit für die Forschung							rechtliche Beurteilung		
	Stilometrie	Distinktivität	Topic Modeling	Sentiment Analyse	Netzwerkanalyse	Text Re-Use	Sprachmodelle	Schutz vor Wiederkennbarkeit	Schutz vor Rekonstruierbarkeit	Unmöglichkeit des Werkzeugennusses
5.1.1. Einfache Term-Dokument-Matrix	+	+	0	-	-	-	-	+	+	+
5.1.2. Segmentweise aufgehobene Sequenzinf.	+	+	+	0	0	-	0	0	0	+
5.1.3. Selektiv reduzierte Information über Tokens	-	0	+	0	0	+	0	0	0	0
5.2.1. N-Gramme auf Teilkorpus-Ebene	-	0	-	-	-	-	+	0	0	+
5.2.2. einfache Wortembeddings	+	-	-	-	-	-	+	+	+	+
5.2.3. kontextualisierte Embeddings	0	0	0	0	0	0	+	+	+	+

Abb. 1: Übersicht über die abgeleiteten Textformate, ihre Nützlichkeit in der Forschung und ihre urheberrechtliche Einschätzung. [Hinzmann/Schöch 2020]

Eine linguistische Annotation zumindest mit der Information über das Lemma und die Wortart erscheint immer wünschenswert und aus rechtlicher Sicht unproblematisch. Die Nutzung eines geeigneten Word Embedding Models im Sinne einer semantisch-syntaktischen Annotation ist ebenso von Vorteil. Der Parameter der Segmentlänge wird voraussichtlich eine Abwägungsfrage bleiben:

⁵² Diese Übersicht kann lediglich der Orientierung dienen. Die exakte Einschätzung für jedes Format und jedes Verfahren bzw. für verschiedene urheberrechtliche Aspekte hängt auch von den jeweiligen Parametern des Formats ab und muss dem Text entnommen werden.

Aus der Perspektive der Analyseverfahren sind kleine Segmentlängen grundsätzlich besser (nicht zuletzt, weil eine Aggregation auf größere Segmente immer möglich, eine Aufteilung in kleinere Segmente hingegen im Nachhinein nicht möglich ist). Aus rechtlicher Sicht steigt aber in der Regel die Sicherheit, mit der ein Format urheberrechtlich irrelevant ist, mit größerer Segmentlänge an.

Nicht alle denkbaren abgeleiteten Textformate sind hier diskutiert worden. Einige sind rechtlich klar bedenklich, andere für die Forschung wenig nützlich. Urheberrechtlich problematisch sind insbesondere Formate, die längere Folgen von Worten beinhalten, also Formate, bei denen die lokale Sequenzinformation nur teilweise aufgehoben ist. Darunter fallen recht klar Formate, die aus vollständigen Sätzen in durcheinandergewirbelter Form bestehen oder aus einem Sample vollständiger Sätze. Beide Formate sind aus der Sicht der Analyseverfahren, die lokale Sequenzinformation erfordern, wie beispielsweise Sentimentanalyse oder Text Re-Use, besonders attraktiv, auch wenn es in den Digital Humanities häufig erforderlich erscheint, dass Analysen auf dem vollständigen relevanten Sprachmaterial beruhen; sie sind aber aus rechtlicher Hinsicht eher problematisch.

Nicht verschwiegen werden sollen einige inhärente Nachteile, die mit dem Modell der abgeleiteten Textformate verbunden sind und die ebenfalls in der Übersicht deutlich werden. Für diese sind sicherlich geeignete Minimierungsstrategien zu entwickeln. Dazu gehört erstens die nicht vollständige Nachvollziehbarkeit der Forschung, weil Analyseprozesse nicht vom Ursprungsmaterial aus nachvollzogen werden können, sondern nur vom verwendeten, abgeleiteten Textformat aus. Dies ist einer der Gründe, warum die Erstellung der abgeleiteten Textformate ein standardisierter und zertifizierter Prozess sein sollte, der die notwendige Vertrauenswürdigkeit der Formate garantiert. Ein weiterer Nachteil ist zweifellos, dass zwar einige, aber eben nicht alle relevanten Analyseverfahren auch mit einem abgeleiteten Textformat umgesetzt werden können. Schließlich ist in Rechnung zu stellen, dass die Erstellung von standardisierten, zertifizierten Beständen an Texten in abgeleiteten Formaten für die Anbietenden mit einem erheblichen Aufwand verbunden sind. Dennoch überwiegen aus unserer Sicht die Vorteile dieser Strategie gegenüber den (eingangs genannten) Alternativen beziehungsweise ist diese Strategie in jedem Fall eine wichtige, komplementäre Maßnahme neben den alternativen Ansätzen. Dies gilt nicht zuletzt auch, weil mit solchen Textbeständen besser als bisher demonstriert werden könnte, welches Potential in der Analyse urheberrechtlich geschützter Textbe-

stände in den DH liegt. Im Kontext eines gesellschaftlichen und rechtlichen Interessenausgleichs zwischen Rechteinhabenden und Anwender*innen von TDM kann dies ein wichtiges Argument sein.

VI. Fazit: Eine Forschungsagenda für abgeleitete Textformate

Abschließend soll hier ein Fazit insbesondere zu der Frage formuliert werden, welche Aufgaben und nächsten Schritte auf verschiedene relevante Akteure zukommen, wenn es gelingen soll, die abgeleiteten Textformate in der Praxis der DH zu verankern. Hier wird insbesondere auf die Rolle der Bibliotheken und Archive, der Informatik und der TDM-Anwender*innen in den DH eingegangen.

1. Bibliotheken und Archive

Bibliotheken und Archiven kommt eine zentrale Rolle bei der Etablierung abgeleiteter Textformate zu, weil sie rechtmäßigen Zugang zu umfangreichen urheberrechtlich geschützten Textbeständen haben. Schon mit dem UrhWissG wurden sie im UrhG als Institutionen benannt,⁵³ die urheberrechtlich geschützte Korpora, die von Dritten angefertigt wurden, aufbewahren und selektiv, nämlich zur Überprüfung wissenschaftlicher Qualität, verfügbar machen dürfen und sollen. Die Umsetzung des 2019 verabschiedeten Art. 3 DSM-RL bis Juni 2021 in nationales Recht⁵⁴ stärkt die Rolle dieser Einrichtungen im Kontext des TDM nochmals enorm und dies in mehrfacher Hinsicht: Bibliotheken und Archive werden nun ausdrücklich und autonom als Akteure des TDM adressiert und für wissenschaftliche Zwecke privilegiert; zugleich wird das Nachnutzungsverbot für die Korpora für Anschlussforschung aufgehoben werden.⁵⁵ Die Frage der wissenschaftlichen Anschlussforschung mittels abgeleiteter Formate tritt also absehbar als eine von dann zwei Säulen neben die Frage der wissenschaftlichen Anschlussforschung an Korpora.⁵⁶ Das lässt eine kohärente,

⁵³ Vgl. §§ 60d Abs. 3 S. 2 iVm 60e, 60f UrhG.

⁵⁴ Nach derzeitigem Stand voraussichtlich in §§ 44b, 60d UrhG-E (Entwurfssfassung).

⁵⁵ Vgl. *Döbl*, Digital Turn – Gedächtnisinstitutionen und Digital Humanities. Zwischenbericht einer Workshop-Reihe der Deutschen Nationalbibliothek, in *Zeitschrift für Bibliotheks- und Bibliographie* 67 (2020), H. 3–4, 213–230.

⁵⁶ Vgl. *Döbl*, Game Changer für Gedächtnisinstitutionen und Digital Humanities? Herausforderungen des neuen Rechts auf wissenschaftliche Nachnutzung von Korpora bei Text

in sich stimmige Entwicklungsstrategie der Bestände von der Warte des TDM aus in den Bereich des Möglichen rücken. Und beides ist verbunden mit einem nun expliziten Auftrag des Gesetzgebers an die Bibliotheken und Archive, sich jedenfalls insoweit mit dem Bereich TDM aktiv auseinanderzusetzen. Denn neben Forschungsorganisationen werden insofern nur Kulturerbeeinrichtungen privilegiert, anders als etwa andere Behörden, Bildungsträger, Journalisten oder Privatwirtschaft. Das ist eine Chance für Bibliotheken und Archive, aber zugleich auch eine Verantwortung.

Es bleibt in diesem Kontext im anstehenden neuen Recht bei der nachvollziehbaren, derzeit impliziten, dann aber auch ausdrücklich im Gesetz stehenden Einschränkung, dass die Nutzung der Korpora einen rechtmäßigen Zugang voraussetzt und in bestimmten Fällen sogar nur in den Räumen der privilegierten Einrichtungen gewährt werden darf. Bibliotheken und Archive werden künftig im Kontext von TDM eine aktivere Rolle in Forschungsprozessen einnehmen, wodurch sich die Funktionen der Einrichtungen selbst langfristig verändern werden.

Bibliotheken und Archive sehen es stets und unverändert als ihre genuine Aufgabe an, die Bestände in einer jeweils zeitgemäßen Form zur Nutzung zur Verfügung zu stellen. Die Bibliotheken und Archive können demnach durch die neuen rechtlichen Rahmenbedingungen und sobald geeignete Spezifikationen für grundsätzlich nützliche abgeleitete Textformate vorliegen, in die Rolle der Anbieter von entsprechenden Datensätzen für die Forschung treten. Abgeleitete Formate haben das Potential, sich hier zu einem wichtigen Arbeits- und Angebotsbereich im Kerngeschäft von Bibliotheken und Archiven zu entwickeln. Abgeleitete Textformate fördern die Perspektive auf eine globale, virtuelle Sammlung für die Forschung, die als Linked Open Data ausgeformt werden kann. Auch daraus ergeben sich weitreichende Konsequenzen für die Rolle, die Funktionen und das institutionelle Selbstverständnis dieser Einrichtungen.⁵⁷ Neben der Erstellung und Bereitstellung der abgeleiteten Textformate selbst wird es auch notwendig sein, für die Forschenden in den DH relevante Metadaten über die Texte bereitzustellen oder Wege zu entwickeln, wie diese erhoben werden können. Denn für die Forschung sind oft andere, forschungsnähere oder fachnähere Metadaten notwendig als diejenigen, die üblicherweise aus bibliothekarischer Perspektive erhoben würden.

und Data Mining, Art. 3 DSM-RL und §§ 44b, 60d UrhG-E, in Recht und Zugang 1 (2020), H. 2.

⁵⁷ Vgl. Wissenschaftliche Bibliotheken 2025. Hg. von Deutscher Bibliotheksverband e.V. Sektion 4 »Wissenschaftliche Universalbibliotheken«, München 2018.

Eine besondere Herausforderung bedeutet die Dokumentation einzelner Analysen unter wissenschaftlichen Anforderungen. In einem Bestand abgeleiteter Daten, der sich durch diverse technische und fachliche Einflüsse in einem fortwährenden Optimierungsprozess⁵⁸ befindet, muss eine besondere Herausforderung auf die Referenzierbarkeit einzelner Datensätze über persistente Identifikatoren sowie deren Echtheitsnachweis und Integrität zukommen. Grundsätzlich bestehen hohe Anforderungen an die Dokumentation, Transparenz und Verlässlichkeit des Prozesses der Erstellung von abgeleiteten Textformaten. Denn die Forschenden, die ein abgeleitetes Textformat für ihre Analysen einsetzen möchten, müssen sich darauf verlassen können, dass die abgeleiteten Textformate wirklich exakt im jeweils dokumentierten Verhältnis zu den Ausgangstexten stehen. Eine direkte Einsicht durch die Nutzenden ist bei den urheberrechtlich geschützten Ausgangstexten im Regelfall ja gerade nicht möglich. Daraus ergibt sich die Notwendigkeit, abgeleitete Textformate durch detaillierte Spezifikationen zu beschreiben, deren Einhaltung entsprechend geprüft werden kann. Dies schließt die jeweils gewählten Konzepte, Verfahren und Werkzeuge für Tokenisierung, Lemmatisierung, Wortartauszeichnung (einschließlich des jeweils verwendeten Sprachmodells und des Tagsets) mit ein, die allerdings sprachspezifisch beziehungsweise auch sprachstufenspezifisch sind. Diese Spezifikationen sollten von der Anbieterseite gemeinsam mit den TDM-Anwender*innen im Sinne eines Community-Standards entwickelt werden. Der Programmiercode, der für den Verarbeitungsprozess von den Ausgangstexten zu den abgeleiteten Textformaten eingesetzt wird, muss darüber hinaus öffentlich und frei zur Verfügung stehen, damit die Nachvollziehbarkeit und Nachnutzbarkeit des Codes garantiert ist. Ideal wäre es, wenn die Pipeline zur Erstellung der Textformate institutionenübergreifend entwickelt, gepflegt und geprüft würde, sodass es auch hier zu einer Standardisierung kommt. Dann könnte auch eine Dokumentation gemeinsam entwickelt werden, die die Verarbeitungsschritte in Prosa erläutert und anhand urheberrechtlich unbedenklicher Textbestände demonstriert. Eine solche Standardisierung der Prozesse und Formate

⁵⁸ Unter Optimierungsprozess ist beispielsweise die Erweiterung des Datenbestandes, die Optimierung von Algorithmen oder die qualitative Anhebung der Daten durch verbesserte Software für die optische Zeichenerkennung zu verstehen.

wird es dann auch erlauben, Bestände abgeleiteter Textformate aus unterschiedlichen anbietenden Institutionen für bestimmte Forschungsvorhaben zu kombinieren.

Aus den genannten Punkten ergeben sich sicherlich auch infrastrukturelle und institutionelle Herausforderungen mit Bezug auf die Personalentwicklung (Data Curation, Rechts- und IT-Expertise), wie sie derzeit vermutlich nur bei größeren Gedächtnisinstitutionen vorhanden sind. Auch bestimmte Rechte und Pflichten der beteiligten Akteure werden in diesem Kontext von Relevanz sein, insbesondere betrifft dies die Service anbietenden Bibliotheken und Archive und die TDM praktizierenden Nutzenden. Hierbei gilt es, unter Berücksichtigung von fachspezifischen, bibliothekarischen, wirtschaftlichen und juristischen Belangen, für alle beteiligten Akteure transparente Rahmenbedingungen zu schaffen. Diese können in einer Data Governance münden und darüber hinaus können inhaltliche Entwicklungen durch Policies gesteuert werden.

2. Informatik und Computerlinguistik

Im Zusammenhang mit den abgeleiteten Textformaten ergeben sich auch für die Informatik und Computerlinguistik neue Tätigkeitsfelder. So kann die Computerlinguistik sicherlich dabei mitwirken, standardisierte und zukunftsfähige Textformate sowie insbesondere standardisierte und zertifizierbare Pipelines für das Erstellen von abgeleiteten Textformaten zu entwickeln. Darüber hinaus könnten diese Fächer in den Bereichen Benchmarking und Rekonstruierbarkeit tätig werden.

Die empirische Überprüfung der Wahrscheinlichkeit beziehungsweise des Grades, mit der ein Ausgangstext auf der Grundlage eines abgeleiteten Textformats rekonstruiert werden kann, wird hier sicherlich eine wichtige Aufgabe der Informatik sein. Welcher Anteil eines Textes kann rekonstruiert werden? Wie lang sind die Fragmente, die auf diese Weise entstehen? Mit welcher Wahrscheinlichkeit sind die rekonstruierten Fragmente auch tatsächlich korrekt? Neben solchen Fragen ist hier insbesondere relevant, dass eine Rekonstruktion auch nicht durch die Kombination von Informationen aus mehreren unterschiedlichen abgeleiteten Textformaten möglich sein sollte. Für die Beantwortung dieser Fragen sind gemeinfreie Textbestände, bei denen Ausgangstexte und abgeleitete Formate gemeinsam vorliegen, einsetzbar beziehungsweise erforderlich.

In Kooperation von Informatik oder Computerlinguistik mit den Anwender*innen in den DH sollten Benchmarking-Analysen und vergleichende Un-

tersuchungen durchgeführt werden. Solche Analysen können auf der Grundlage von Datenbeständen, die Ausgangstexte und abgeleitete Textformate gleichermaßen beinhalten, überprüfen, wie groß die Unterschiede in der Performance bestimmter Analysemethoden sind, wenn man sie auf den Ausgangstexten einerseits, verschiedenen abgeleiteten Textformaten andererseits, einsetzt. Derartige Analysen wiederum sind schon jetzt von großem Interesse für die Definition und Spezifikation der abgeleiteten Textformate und können auch auf urheberrechtlich unbedenklichen (beispielsweise gemeinfreien) Beständen durchgeführt werden, bevor eine verlässliche rechtswissenschaftliche Einschätzung eines bestimmten Textformats vorliegt.

3. Digital Humanities

Eine wesentliche Aufgabe der DH als die am unmittelbarsten betroffenen Stakeholder wird es sein, eine enge Abstimmung zwischen den verschiedenen Akteuren und Communities sicherzustellen. Denn nur, wenn ein mit allen relevanten Akteuren (insbesondere aus DH, Rechtswissenschaften, Gedächtnisinstitutionen und Informatik) abgestimmtes Inventar von abgeleiteten Textformaten entwickelt wird, können auch entsprechend standardisierte und zertifizierte Pipelines für das Erstellen der Textformate entwickelt und angeboten werden. Diese tragen wiederum entscheidend zur Verlässlichkeit und Nützlichkeit der in dieser Form angebotenen Textbestände für die Forschung in den DH bei. Mit dem vorliegenden Beitrag haben wir das Ziel verfolgt, erste Schritte zu einer solchen Konsensbildung innerhalb der Fachcommunity zu gehen. Anliegen war uns demnach, die Grundidee der abgeleiteten Textformate und die damit einhergehenden Chancen und Herausforderungen insbesondere in der Community der DH sowie der Bibliotheken und Archive vorzustellen und eine Diskussion über sie anzuregen. Wir hoffen, dass unser Beitrag zeigen konnte, dass das Erstellen, Veröffentlichen und Nutzen solcher abgeleiteten Textformate grundsätzlich möglich und wünschenswert ist. Darüber hinaus hoffen wir, dass es gelungen ist, der Realisierung solcher Datenbestände näher zu kommen und die nächsten notwendigen Schritte unterschiedlicher Akteure herauszuarbeiten.

Teil 4: Gesetzlich erlaubte Handlungen

A. Die Text und Data-Mining-Schranken und ihr Rahmen für Textanalysen in den Digital Humanities

Katharina Erler-Fridgen

Version 1.0 (02.06.2022), CC BY-SA 4.0.

Werden Informationen aus urheberrechtlich geschützten Texten¹ extrahiert und zu diesem Zweck Quellen, beispielsweise aus Datenbanken,² gesammelt, aufbereitet und die Ergebnisse der Textanalyse sowie die Ausgangstexte schließlich aufbewahrt, so werden Vervielfältigungshandlungen oder Entnahmen vorgenommen, die einer urheberrechtlichen Gestattung bedürfen. Die Text und Data Mining-Schranken in § 44b UrhG und § 60d UrhG schaffen hierfür einen Rahmen, der im Folgenden erläutert werden soll.

I. Hintergrund des Text und Data Mining in den Digital Humanities³

Text und Data Mining wird eingesetzt, um Informationen aus Texten zu gewinnen. Verfahren des Text und Data Mining können etwa das Information Retrieval⁴, die Named Entity Recognition⁵ oder das Topic Modeling⁶ sein.⁷ Beispielsweise können mit diesen Verfahren Themenaussagen in Sekundärliteratur identifiziert oder eine Extraktion von Figuren und Handlungsorten aus Romanen vorgenommen werden. Dabei können die zugrundeliegenden literarischen

¹ Zu den Kriterien der urheberrechtlichen Schutzfähigkeit siehe *Erler-Fridgen*, Teil 2, A. in diesem Band.

² Zum Schutz von Datenbanken und der Entnahme hieraus siehe *Erler-Fridgen*, Teil 2, D. in diesem Band.

³ Dank gilt Prof. Dr. Christof Schöch und Dr. Maria Hinzmann (MiMoText, Universität Trier) für den Austausch zum Thema und die Bereitstellung und Diskussion von Beispielen.

⁴ *Ignatow/Mihalcea*, Text Mining, 137 ff.

⁵ *Neumann*, 5.3 Text-basiertes Informationsmanagement, in K.-U. Carstensen, Ch. Ebert, C. Ebert, S. Jekat, R. Klabunde, H. Langer (Hrsg.), *Computerlinguistik und Sprachtechnologie*, 3. Aufl., 596 ff.

⁶ *Blei*, Probabilistic Topic Models, <http://www.cs.columbia.edu/~blei/papers/Blei2012.pdf>, zuletzt abgerufen am 20.07.2021.

⁷ Zu unterschiedlichen Verfahrensmethoden siehe *Schöch*, Quantitative Analyse, in Jannidis/Kohle/Rehbein (Hrsg.), *Digital Humanities*, 279 ff.

Quellen aus Datenbanken heraus eingesetzt worden sein. Außerdem sind zur Umsetzung der Informationsextraktion regelmäßig zahlreiche Kopierschritte notwendig: Zunächst muss häufig ein Scan vorgenommen werden oder eine bereits vorliegende digitale Ressource eingesetzt werden. Die dann vorliegende Datei (Bild/PDF-Datei) wird mittels Optical Character Recognition (OCR)⁸ oder dem Double Keying Verfahren⁹ in eine auslesbare Textdatei umgewandelt. In der üblicherweise hieran anschließenden Phase des Preprocessings werden die Texte oftmals korrigiert, vereinheitlicht, in ein geeignetes Datenformat transformiert und gegebenenfalls auch segmentiert oder gefiltert. Weiterhin kann die Datei in ein Annotationstool importiert, dort bearbeitet und dann exportiert werden. Schließlich können quantitative Verfahren wie die Named Entity Recognition oder Verfahren des maschinellen Lernens angewandt werden.

Für die Digital Humanities können im Bereich des Text und Data Mining unterschiedliche Arten von Textquellen interessant sein. Zum einen können für die Digital Humanities Textsammlungen hilfreich sein, die durch die Universitätsbibliothek etwa über Lizenzen bereitgestellt werden. Ein Beispiel hierfür wäre das Corpus of Contemporary American English (COCA)¹⁰, das aus Sicht der Digital Humanities für die Analyse auch mit Verfahren des Text und Data Mining geeignet ist. Zum anderen gibt es im Internet zugängliche Textbestände wie beispielsweise das wissenschaftliche Blogportal Hypotheses.¹¹ Hier sind die Texte teilweise über Creative-Commons lizenziert. Im Fall der Sammlung literarischer Texte auf der Plattform Teatro Espanol del Siglo de Oro (TESO)¹² sind weiterhin gemeinfreie Texte vorgehalten; der Download aller oder Teile des Bestands, etwa zur Erstellung einer Sammlung, werden in der Lizenz der Sammlung untersagt.

⁸ Verfahren zur Textdigitalisierung, in dem der Prozess der Texterfassung automatisiert vorgenommen wird, siehe <http://digitalhumanities.berkeley.edu/resources/digitization-workflows-scanning-ocr-and-audio-transcription>, zuletzt abgerufen am 03.08.2021.

⁹ Double Keying Verfahren: Double Keying beschreibt ein Verfahren zur Textdigitalisierung, in dem der Prozess der Texterfassung manuell vorgenommen wird, siehe <http://www.dhmuseum.uni-trier.de/node/49>, zuletzt abgerufen am 03.08.2021.

¹⁰ <https://www.english-corpora.org/coca/>, zuletzt abgerufen am 18.05.2022.

¹¹ <https://hypotheses.org/about-hypotheses>, zuletzt abgerufen am 18.05.2022.

¹² <https://www.nationallizenzen.de/angebote/nlproduct.2006-03-10.7433818373>, zuletzt abgerufen am 18.05.2022.

II. Rechtlicher Rahmen des Text und Data Mining

Grundsätzlich ist das Text und Data Mining als solches keine urheberrechtlich relevante Handlung.¹³ Semantische Informationen in Texten – mit Ausnahme von fiktionalen Geschichten – sind urheberrechtlich nicht geschützt.¹⁴ Die Extraktion von Informationen aus urheberrechtlich geschützten Quellen als solche ist daher nicht vom Zuweisungsgehalt des Urheberrechts umfasst.¹⁵ Auch wenn die Analyse mit technischen Mitteln erfolgt, ändert sich auch nichts daran, dass die Analyse und das Herausfiltern von Informationen aus einem Text nicht vom Zuweisungsgehalt des Urheberrechts erfasst ist.¹⁶ Selbst wenn die automatisierte Auswertung als solche keine urheberrechtlich relevante Handlung darstellt, so sind die Informationen regelmäßig in einer urheberrechtlich geschützten Hülle¹⁷ oder in einer geschützten Datenbank¹⁸ enthalten und es können urheberrechtlich relevante Vervielfältigungen (§ 16 UrhG) oder Entnahmen (§ 87a ff. UrhG) als Vorstufe, zur Durchführung oder im Nachgang von Textanalysen notwendig sein.¹⁹ Denn insbesondere das Vervielfältigungsrecht in Art. 2 InfoSoc-RL (2001/29/EG) bzw. § 16 UrhG ist weit auszulegen und erfasst auch das vorübergehende Hochladen in den Arbeitsspeicher^{20, 21}

Um in diesen Fällen Rechtssicherheit zu gewähren, führte bereits die Text und Data Mining-Schranke für wissenschaftliche Forschung (§ 60d UrhG alte Fassung (a.F.), zur Umsetzung von Art. 3 DSM-RL ((EU) 2019/790)²² siehe § 60d UrhG) für solche Handlungen im Umfeld des Mining eine Ausnahme ein.²³ Als

¹³ Erwägungsgrund 9 der Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG (DSM-RL).

¹⁴ *Raue*, ZUM 2019, 684, 685 mwN., siehe insbesondere BGH GRUR 1981, 352, 353 – Staatsexamensarbeit.

¹⁵ *Raue*, ZUM 2019, 684, 686.

¹⁶ *Raue*, GRUR 2017, 11, 13.

¹⁷ Zu den Kriterien der urheberrechtlichen Schutzfähigkeit siehe *Erler-Fridgen*, Teil 2, A. in diesem Band.

¹⁸ Zum Datenbankschutz siehe *Erler-Fridgen*, Teil 2, D. in diesem Band.

¹⁹ *Raue*, ZUM 2019, 684, 685, zu den nötigen Kopiervorgängen siehe oben I.

²⁰ BGH GRUR 2011, 418 Rn. 13 – Used Soft; *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 16 Rn. 13 mwN.

²¹ *Raue*, GRUR 2017, 11, 13.

²² Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG (DSM-RL).

²³ *Raue*, GRUR 2017, 11, 13.

Schrankenbestimmung erfüllte § 60d UrhG a.F. die Funktion, Handlungen des Text und Data Mining zustimmungsfrei möglich zu machen, sofern dies zum Zweck der wissenschaftlichen Forschung erfolgte²⁴ und die handelnde Person bereits rechtmäßigen Zugang zu den betroffenen Werken hatte.²⁵ Die bisher bestehende Schranke für das Text und Data Mining in der wissenschaftlichen Forschung in § 60d UrhG a.F. stützte sich auf Art. 5 Abs. 3 lit. a InfoSoc-RL (2001/29/EG)²⁶ und Art. 6 Abs. 2 lit. b Datenbank-RL²⁷ und war insofern eine Innovation des deutschen Gesetzgebers^{28,29} Mit ihr griff der deutsche Gesetzgeber der damals noch unsicheren Verabschiedung der DSM-RL ((EU) 2019/790) und der in deren Art. 3 DSM-RL-E³⁰ vorgeschlagenen Text und Data Mining-Schranke für die wissenschaftliche Forschung vor.³¹ Nunmehr enthält die jüngst umgesetzte³² DSM-RL ((EU) 2019/790) in Art. 4 eine allgemeine Text und Data Mining-Schranke sowie in Art. 3 eine weitere TDM-Schranke zum Zwecke der wissenschaftlichen Forschung. Das UrhG beinhaltet nach der Umsetzung in § 44b und § 60d entsprechende Schrankenregelungen. Nun gibt es also zusätzlich eine allgemeine Text und Data Mining-Schranke (§ 44b UrhG in Umsetzung von Art. 4 DSM-RL ((EU) 2019/790)), die unter gesonderten Bedingungen das Text und Data Mining für alle freistellt, sofern rechtmäßiger Zugang zu den Texten besteht.³³ Die Bedeutung dieser beiden TDM-Schranken für die Qualität von Textanalysen ist wesentlich: Zwar können vorübergehende Vervielfältigungen bei Textanalysen³⁴ weiterhin auf § 44a UrhG gestützt werden.³⁵ Jedoch ist etwa für die Zusammenstellung, Nor-

²⁴ Zum allgemeinen Text und Data Mining-Schranke siehe unten III.

²⁵ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 60d Rn. 1; siehe auch Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 60d Rn. 2.

²⁶ Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft (Info-Soc-RL).

²⁷ Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11. März 1996 über den rechtlichen Schutz von Datenbanken (Datenbank-RL).

²⁸ Spindler, CR 2019, 277 Rn. 2.

²⁹ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 60d Rn. 2.

³⁰ COM(2016) 593 final.

³¹ Raue, ZUM 2021, 793, 794.

³² Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, in der Fassung der Bekanntmachung vom 04. Juni 2021, BGBl. I S. 1204 ff.

³³ Heesen/Jüngels, RuZ 2021, 45, 46 f. unter einer CC BY-SA Lizenz; zur Voraussetzung des rechtmäßigen Zugangs siehe unten IV.

³⁴ Etwa solche, die lediglich vorübergehend in den Arbeitsspeicher eines Computers hochgeladen werden: Raue, ZUM 2021, 793, 795.

³⁵ Raue, ZUM 2021, 793, 795.

malisierung und Annotation eines Datenkorpus bei fundierten Textanalysen regelmäßig eine dauerhafte Speicherung notwendig, die einer über § 44a UrhG hinausgehenden urheberrechtlichen Gestattung bedarf.³⁶

III. Text und Data Mining § 44 b UrhG

Die neu eingeführte allgemeine Text und Data Mining-Schranke in § 44b UrhG stellt das Text und Data Mining für alle frei und damit auch für kommerzielle Zwecke.³⁷ Diese allgemeine TDM-Schranke trägt dem Umstand Rechnung, dass Text und Data Mining-Verfahren nicht nur bedeutende Instrumente für die wissenschaftliche Forschung sind, sondern auch bei privaten sowie staatlichen Institutionen Einsatz finden.³⁸ Sie erfasst Vervielfältigungen von Werken und Datenbankwerken und ist außerdem auf Leistungsschutzrechte wie etwa das Leistungsschutzrecht des Datenbankherstellers³⁹ (siehe § 87c Abs. 1 Nr. 4 UrhG) sowie durch den Verweis auf Schranken für Werke beispielsweise auch auf das Leistungsschutzrecht für wissenschaftliche Ausgaben (§ 70 UrhG)⁴⁰ anwendbar.⁴¹

Zu diesem Zweck definiert sie Text und Data Mining in Absatz 1 als „die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen“. Diese Definition entspricht im Wesentlichen der Definition des Text und Data Minings in Art. 2 DSM-RL ((EU) 2019/790).⁴² Sie stellt jedoch klar, dass auch die Analyse von einzelnen Werken unter das Text und Data Mining fallen kann.⁴³ Generell ist es hiernach möglich, ein Korpus

³⁶ *Raue*, ZUM 2021, 793, 795; siehe insbesondere *Raue*, IIC 2018, 379, 381: vorbereitende Handlungen wie die Digitalisierung analoger Quellen oder die Normalisierung und Annotierung des Korpus verursachen üblicherweise eine längere Speicherdauer der Vervielfältigungen als sie von Art. 5 Abs. 1 InfoSoc-RL/§ 44a UrhG freigestellt ist.

³⁷ *Heesen/Jüngels*, RuZ 2021, 45, 46 f. unter einer CC BY-SA Lizenz.

³⁸ *Raue*, ZUM 2021, 793, 795 unter Verweis auf Erwägungsgrund 18 S. 1 DSM-RL.

³⁹ Hierzu ausführlich *Erler-Fridgen*, Teil 2, D. in diesem Band.

⁴⁰ Zum Leistungsschutzrecht für wissenschaftliche Ausgaben siehe: *Erler-Fridgen*, Teil 2, C. in diesem Band.

⁴¹ *Raue*, ZUM 2021, 793, 796; für § 70 UrhG siehe *Loewenbeim*, in *Schricker/Loewenbeim*, Urheberrecht, 6. Auflage 2020, § 70 Rn. 9.

⁴² Art. 2 Nr. 2 DSM-RL: Text und Data Mining bezeichnet eine Technik, für die automatisierte Analyse von Texten und Daten in digitaler Form, mit deren Hilfe Informationen unter anderem – aber nicht ausschließlich – über Muster, Trends und Korrelationen gewonnen werden können.

⁴³ *Raue*, ZUM 2020, 172, 172.

zum Zwecke des Text und Data Mining zu erstellen, wie es § 60d UrhG a.F. ausdrücklich regelte.⁴⁴ Digitale Werke in der Definition sind dabei nach der Begründung des Gesetzesentwurfs solche, die von vornherein in digitaler Form vorliegen, wohingegen digitalisierte Werke erst im Zuge des Text und Data Mining digitalisiert werden.⁴⁵ Das heißt, auch die Digitalisierung von analogen Werken wird als Vorstufe der Text und Data Analyse in § 44d Abs. 2 S. 1 UrhG erlaubt.⁴⁶ Nicht erfasst ist hingegen die Digitalisierung von Texten, die allein darauf abzielt, ein digitales Archiv zu erstellen.⁴⁷

Zulässig sind nach § 44b Abs. 2 UrhG Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining. Dieses Kriterium des rechtmäßigen Zugangs rechtfertigt die Schranke und ermöglicht es den Rechteinhabern, den Zugang zu ihren Werken zu kontrollieren und vergütungspflichtig zu stellen.⁴⁸ Rechtmäßiger Zugang liegt etwa bei entsprechenden vertraglichen Vereinbarungen zwischen Rechteinhabern und Nutzenden vor, besteht aber auch bei frei im Internet verfügbaren Gegenständen oder solchen, die im Wege des Open Access frei zugänglich sind.⁴⁹

Umfasst werden Vervielfältigungen für das Text und Data Mining.⁵⁰ Dabei ist zu beachten, dass das im Diskussionsentwurf des BMJV enthaltene Erforderlichkeitskriterium (es lautete in § 44b Abs. 2: „... sofern sie für das Text und Data Mining erforderlich sind ...“) entfallen ist⁵¹ und damit keine enge Auslegung der Verbindung zwischen Vervielfältigung und Text und Data Mining vorgenommen werden muss.⁵² Dies ermöglicht es nun, den Zweck der Vervielfältigung für das Text und Data Mining wie es auch in Art. 4 Abs. 1 DSM-RL ((EU)

⁴⁴ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 88.

⁴⁵ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 88.

⁴⁶ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 88.

⁴⁷ *Raue*, ZUM 2021, 793, 795 unter Hinweis auf Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 88.

⁴⁸ *Raue*, ZUM 2019, 684, 686.

⁴⁹ *Raue*, ZUM 2021, 793, 796 sowie *Schack*, GRUR 2021, 904, 907.

⁵⁰ Siehe auch *Raue*, ZUM 2019, 684, 686.

⁵¹ Siehe § 44b Abs. 2 Disk-E des Bundesministeriums der Justiz und für Verbraucherschutz vom 15.01.2020.

⁵² *Raue*, ZUM 2020, 172, 173.

2019/790) („zum Zweck des Text und Data Mining“) – zu Gunsten der angestrebten Rechtssicherheit für die Forschung durch die Normen – breiter als im ursprünglichen Entwurf auszulegen.⁵³

Was die Nachnutzbarkeit und Aufbewahrung der erzeugten Vervielfältigungsstücke angeht, so sieht § 44b UrhG vor, dass diese zu löschen sind, wenn sie für das Text und Data Mining nicht mehr erforderlich sind. Die umzusetzende DSM-RL ((EU) 2019/790) spricht in Art. 4 Abs. 2 in diesem Kontext davon, dass Vervielfältigungen und Entnahmen so lange aufbewahrt werden dürfen, wie sie für Zwecke des TDM notwendig sind. Eine solche Löschpflicht nach Abschluss der Forschungsarbeiten war in der Vorgängervorschrift zum Text und Data Mining in der Forschung in § 60d UrhG a.F. ebenfalls enthalten, ist jedoch im neuen § 60d UrhG nicht mehr vorgesehen. Damit ist in § 44b UrhG keine dauerhafte Möglichkeit zur Aufbewahrung geregelt,⁵⁴ sondern eine Löschpflicht nach Erfüllung ihres Zweckes, der Erforderlichkeit für das Text und Data Mining, vorgesehen.

Es soll den Rechteinhabern bei der kommerziellen Forschung in § 44b UrhG möglich sein, Nutzungen vorzubehalten nach § 44b Abs. 3 S. 1 UrhG (in Umsetzung von Art. 4 Abs. 3 DSM-RL ((EU) 2019/790)). Das bedeutet, dass Mining-Handlungen von Rechteinhabern aktiv verboten werden müssen.⁵⁵ Diese Regelung verschafft Rechteinhabern die Möglichkeit, weiterhin Lizenzen für das Text und Data Mining einzuräumen.⁵⁶ Nach § 44b Abs. 3 S. 2 UrhG ist ein solcher Nutzungsvorbehalt bei online zugänglichen Werken jedoch nur dann wirksam, wenn er in maschinenlesbarer Form erfolgt. Dies setzt auf der Voraussetzung in Art. 4 Abs. 3 DSM-RL ((EU) 2019/790) auf, dass die Rechteinhaber ausdrücklich und in angemessener Weise den Nutzungsvorbehalt vorsehen müssen.⁵⁷ Der Nutzungsvorbehalt nach § 44b Abs. 3 S. 1 UrhG wirkt jedoch lediglich für die Zukunft (*ex nunc*) und berührt Textanalysen nicht, die sich auf die Schranken nach § 44a UrhG (bei vorübergehenden Vervielfältigungen) oder auf § 60d UrhG stützen.⁵⁸

⁵³ Zum Disk-E des BMJV vom 15.01.2020 siehe noch *Raue*, ZUM 2020, 172, 173.

⁵⁴ *Heesen/Jüngels*, RuZ 2021, 45, 46 unter einer CC BY-SA Lizenz.

⁵⁵ *Heesen/Jüngels*, RuZ 2021, 45, 46 unter einer CC BY-SA Lizenz.

⁵⁶ *Raue*, ZUM 2021, 793, 796; Erwägungsgrund 18 S. 2 DSM-RL.

⁵⁷ Hierzu auch *Raue*, ZUM 2020, 172, 173.

⁵⁸ *Raue*, ZUM 2021, 793, 796.

Für das Text und Data Mining sieht § 44b UrhG keine Vergütung vor, da dies nach der Begründung des Gesetzesentwurfes der Rechtssicherheit diene.⁵⁹ Denn diese Rechtssicherheit könne nicht erreicht werden, wenn ansonsten in Zukunft immer ermittelt werden müsse, ob eine flüchtige Vervielfältigung (dann in jedem Fall vergütungsfrei nach § 44a UrhG) oder eine sonstige Vervielfältigung vorliege (anderenfalls zweifelhaft, ob vergütungsfrei).⁶⁰ Diese Abgrenzungsproblematik entfalle demnach durch die Vergütungsfreistellung.⁶¹

IV. Text und Data Mining zum Zwecke der wissenschaftlichen Forschung

§ 60 d UrhG setzt Art. 3 DSM-RL ((EU) 2019/790) um und modifiziert die Regelungen in § 44b UrhG (Art. 4 DSM-RL) für das allgemeine Text und Data Mining, wenn es für die wissenschaftliche Forschung eingesetzt wird.⁶² Das bedeutet, dass nach § 60d UrhG Forschende, Forschungsorganisationen und Einrichtungen des Kulturerbes urheberrechtlich geschützte Gegenstände automatisiert analysieren dürfen, wenn sie rechtmäßigen Zugang zu diesen haben.⁶³ Bereits seit 2018 existierte in § 60d UrhG a.F. eine auf Grundlage der InfoSoc-RL⁶⁴ eingeführte deutsche Text und Data Mining-Schranke für nicht kommerzielle – wissenschaftliche – Zwecke.⁶⁵ Weitergehende Regelungen⁶⁶ aus § 60d UrhG

⁵⁹ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 88. Art. 4 DSM-RL enthält dazu keine Aussage, es fehlt auch an einer ausdrücklichen Option für die Mitgliedsstaaten eine Vergütungspflicht vorzusehen; ausführlich dazu auch *Raue*, ZUM 2020, 172, 173.

⁶⁰ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 88.

⁶¹ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 88.

⁶² Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 95.

⁶³ *Raue*, ZUM 2021, 793, 797.

⁶⁴ Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft.

⁶⁵ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 95.

⁶⁶ Insbesondere wird in § 60d UrhG-E der Berechtigtenkreis über Forschungseinrichtungen oder Einrichtungen des Kulturerbes hinaus u.a. auf Einzelforscher erstreckt und die öffentliche Zugänglichkeit an Forschergruppen aufgeführt, *Raue*, ZUM 2020, 172, 173.

a.F. werden auch nach Umsetzung von Art. 3 DSM-RL ((EU) 2019/790) beibehalten, was unionsrechtlich ausdrücklich gestattet ist.⁶⁷

Für die Definition des Text und Data Mining verweist § 60d Abs. 1 UrhG auf § 44b UrhG und legt zudem fest, dass in Abgrenzung zu den allgemeinen Regelungen in § 44b UrhG für TDM zum Zwecke der wissenschaftlichen Forschung die Bestimmungen des § 60d Abs. 2 - 6 UrhG maßgeblich sind. § 44b Absatz 1 UrhG definiert Text und Data Mining wie beschrieben als „die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen.“ Zulässig sind Vervielfältigungen zum Zweck des Text und Data Minings, sodass neben der automatisierten Analyse, die Quellen auch zuvor gespeichert, normalisiert, annotiert oder auf sonstige Weise bearbeitet (§ 23 Abs. 3 UrhG) werden können.⁶⁸

Voraussetzung für die Zulässigkeit von Vervielfältigungen für TDM ist wie in § 44b Abs. 2 S. 1 UrhG der rechtmäßige Zugang zu den zugrundeliegenden Werken. Dabei gilt der Zugang zu frei im Internet verfügbaren Inhalten als rechtmäßig.⁶⁹ Außerdem kann hierunter auch nach Erwägungsgrund 14 der DSM-RL ((EU) 2019/790) der Zugang durch eine Open-Access Strategie, durch vertragliche Vereinbarungen zwischen Rechteinhabern und Forschungsinstitutionen wie Abonnements oder sonstige rechtmäßige Mittel fallen. Außerdem besteht rechtmäßiger Zugang zu Gegenständen, die nach Erwägungsgrund 14 S. 4 der DSM-RL ((EU) 2019/790) frei im Internet verfügbar sind sowie auch nach dem Erwerb oder der Fernleihe von analogen Medien.⁷⁰ Beispielsweise werden in Projekten digitale Zeitschriften oder E-Books, die über die Universitätsbibliotheken entsprechend lizenziert sind, eingesetzt. Auch sonstige, rechtmäßig zugängliche Datenbanken, können die Quelle für digitale Ressourcen sein.

⁶⁷ Art. 5 Abs. 3 lit. a InfoSoc-RL in der Neufassung durch Art. 24 Abs. 2 lit. b DSM-RL in Verbindung mit Art. 25 DSM-RL gestattet dies nach der Gesetzesentwurfsbegründung, siehe Gesetzentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 95.

⁶⁸ *Raue*, ZUM 2021, 793, 797 und 798.

⁶⁹ Erwägungsgrund 18 UAbs. 2 S. 1 DSM-RL, Gesetzentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 96; wobei die Art der Nutzung (Download oder anderes) bei im Internet frei verfügbaren Inhalten im Rahmen der Frage nach dem rechtmäßigen Zugang wohl zunächst zu keiner Änderung führen dürfte, die übrigen Voraussetzungen des TDM müssen berücksichtigt werden; insbesondere muss wissenschaftliche Forschung bezweckt werden.

⁷⁰ *Raue*, ZUM 2021, 793, 800; *Spindler* CR 2019, 277, 280.

Zur Vervielfältigung berechtigt sind nach Abs. 2 Forschungsorganisationen, die hier legaldefiniert werden als „Hochschulen, Forschungsinstitute oder sonstige Einrichtungen, die wissenschaftliche Forschung betreiben, sofern sie 1) nicht kommerzielle Zwecke verfolgen, 2) sämtliche Gewinne in die Forschung reinvestieren oder 3) im Rahmen eines staatlich anerkannten Auftrags im öffentlichen Interesse tätig sind“. Der Begründung des Gesetzesentwurfs nach entspricht diese Definition inhaltlich derjenigen in Art. 2 Nr. 1 DSM-RL ((EU) 2019/790) und ist lediglich in ihrer Formulierung gestrafft.⁷¹ Dem Wortlaut nach („oder“) genügt es, wenn eine von den drei genannten Voraussetzungen erfüllt ist.⁷²

Der Betrieb von „wissenschaftlicher Forschung“ in dieser Definition bezieht sich nach Erwägungsgrund 12 S. 2 DSM-RL sowohl auf die Naturwissenschaften als auch auf die Geisteswissenschaften.⁷³ Die in der DSM-RL ((EU) 2019/790) ausdrücklich genannte Lehrtätigkeit (auch in Verbindung mit wissenschaftlicher Forschung) wird in der Definition von Forschungsorganisationen nicht aufgeführt.⁷⁴ Jedoch betreiben mit Lehre befasste Institutionen ganz regelmäßig auch Forschung und etwa Fachhochschulen werden zutreffenderweise wohl auch unter die Definition gefasst.⁷⁵ Ist die Forschung jedoch gänzlich der Lehre untergeordnet, etwa bei Schulen, befindet sich dies außerhalb der Definition der Richtlinie.⁷⁶

Für die Qualifikation als Forschungsorganisation ist es jedoch unerheblich, welche Rechtsform oder Struktur die Forschungseinrichtung prägt.⁷⁷ Jedenfalls darf kein kommerzieller Zweck verfolgt werden (Nr. 1), der eine Gewinnorientierung der Einrichtung darstellen würde.⁷⁸ Zur Annahme einer Forschungsorganisation kommt man hingegen auch, wenn die Einrichtung ihre Gewinne vollständig in die Forschung reinvestiert (Nr. 2).⁷⁹ Auch werden Forschungsein-

⁷¹ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 96.

⁷² So auch: Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 96.

⁷³ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 96.

⁷⁴ Art. 2 Nr. 1 DSM-RL; *Raue*, ZUM 2020, 172, 173.

⁷⁵ *Raue*, ZUM 2020, 172, 173; *Spindler*, CR 2019, 277 Rn. 6.

⁷⁶ *Spindler*, CR 2019, 277 Rn. 6.

⁷⁷ *Raue*, ZUM 2021, 793, 800 unter Verweis auf Erwägungsgrund 12 S. 4 DSM-RL.

⁷⁸ *Raue*, ZUM 2021, 793, 800: siehe Art. 2 Nr. 1 lit. a DSM-RL.

⁷⁹ Zur dabei zulässigen Preisbildung für Forschungsergebnisse siehe *Raue*, ZUM 2021, 793, 800 mwN.

richtungen von der Schranke umfasst, die im Rahmen eines staatlich anerkannten Auftrags im öffentlichen Interesse tätig sind (Nr. 3). Typisch für einen staatlich anerkannten Auftrag im öffentlichen Interesse ist beispielsweise die Finanzierung durch die öffentliche Hand oder Bestimmungen zum öffentlichen Interesse in Rechtsvorschriften oder Aufträgen der öffentlichen Hand.⁸⁰

Nicht berechtigt sind hingegen nach § 60d Abs. 2 S. 2 UrhG Forschungsorganisationen, die mit einem privaten Unternehmen zusammenarbeiten, das einen bestimmenden Einfluss auf die Forschungsorganisation und einen bevorzugten Zugang zu den Ergebnissen der wissenschaftlichen Forschung hat. Bei sogenannten Public Private Partnerships und damit verbundenen Kooperation mit privaten Personen beim Text und Data Mining soll es demnach zwar möglich sein, sich auf die Text und Data Mining Schranke zu berufen, dies aber entsprechend eingeschränkt werden, wenn ein bestimmender Einfluss oder ein bevorzugter Zugang vorliegt.⁸¹ Beispielsweise kann ein bestimmender Einfluss eines privaten Unternehmens gegeben sein, wenn es aufgrund struktureller Gegebenheiten als Anteilseigner Kontrolle ausüben kann und hierdurch auch bevorzugten Zugang zu Forschungsergebnissen erhält.⁸²

Ferner berechtigt sind nach Absatz 3 Bibliotheken und Museen, wenn sie öffentlich zugänglich sind, sowie Archive und Einrichtungen im Bereich des Film- oder Tonerbes (Kulturerbe-Einrichtungen). Diese Legaldefinition setzt die neue Definition der Kulturerbe-Einrichtungen aus Art. 2 Nr. 3 der DSM-RL ((EU) 2019/790) um, der diese umreißt als „eine öffentlich zugängliche Bibliothek oder Museum, Archiv oder eine im Bereich des Film- oder Tonerbes tätige Einrichtung“.⁸³ In Abgrenzung zu den nach § 60d Abs. 2 S. 2 UrhG berechtigten Forschungsorganisationen, handelt es sich bei den Kulturerbe-Einrichtungen nicht um Forschungsinstitute oder andere Forschungsorganisationen, sondern beispielsweise um Universitätsbibliotheken und sonstige Bibliotheken, jedoch

⁸⁰ Gesetzentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 96; Erwägungsgrund 12 S. 6 DSM-RL.

⁸¹ Erwägungsgrund 11 DSM-RL, Gesetzentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 96; siehe auch *Spindler*, CR 2019, 277 Rn. 10.

⁸² Erwägungsgrund 12 S. 7 DSM-RL, Gesetzentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 96.

⁸³ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 60d Rn. 6; *de la Durantaye/Kuschel*, ZUM 2021, 785, 786.

unter der Voraussetzung des öffentlichen Zugangs zu diesen Bibliotheken.⁸⁴ Rein privat zugängliche Bibliotheken sind damit beispielsweise im Umkehrschluss nicht erfasst.⁸⁵

Weitergehend als Art. 3 DSM-RL ((EU) 2019/790) berechtigt Absatz 3 (wie schon die Vorgängervorschrift § 60d UrhG a.F. s.o.) auch Einzelforscher, sofern sie nicht kommerzielle Zwecke verfolgen. Die Berechtigung von Einzelforschern geht über die Umsetzung von Art. 3 DSM-RL hinaus und kann aber über Art. 25 DSM-RL und Art. 5 III lit. a InfoSoc-RL gefasst werden.⁸⁶

Die Berechtigten dürfen die Vervielfältigungen auch einem bestimmten Personenkreis für bestimmte Zwecke öffentlich zugänglich machen: Nach § 60d Abs. 4 Nr. 1 UrhG soll eine gemeinsame Forschung in Forschendengruppen ermöglicht werden und zu diesem Zweck die öffentliche Zugänglichmachung von Vervielfältigungen an einen bestimmt abgegrenzten Personenkreis gestattet sein.⁸⁷ Zur Überprüfung der Qualität von wissenschaftlicher Forschung erlaubt § 60d Abs. 4 Nr. 2 UrhG es, einzelnen Dritten die Vervielfältigungen öffentlich zugänglich zu machen. Nach Abschluss der Forschungsarbeiten bzw. der Überprüfung der Qualität wissenschaftlicher Forschung, ist die öffentliche Zugänglichmachung nach § 60d Abs. 4 UrhG zu beenden.

Die öffentliche Zugänglichmachung zur Qualitätsüberprüfung war bereits in der Vorgängervorschrift § 60d UrhG a.F. enthalten und deren Fortgeltung beruht wie andere weitergehende Regelungen (s.o.) auf Art. 25 und Erwägungsgrund 5 S. 1 DSM-RL ((EU) 2019/790).⁸⁸ Damit war bereits in der Vorgängerregelung während des Peer-Review-Verfahren vor einer Veröffentlichung die

⁸⁴ Näher hierzu *Raue*, ZUM 2021, 793, 801.

⁸⁵ Anderes gelte nach dem deutschen Gesetzeswortlaut für Archive und Einrichtungen des Film- oder Tonerbes: *Raue*, ZUM 2021, 793, 801.

⁸⁶ *Raue*, ZUM 2020, 172, 174; so auch *Heesen/Jüngels*, RuZ 2021, 45, 48 unter einer CC BY-SA Lizenz.

⁸⁷ Die Regelung des § 60d Abs. 4 UrhG ist an dieser Stelle nicht ganz klar, denn „öffentliche Zugänglichmachung“ im Sinne des § 19a UrhG/Art. 3 Abs. 2 Info-Soc-RL (2001/29/EG) wird nach der Rechtsprechung des EuGH an einen unbegrenzten Personenkreis gerichtet, *Spindler*, CR 2019, 277, 280 mit Hinweis u.a. auf EuGH C-527/15, ECLI:EU:C:2017:300, GRUR 2017, 610 Rn. 32 – Stichting Brein; Filmspeler. In § 60d Abs. 4 UrhG wird hingegen explizit auf einen begrenzten Personenkreis abgestellt; hierzu: *Raue*, CR 2017, 656, 660 sowie *Raue*, ZUM 2021, 793, 799; gegebenenfalls könne auch ein Zugänglichmachen vorliegen, auch wenn es sich um ein Zugänglichmachen im Sinne von § 19a UrhG handelt: *Spindler*, CR 2019, 277, 280.

⁸⁸ Gesetzentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 97.

öffentliche Zugänglichmachung zur Qualitätssicherung an einzelne Dritte möglich.⁸⁹ Eine – ggf. über diesen rechtlichen Rahmen hinausgehende – Möglichkeit zur Qualitätsüberprüfung ist (nicht nur) aus Sicht der Digital Humanities ein wesentliches Desiderat: Beispielsweise wäre die Bereitstellung von Textteilen zur Qualitätsüberprüfung von Verfahren wie etwa der Named Entity Recognition aus Sicht der Digital Humanities ein hilfreicher Beleg für die Verlässlichkeit der Ergebnisse der Analyseverfahren.⁹⁰

Neu ist, dass Berechtigte – nicht hingegen Einzelforscher ohne Anbindung an eine Forschungsinstitution – nach § 60d Abs. 5 UrhG die Vervielfältigungen mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufbewahren dürfen, solange sie für die Zwecke der wissenschaftlichen Forschung oder zur Überprüfung der wissenschaftlichen Qualität erforderlich sind. In der Vorgängervorschrift § 60d Abs. 3 UrhG a.F. war zuvor eine Löschpflicht nach Abschluss der Forschungsarbeiten vorgesehen (s.o.).⁹¹ Nach dieser Regelung ist es nun, anders als zuvor, möglich, dass die Forschungsorganisationen die Korpora selbst aufbewahren können, solange dies erforderlich ist. Das Merkmal „erforderlich“ ist an dieser Stelle enger als die Anforderungen der DSM-Richtlinie, nach der eine Bindung der Speicherung an den Zweck der wissenschaftlichen Forschung sowie auch an die Überprüfung der wissenschaftlichen Erkenntnisse ausreicht.⁹² Demnach ist auch hier die Dauer der Aufbewahrung der Forschungsergebnisse eingeschränkt – nur bis zu dieser Erforderlichkeitsgrenze – geregelt. Jedenfalls zehn Jahre lang kann die Aufbewahrung zur Überprüfung wissenschaftlicher Erkenntnisse nach den Grundsätzen guter wissenschaftlicher Praxis für erforderlich gehalten werden.⁹³ Die Erforderlichkeit einer längeren Aufbewahrungsdauer für die wissenschaftliche Forschung – gestützt auf die Wissenschaftsfreiheit – müsste seitens der Forschenden etwa unter Verweis auf

⁸⁹ Zum alten Recht: *Raue*, CR 2017, 656, 695, Gesetzesentwurfsbegründung BT-Drucks. 18/12329, 41.

⁹⁰ Hierzu ausführlich *Erler-Fridgen*, Teil 6, A. in diesem Band.

⁹¹ *Raue*, ZUM 2020, 172, 174.

⁹² *Raue*, ZUM 2020, 172, 174.

⁹³ *Raue*, ZUM 2021, 793, 799 nach Erörterung etwa von: DFG, Leitlinien zur Sicherung guter wissenschaftlicher Praxis, Kodex, 2019, Leitlinie 12, 13, insbesondere 17; Argumentation für eine längere Aufbewahrungsdauer: *Heesen/Jüngels*, RuZ 2021, 45, 50 unter einer CC BY-SA Lizenz.

Anschlussforschung ausreichend plausibel gemacht werden.⁹⁴ Nach der Begründung des Gesetzesentwurfs kann auch eine dauerhafte Speicherung erforderlich und damit zulässig sein, insbesondere, wenn sie durch die Kulturerbe-Einrichtungen und nicht die Forschungseinrichtung selbst erfolgt.⁹⁵ Demnach und ausweislich seines Wortlauts scheint § 60 d Abs. 5 UrhG die Option der Vorgängervorschrift § 60d Abs. 3 S. 2 UrhG a.F. zu eröffnen, die Archivierung auch bei einer Kulturerbe-Einrichtung zentral vorzunehmen.⁹⁶ Voraussetzung der Aufbewahrung sind angemessene Sicherheitsvorkehrungen gegen unbefugte Benutzung nach § 60d Abs. 5 UrhG. Es wird aus diesem Grund in der Literatur empfohlen, diese Daten bei Forschungsdatenrepositorien aufzubewahren, die institutionell dafür Sorge tragen können, dass diese Sicherheitsstandards eingehalten werden.⁹⁷

Die Nutzung zum Text und Data Mining für die wissenschaftliche Forschung wird durch die Änderung des § 60h Abs. 2 UrhG vergütungsfrei gestellt. Denn in Erwägungsgrund 17 S. 2 DSM-RL ((EU) 2019/790) ist festgehalten, dass das Text und Data Mining zu Zwecken der wissenschaftlichen Forschung künftig entsprechend ausgestaltet werden soll.⁹⁸ Dies entspricht auch der zugrundeliegenden Wertung der Voraussetzung des rechtmäßigen Zugangs, der eine ausreichende Möglichkeit der Rechteinhaber zur vergütungspflichtigen Ausgestaltung des Zugangs bietet.⁹⁹ Schließlich können vertragliche Einschränkungen nach § 60g Abs. 1 UrhG und technische Schutzmaßnahmen nach § 95b Abs. 1

⁹⁴ *Raue*, ZUM 2021, 793, 799: Forschende haben hier eine Einschätzungsprärogative, die gerichtlich lediglich eingeschränkt auf Missbrauch überprüft werden kann; bereits zur Definition des Forschungsprojektes durch die Forschenden unter § 60d UrhG a.F. siehe *Stieper*, in *Schricker/Loewenheim*, Urheberrecht, 6. Auflage 2020, § 60d Rn. 17 und *Raue*, CR 2017, 656, 659.

⁹⁵ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 97.

⁹⁶ Zu den Vorteilen der zentralen Archivierung siehe *Kleinkopf/Jacke/Gärtner*, MMR 2021, 196, 197, 198.

⁹⁷ *Raue*, ZUM 2021, 793, 799; Erwägungsgrund 15 S. 3 DSM-RL führt diesbezüglich sogenannte vertrauenswürdige Stellen ein, die der deutsche Gesetzgeber entsprechend Erwägungsgrund 15 S. 3 DSM-RL benennen kann: *Heesen/Jüngels*, RuZ 2021, 45, 52 unter einer CC BY-SA Lizenz.

⁹⁸ Gesetzesentwurf der Bundesregierung, Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, BT-Drs. 19/27426, 100.

⁹⁹ *Raue*, ZUM 2021, 793, 798.

Nr. 11 und Abs. 3 Nr. 5 UrhG die Schranke des § 60d UrhG nicht beeinträchtigen.¹⁰⁰ Der Vorrang der Schranke § 60d UrhG¹⁰¹ nach § 60g gilt nach § 137o UrhG jedoch nicht für Verträge, die vor dem 01. März 2018 geschlossen wurden. Neben der Wissenschaftsschranke § 60d UrhG können sich Forschende auch auf die Schranke für vorübergehende Vervielfältigungshandlungen nach § 44a UrhG sowie auf die allgemeine Text und Data Mining-Schranke nach § 44b UrhG berufen.¹⁰²

V. Ergebnis

Eine neu eingeführte allgemeine Text und Data Mining-Schranke in § 44b UrhG stellt das Text und Data Mining für alle und auch für kommerzielle Zwecke frei. Zulässig sind nach § 44b Abs. 2 S. 1 UrhG Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining. Erstellte Vervielfältigungen sind jedoch nach § 44b Abs. 2 UrhG zu löschen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind. Das Text und Data Mining ist vergütungsfrei, es können jedoch Nutzungen durch die Rechteinhaber nach § 44b Abs. 3 UrhG vorbehalten werden.

§ 60 d UrhG setzt Art. 3 DSM-RL ((EU) 2019/790) um und stellt das Text und Data Mining frei, wenn es für die wissenschaftliche Forschung eingesetzt wird. Voraussetzung für die Zulässigkeit von Vervielfältigungen für TDM ist wie in § 44b Abs. 2 S. 1 UrhG der rechtmäßige Zugang zu den zugrundeliegenden Werken. Berechtigt zur Durchführung von TDM sind hiernach Forschungsorganisationen und ihre Mitglieder (Abs. 2) sowie Bibliotheken, Museen und Einzelforscher (Abs. 3). Öffentlich zugänglich¹⁰³ gemacht werden dürfen die Vervielfältigungen einem abgegrenzten Personenkreis für die gemeinsame wissenschaftliche Forschung sowie einzelnen Dritten zur Qualitätsüberprüfung. Eine Präsentation der Vervielfältigungen über diese Grenzen hinaus wird durch die TDM-Schranke nicht ermöglicht. Berechtigte – nicht hingegen Einzelforscher – dürfen nach § 60d Abs. 5 UrhG die Vervielfältigungen mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufbewahren, solange sie

¹⁰⁰ Ausführlich hierzu *Raue*, ZUM 2021, 793, 801: Der Rechteinhaber muss nach § 95b Abs. 1 Nr. 11 UrhG den Begünstigten die notwendigen Mittel zur Verfügung zu stellen, damit sie von der TDM-Schranke trotz technischer Schutzmaßnahmen in dem erforderlichen Maße Gebrauch machen können.

¹⁰¹ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 60g Rn. 3.

¹⁰² *Raue*, ZUM 2021, 793, 798.

¹⁰³ Zum Begriff des „Öffentlich Zugänglichmachens“ in § 60d Abs. 4 UrhG siehe oben.

für die Zwecke der wissenschaftlichen Forschung oder zur Überprüfung der wissenschaftlichen Qualität erforderlich sind. Auch das Text und Data Mining für die wissenschaftliche Forschung wird nach § 60h Absatz 2 Nr. 2 UrhG vergütungsfrei gestellt.

Wer Vervielfältigungsstücke über die beschriebenen Freistellungen der TDM-Schranken hinaus nutzen oder präsentieren möchte, muss die Zustimmung der Rechteinhaber einholen oder sich auf eine der übrigen Schranken der §§ 44a ff. UrhG berufen können. Abgeleitete Textformate können einen Ansatz für Lösungen des urheberrechtskonformen Zugangs zu großen Textkorpora außerhalb des Rahmens der Text und Data-Mining Schranken bieten.¹⁰⁴

Leseempfehlungen zur vertiefenden Lektüre: *Raue*, Die Freistellung von Datenanalysen durch die neuen Text und Data Mining-Schranken (§§ 44b, 60d UrhG), *Zeitschrift für Urheber- und Medienrecht (ZUM)* 2021, 793, *Heesen/Jüngels*, Der Regierungsentwurf der Text und Data Mining-Schranken (§§44b und 60d UrhG-E), *Recht und Zugang (RuZ)* 2021, 45 unter einer CC BY-SA Lizenz, *Spindler*, Die neue Urheberrechts-Richtlinie der EU, insbesondere „Upload-Filter“ – Bittersweet?, *Computer und Recht (CR)* 2019, 277, *Raue/Schöch*, Zugang zu großen Textkorpora des 20. und 21. Jahrhunderts mit Hilfe abgeleiteter Textformate – Versöhnung von Urheberrecht und textbasierter Forschung, Teil 3, A. in diesem Band.

¹⁰⁴ *Raue/Schöch*, Teil 3, A. in diesem Band, 129, 131f.

B. Text und Data Mining in Einrichtungen des Kulturerbes

Die neuen Möglichkeiten des § 60d UrhG n.F. aus Sicht von Gedächtniseinrichtungen *

Benjamin Raue

Version (2022), CC BY-SA 4.0.

I. Ungehobene Wissensschätze in Einrichtungen des Kulturerbes

Als Zwerge stehen wir auf den Schultern von Riesen.¹ Dieses Gleichnis veranschaulicht, dass wir auf den Erkenntnissen, dem Wissen vergangener Generationen aufbauen und dieses fortschreiben. In der Vergangenheit war es vor allem die schöpferische Kraft des menschlichen Geistes, die aus vorhandenem Wissen neues hervorgebracht hat. Weil immer mehr Wissensquellen digital vorliegen oder digitalisiert werden, können wir nunmehr für die Generierung neuen Wissens auf die Unterstützung von Algorithmen zurückgreifen und mithilfe von Algorithmen „Muster, Trends und Korrelationen“² in großen Textmengen und anderen Datensätzen erkennen. Im Urheberrecht wird dies als Text und Data Mining bezeichnet (§ 44b Abs. 1 UrhG).

Große, in weiten Teilen ungehobene Wissensschätze schlummern in Bibliotheken, Museen und Archiven. Es gehört zum Selbstverständnis dieser Gedächtnisinstitutionen, die von ihnen gesammelten Objekte nicht nur zu verwahren, sondern zur Gewinnung und Erweiterung von neuem Wissen einzusetzen.³ Deswegen ist es konsequent, dass die neue Text und Data Mining-Schranke in § 60d

* Dieser Betrag wurde zuerst veröffentlicht als *Raue RuZ 2022*, 4-18 unter einer CC BY-SA Lizenz.

¹ Bekannt ist die Verwendung des Gleichnisses durch *Isaac Newton* in einem Brief an *Robert Hooke*. Sie ist aber auch schon früher verwendet worden, etwa von *Johannes von Salisbury*, dazu etwa *Leuker*, *Mittelalterliches Jahrbuch 1997*, 71 m.w.N.

² So dessen Definition in § 44b Abs. 1 UrhG und Art. 2 Nr. 2 DSM-RL.

³ Vgl. für Museen etwa *ICOM*, *Ethische Richtlinien für Museen* von ICOM, 2. Aufl. 2010, Richtlinie 3, abrufbar unter https://icom-deutschland.de/images/Publikationen_Buch/Publikation_5_Ethische_Richtlinien_dt_2010_komplett.pdf, zuletzt abgerufen

UrhG ausdrücklich Bibliotheken und Museen sowie Archive und Einrichtungen im Bereich des Film- oder Tonerbes privilegiert (und sie in Abs. 3 Nr. 1 als Kulturerbe-Einrichtungen legaldefiniert). Mit der prominenten gesetzgeberischen Beachtung und den neu geschaffenen Möglichkeiten geht nun aber auch eine gewisse Verantwortung der Gedächtniseinrichtungen einher, diese mit Leben zu füllen.⁴

II. Möglichkeit und Aufgabe für Einrichtungen des Kulturerbes

Viele Informationen sind in Texten, Bildern, Videos oder anderen Datensammlungen enthalten, die urheberrechtlich geschützt sind („urheberrechtliche Hülle“). Zwar sind die in urheberrechtlich geschützten Gegenständen enthaltenen (Sach-) Informationen nicht vom Zuweisungsgehalt des Urheberrechts gedeckt.⁵ Wenn und soweit für die Informationsextraktion allerdings Vervielfältigungen erforderlich sind, fallen sie wegen § 16 UrhG in dessen Anwendungsbereich, weil das UrhG auch vorübergehende Vervielfältigungen grundsätzlich dem Urheber vorbehält.⁶ Auch wenn die Inhalte in einer nach § 87a ff. UrhG geschützten Datenbank enthalten und für weitere Analyse vervielfältigt werden müssen, kann das Urheberrecht der Forschung rechtliche Grenzen setzen. Jedenfalls für die nicht-kommerzielle, wissenschaftliche Forschung hat § 60d UrhG a.F. aus dem Jahr 2017 die Informationen bildlich gesprochen vom Urheberrecht „befreit“ und Vervielfältigungshandlungen zum Zweck des Text und Data Mining freigestellt, soweit Forschende rechtmäßigen Zugang zu den Werken hatten.⁷

am 05.08.2022 und für Archive § 3 Abs. 1 S. 1 Bundesarchivgesetz: „Das Bundesarchiv hat die Aufgabe, das Archivgut des Bundes auf Dauer zu sichern, nutzbar zu machen und wissenschaftlich zu verwerten.“

⁴ *Döhl*, RuZ 2020, 195, 209 unter einer CC BY-SA Lizenz. Zu den „Herausforderungen und Möglichkeiten für Bibliotheken“ beim Text und Data Mining, *Drees*, Perspektive Bibliothek 5.1 (2016), 49 ff., insb. 62 ff.

⁵ Vgl. nur BGH, GRUR 1981, 352, 353, 355 – Staatsexamensarbeit; GRUR 2011, 803 Rn. 49 f. – Lernspiele; *GA Szpunar*, C-469/17, ECLI:EU:C:2018:870 Tz. 16, 19 – Funke Medien/NRW; *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 196; *Obergfell*, FS Büscher, 223, 225; *Dreier/Schulze*, UrhG, 6. Aufl. 2018, § 60d Rn. 1; *Raue*, GRUR 2017, 10, 14.

⁶ Dazu *Spindler*, GRUR 2016, 1112, 1113 ff.; *ders.*, ZGE 2018, 273, 276; *Raue*, CR 2017, 656.

⁷ Zu § 60d UrhG a.F. etwa *Dreier/Schulze*, UrhG, 6. Aufl. 2018, § 60d; *Obergfell*, FS Büscher, 223, 228 ff.; *Raue*, CR 2017, 656 ff.; *Spindler*, ZGE 2018, 273, 277 ff.; *Specht*, OdW 2018, 285; *Schricker/Loewenheim-Stieper*, Urheberrecht, 6. Aufl. 2020, § 60d.

Die Richtlinie über das Urheberrecht im digitalen Binnenmarkt (DSM-RL)⁸ hat die Schrankenregelungen für das Text und Data Mining in der EU abschließend harmonisiert und verpflichtet alle Mitgliedstaaten, entsprechende Schranken in ihre Urhebergesetze einzuführen. Anders als die bisherige deutsche Urheberrechtsschranke in § 60d UrhG a.F. stellt Art. 3 DSM-RL nicht auf den einzelnen Forscher ab, sondern verfolgt einen institutionenorientierten Ansatz. Die Kommission wollte in ihrem ersten Richtlinien-Entwurf⁹ von Art. 3 DSM-RL-E das Text und Data Mining nur für Forschungsorganisationen freistellen.¹⁰ Mit der Richtlinie will die Europäische Union jedoch ausdrücklich die kulturelle Vielfalt bewahren, fördern und gleichzeitig das gemeinsame kulturelle Erbe Europas hervorheben (Erwgr. 2 DSM-RL). Das schließt auch die Erforschung des kulturellen Erbes ein. Daher ist es nur konsequent, dass in dem verabschiedeten Richtlinienentwurf mit Museen, Archive und Bibliotheken auch solche Institutionen in den Anwendungsbereich der Text und Data Mining-Schranke eingeschlossen werden, die über große Wissensschätze verfügen.

Denn auch die neue Text und Data Mining-Schranke gilt nur für Schutzgegenstände, zu denen der Nutzer rechtmäßigen Zugang hat (§§ 60d Abs. 1 iVm. § 44b Abs. 2 UrhG).¹¹ Die Schranke gewährt ihm aber kein Zugangsrecht zu Informationen. Daher wäre es bei dem institutionenorientierten Ansatz wenig stimmig gewesen, gerade solche Institutionen vom Anwendungsbereich der Schranke auszuschließen, die über große und wertvolle Wissensressourcen verfügen. § 60d UrhG n.F. und Art. 3 DSM-RL privilegieren daher nun ausdrücklich auch „Einrichtungen des Kulturerbes“ bei urheberrechtlichen Nutzungshandlungen zum Zwecke des Text und Data Mining, solange dies der wissenschaftlichen Forschung dient.

Im Folgenden wird unter III. erläutert, welche Institutionen § 60d Abs. 3 Nr. 1 UrhG als „Einrichtung des Kulturerbes“ ansieht und welcher Personenkreis die von der Schranke freigestellten Handlungen vornehmen darf. Unter IV. werden die von § 60d UrhG freigestellten Handlungen und anschließend unter V. das Konzept einer „vertrauenswürdige Stelle“ vorgestellt, die von den Mitgliedstaaten für die sichere Aufbewahrung von Daten Korpora benannt werden können.

⁸ Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG, ABl. L 130, 92.

⁹ *Kommission*, Vorschlag für eine Richtlinie des Europäischen Parlaments und des Rates über das Urheberrecht im digitalen Binnenmarkt, COM/2016/0593 final.

¹⁰ Zu Art. 3 DSM-RL-E *Raue*, GRUR 2017, 11, 12 ff.

¹¹ Das war auch für die frühere Fassung von § 60d UrhG anerkannt, RegE, BT-Drucks. 18/12329, 41; *Raue*, CR 2017, 656, 658; *Spindler*, ZGE 2018, 273, 281.

III. Kreis der begünstigten Gedächtniseinrichtungen

Zu den begünstigten Gedächtniseinrichtungen gehören nach § 60d Abs. 3 Nr. 1 UrhG Bibliotheken, Museen, Archive und Einrichtungen im Bereich des Film- und Tonerbes (dazu 1.). Bibliotheken von Forschungseinrichtungen nehmen zudem an der Privilegierung ihrer Forschungseinrichtung teil (dazu 2.).

1. Einrichtungen des Kulturerbes (§ 60d Abs. 3 Nr. 1 UrhG)

a) Bibliotheken, Museen, Archive und Einrichtungen des Film- oder Tonerbes

§ 60d Abs. 3 Nr. 1 UrhG zählt die als Kulturerbe-Einrichtungen begünstigten Gedächtnisinstitutionen auf: Bibliotheken, Museen, Archive und Einrichtungen im Bereich des Film- und Tonerbes. Die Definition geht auf Art. 2 Nr. 3 DSM-RL zurück und muss daher richtlinienkonform ausgelegt werden. Erwgr. 13 DSM-RL erläutert die Definition, enthält aber nur punktuelle Präzisierungen. Die einzelnen Institutionen definiert die DSM-RL jedoch nicht näher. Teilweise kann hierfür auf andere EU-Richtlinien zurückgegriffen werden, die die genannten Einrichtungen in unterschiedlichen Kontexten privilegieren.¹² Weil bei diesen allerdings kein kohärentes Konzept erkennbar ist, müssen die Begriffe in erster Linie im Kontext der DSM-RL und in zweiter Linie in Bezug auf die jeweilige Schrankenregelung ausgelegt werden. Weil die Vorgaben der DSM-RL vollharmonisierend und Tatbestandsmerkmale des Unionsrechts autonom auszulegen sind,¹³ können die Definitionen des UrhG sowie Konkretisierungen der deutschen Rechtsprechung¹⁴ nur mit Vorsicht herangezogen werden.

aa) Museen

Museen können ganz allgemein als auf Dauer angelegte Einrichtungen definiert werden, die naturwissenschaftlich, geschichtlich, gesellschaftlich oder kul-

¹² Etwa Art. 1 RL 2012/28/EU; Art. 5 Abs. 2 lit. c, Abs. 3 lit. n RL 2001/29/EG.

¹³ Vgl. nur EuGH, Urt. v. 29.07.2019 – C 516/17, ECLI:EU:C:2019:625, Tz. 62 – Spiegel Online/Volker Beck m.w.N.

¹⁴ Etwa die Legaldefinition von Bibliotheken in § 60e Abs. 1 UrhG sowie die von der Literatur und teilweise von der Rechtsprechung definierten Begriffe aus §§ 53 Abs. 2 S. 1 Nr. 2, 60f Abs. 2, 61 Abs. 2 UrhG.

turell bedeutende Objekte sammeln, bewahren, erschließen und zugänglich machen.¹⁵ Offen ist, ob auch solche Kunsthallen von den urheberrechtlichen Privilegierungen erfasst werden, die keine eigene Sammlung haben, sondern sich darauf beschränken, wechselnde Ausstellungen mit geliehenen Werken zu organisieren. Erwgr. 13 scheint ein gesetzliches Leitbild eines Museums mit eigener „Sammlung“ (engl.: „*their permanent collection*“; frz. „*leurs collections permanentes*“) zugrunde zu legen. Allerdings stellt die Text und Data Mining-Schranke im Übrigen nicht darauf ab, dass Werke dauerhaft im Besitz der berechtigten Institution stehen. Ausreichend ist, dass sie rechtmäßigen Zugang zu den Inhalten hat. Ein solch rechtmäßiger Zugang kann nach Erwgr. 14 auch darin bestehen, dass ein (im Grundsatz zeitlich beschränkter) Zugang zu Inhalten aufgrund eines Datenbank-Abonnements besteht.

bb) Bibliotheken

Bibliotheken sind Einrichtungen, die planmäßig veröffentlichte Informationen, traditionell von Büchern, sammeln, ordnen und zugänglich machen.¹⁶ Der BGH versteht darunter eine Organisationseinheit, in der „ein systematisch gesammelter und Benutzern zentral zur Verfügung gestellter Bibliotheksbestand vorhanden [ist], der nach seiner Größe und dem Umfang seiner Benutzung einer besonderen Verwaltung (unter anderem auch in Form einer Katalogisierung) bedarf“.¹⁷ Nach der Legaldefinition in § 60e Abs. 1 UrhG müssen sie öffentlich zugänglich sein und dürfen keine unmittelbaren oder mittelbaren kommerziellen Zwecke verfolgen. Ob letzteres auch Voraussetzung für die Privilegierung in § 60d UrhG ist, ist unklar, weil die (vollharmonisierende) Definition der Bibliotheken in Art. 2 Nr. 3 DSM-RL keine solche Einschränkung enthält (dazu unten c).¹⁸

¹⁵ Vgl. Dreier/Schulze, 7. Aufl. 2022, § 60f Rn. 5 sowie die Museumsdefinition des Internationalen Museumsrats ICOM, Ethische Richtlinien für Museen von ICOM, 2. Aufl. 2010, S. 29, abrufbar unter https://icom-deutschland.de/images/Publikationen_Buch/Publikation_5_Ethische_Richtlinien_dt_2010_komplett.pdf, zuletzt abgerufen am 05.08.2022: „Ein Museum ist eine gemeinnützige, auf Dauer angelegte, der Öffentlichkeit zugängliche Einrichtung im Dienste der Gesellschaft und ihrer Entwicklung, die zum Zwecke des Studiums, der Bildung und des Erlebens materielle und immaterielle Zeugnisse von Menschen und ihrer Umwelt beschafft, bewahrt, erforscht, bekannt macht und ausstellt.“

¹⁶ Vgl. Ewert/Umstätter, Bibliotheksdienst 1999, 957, 966; Gantert, Bibliothekarisches Grundwissen, 9. Aufl. 2016, 6 f.; Dreier/Kotthoff/Meckel/Hentsch, Urheberrecht, 4. Aufl. 2018, § 60e Rn. 5.

¹⁷ BGH Urt. v. 20.02.1997, I ZR 13/95, NJW 1997, 3440 (3443) – Betreibervergütung (zu § 54 Abs. 2 Satz 2 UrhG a.F.).

¹⁸ Ablehnend auch Dreier/Schulze, UrhG, 7. Aufl. 2022, § 60d Rn. 6.

Auf die Art der gesammelten und zugänglich gemachten Werke kommt es nicht an.¹⁹ Daher zählen auch Bibliotheken mit Musik-, Video- oder anderen Sammlungen zu Einrichtungen des Kulturerbes.²⁰ Noch ungeklärt ist, ob auch rein virtuelle Bibliotheken mit einer ausschließlich digitalen Sammlung ohne räumliche Anlaufstelle als Bibliotheken i.S.d. UrhG anzusehen sind.²¹

cc) Archive

Ein Archiv ist klassischerweise eine Einrichtung, die bedeutende Dokumente sammelt, systematisiert aufbewahrt, erhält und zugänglich macht.²² Ob darüber hinaus auch Spartenarchive wie Literatur-, Film-, Ton-, Bild- oder sonstige Archive unter den Archivbegriff der Richtlinie fallen, ist offen. Soweit Archive veröffentlichte Werke sammeln und zugänglich machen, fallen sie in jedem Fall unter die Bibliotheksdefinition; Film- und Tonarchive sind als Einrichtungen des Film- oder Tonerbes erfasst (siehe sofort dd).

dd) Einrichtungen des Film- oder Tonerbes

Einrichtungen des Film- oder Tonerbes definiert Erwgr. 20 der Verwaiste-Werke-RL 2012/28/EU als Einrichtungen, „zur Sammlung, Katalogisierung, Erhaltung und Restaurierung von Filmen und anderen audiovisuellen Werken oder Tonträgern“. Diese allgemein gehaltene Definition kann auch für die Zwecke der DSM-RL und damit für § 60d Abs. 3 Nr. 1 UrhG herangezogen werden.

b) Öffentliche Zugänglichkeit

aa) Betroffene Institutionen

Einschränkend werden Bibliotheken und Museen nach § 60d Abs. 3 Nr. 1 UrhG nur dann als Einrichtungen des Kulturerbes privilegiert, wenn sie öffentlich zugänglich sind. Archive und Einrichtungen des Film- oder Tonerbes müssen im Umkehrschluss dagegen nicht öffentlich zugänglich sein.

¹⁹ Erwgr. 13 S. 1 DSM-RL.

²⁰ Schricker/Loewenheim-*Stieper*, Urheberrecht, 6. Aufl. 2020, § 60e Rn. 4. So schon zum Begriff der Bibliothek in Art. 5 II lit. c InfoSoc-RL Walther/von Lewinski-*Walther/von Lewinski*, European Copyright Law, 2010, Art. 5 Rn. 11.5.37.

²¹ Dafür Dreyer/Kotthoff/Meckel/*Hentsch*, Urheberrecht, 4. Aufl. 2018, § 60e Rn. 5. Dagegen Henke, E-Books im Urheberrecht – Kollision von Buchkultur und digitaler Wissensgesellschaft, 2018, 118 ff.; Schricker/Loewenheim-*Stieper*, Urheberrecht, 6. Aufl. 2020, § 60e Rn. 4.

²² Vgl. Duden online, # „Archiv“, abrufbar unter: <https://www.duden.de/rechtschreibung/Archiv>, zuletzt abgerufen am 05.08.2022.

Der deutsche Gesetzgeber weicht insofern von der deutschen Fassung der entsprechenden Definition in Art. 2 Nr. 3 DSM-RL ab, bei der sich die Voraussetzung des Zugangs für die Öffentlichkeit auf drei der vier genannten Institutionen zu erstrecken scheint. Dennoch ist die Präzisierung unionsrechtskonform. Denn wie schon in der InfoSoc-RL²³ und Verwaiste-Werke-RL²⁴ bezieht sich das einschränkende Merkmal „öffentlich zugänglich“ in den unterschiedlichen Sprachfassungen auf einen unterschiedlichen Kreis der aufgeführten Einrichtungen.²⁵

Zur besseren Nachvollziehbarkeit sind im Folgenden die deutsche, englische und französische Fassung von Art. 2 Nr. 3 DSM-RL nebeneinander abgedruckt, wobei die kursive Hervorhebung hinzugefügt wurde.

Deutsche Fassung	Englische Fassung	Französische Fassung
„Einrichtung des Kulturerbes“ bezeichnet eine <i>öffentlich zugängliche</i> Bibliothek oder Museum, Archiv oder eine im Bereich des Film- oder Tonerbes tätige Einrichtung	‘cultural heritage institution’ means a <i>publicly accessible</i> library or museum, an archive or a film or audio heritage institution	«institution du patrimoine culturel», une bibliothèque <i>accessible au public</i> , un musée, des archives ou une institution dépositaire d'un patrimoine cinématographique ou sonore

Auch wenn im Übrigen grammatikalisch misslungen („eine“) scheint sich das „öffentlich zugänglich“ auf Bibliothek, Museum und Archiv, wegen des erneuten unbestimmten Artikels ohne Nennung des Attributs nicht aber auf die Einrichtung des Film- oder Tonerbes zu beziehen. Aus demselben Grund bezieht sich das „*publicly accessible*“ in der englischen Sprachfassung wohl nur auf die Bibliotheken und Museen, in der französischen Fassung das „*accessible au public*“ eindeutig nur auf Bibliotheken.

²³ Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft, ABl. L 167/10.

²⁴ Richtlinie 2012/28/EU des Europäischen Parlaments und des Rates vom 25. Oktober 2012 über bestimmte zulässige Formen der Nutzung verwaister Werke, ABl. L 299/5.

²⁵ Zu den sprachlichen Ungenauigkeiten von InfoSoc- und Verwaiste-Werke-RL etwa Schrickler/Loewenheim-*Spindler*, Urheberrecht, 6. Aufl. 2020, § 61 Rn. 22.

Etwas eindeutiger, allerdings auch nicht widerspruchsfrei ist Erwgr. 13 DSM-RL, der zum besseren Verständnis ebenfalls in den drei Sprachfassungen abgedruckt ist:

Deutsche Fassung	Englische Fassung	Französische Fassung
<p>Als Einrichtungen des kulturellen Erbes sollten <i>öffentlich zugängliche</i> Bibliotheken und Museen unabhängig von der Art der dauerhaft in ihren Sammlungen befindlichen Werke oder sonstigen Schutzgegenstände sowie Archive und im Bereich des Film- oder Tonerbes tätige Einrichtungen gelten. Als solche sollten unter anderem auch Nationalbibliotheken und Nationalarchive gelten sowie die Archive und die <i>öffentlich zugänglichen</i> Bibliotheken von Bildungseinrichtungen, Forschungsorganisationen und öffentlich-rechtlichen Rundfunkanstalten.</p>	<p>Cultural heritage institutions should be understood as covering <i>publicly accessible</i> libraries and museums regardless of the type of works or other subject matter that they hold in their permanent collections, as well as archives, film or audio heritage institutions. They should also be understood to include, inter alia, national libraries and national archives, and, as far as their archives and <i>publicly accessible</i> libraries are concerned, educational establishments, research organisations and public sector broadcasting organisations.</p>	<p>Les institutions du patrimoine culturel devraient s'entendre comme couvrant les bibliothèques <i>accessibles au public</i> et les musées, quels que soient les types d'œuvres ou autres objets protégés qu'ils détiennent dans leurs collections permanentes, de même que les archives et les institutions dépositaires du patrimoine cinématographique ou sonore. Elles devraient aussi s'entendre comme englobant, entre autres, les bibliothèques nationales et les archives nationales et, dans la mesure où leurs archives et leurs bibliothèques <i>accessibles au public</i> sont concernées, les établissements d'enseignement, les organismes de recherche et les organismes publics de radiodiffusion.</p>

Hier sind die deutsche und die englische Sprachfassung klarer. Die öffentliche Zugänglichkeit bezieht Erwgr. 13 S. 1 DSM-RL nur auf Bibliotheken und Museen, nicht aber auf Archive und Einrichtungen des Film- und Tonerbes. In der französischen Fassung erstreckt sich die Zugänglichkeit wiederum nur auf Bibliotheken. In allen drei Sprachfassungen bleibt unklar, ob Nationalbibliotheken und Nationalarchive unabhängig von ihrer öffentlichen Zugänglichkeit als Einrichtungen des Kulturerbes gelten.

Bei Unterschieden in den gleichermaßen verbindlichen Sprachfassungen stellt der EuGH im Regelfall auf die Systematik und den Zweck der Regelung ab.²⁶ Der spricht für die englische Definition in Art. 2 Nr. 3 DSM-RL, die zudem von der deutschen und englischen Fassung von Erwgr. 13 unterstützt wird. Museen sind nach allgemeinen Definitionen darauf angelegt, jedenfalls Teile ihrer Sammlung öffentlich zugänglich zu machen. Reine Privatmuseen, die ihre Inhalte nur ausgewählten Personen zugänglich machen, sind dagegen nicht privilegierungswürdig. Anders wiederum dienen Archive und Einrichtungen des Film- und Tonerbes in erster Linie der Aufbewahrung und Erhaltung von Archivgut, nicht aber der Zugänglichmachung einer breiten Öffentlichkeit. Sie sind daher auch dann privilegierungswürdig, wenn sie nur einem eingeschränkten Personenkreis zugänglich sind. Im Gegenzug muss die privilegierte Forschung dann aber einen gewissen Gemeinwohlbezug haben (dazu unten c). Vor dem Hintergrund ist es sehr zu begrüßen, dass der deutsche Gesetzgeber den Bezugspunkt des notwendigen Zugangs für die Öffentlichkeit klargestellt hat.²⁷

bb) Begriff der Öffentlichkeit

Bibliotheken und Museen sind jedenfalls dann öffentlich zugänglich, wenn sie der allgemeinen Öffentlichkeit offenstehen. Davon unberührt bleibt die Möglichkeit, Eintrittsgelder zu erheben oder den Zugang von einer (jedem zustehenden) Mitgliedschaft abhängig zu machen.²⁸ Ob dagegen eine Teilöffentlichkeit ausreicht, etwa alle Mitglieder einer Universität oder einer öffentlich-

²⁶ EuGH, Urt. v. 27.09.2017 – verb. Rs. C-24/16 und C-25/16, ECLI:EU:C:2017:724 Tz. 72 – Nintendo/BigBen m.w.N.

²⁷ So sollte auch § 61 Abs. 2 UrhG verstanden werden, obwohl dieser auf einer anderen unionsrechtlichen Grundlage beruht (RL 2012/28/EU).

²⁸ BeckOK UrhR-*Hagemeyer*, 26. Ed. 15.7.2019, UrhG § 60e Rn. 19.

rechtlichen Rundfunkanstalt, wird für die deutschen Umsetzungen von InfoSoc- und Verwaiste-Werke-RL unterschiedlich ausgelegt.²⁹ Erwgr. 13 S. 2 DSM-RL spricht dagegen, weil andernfalls nicht erforderlich gewesen wäre, für Bibliotheken von „Bildungseinrichtungen, Forschungsorganisationen und öffentlich-rechtlichen Rundfunkanstalten“ die öffentliche Zugänglichkeit zu fordern, weil diese Einrichtungen typischerweise so groß sind, dass ihre Mitglieder sonst eine ausreichend große Teilöffentlichkeit bilden würden.

c) Nicht gewinnorientiert bzw. im öffentlichen Interesse tätig?

Forschungsorganisationen dürfen nicht gewinnorientiert sein oder müssen ihre Gewinne in die Forschung reinvestieren bzw. im öffentlichen Interesse tätig sein (§ 60d Abs. 2 S. 2 UrhG). Eine entsprechende Einschränkung fehlt in der Definition der Einrichtungen des Kulturerbes in § 60d Abs. 3 Nr. 1 UrhG sowie in Art. 2 Nr. 3 DSM-RL. Das lädt auf den ersten Blick zu einem Umkehrschluss ein, dass auch kommerziell tätige Kulturerbe-Einrichtungen privilegiert werden. Für eine teleologische Reduktion auf nicht-kommerzielle Zwecke spricht jedoch der systematische Einklang mit den Forschungsorganisationen (§ 60d Abs. 2 UrhG) und den Einzelforschern (§ 60d Abs. 3 Nr. 2 UrhG) sowie die Abgrenzung zur allgemeinen Text und Data Mining-Schranke in § 44b UrhG, die Nutzungen nur unter deutlich eingeschränkteren Voraussetzungen ermöglicht.

Es spricht viel dafür, dass der Unionsgesetzgeber dieses Merkmal deswegen nicht in die Definition des Art. 2 Nr. 3 DSM-RL aufgenommen hat, weil Art. 8 DSM-RL die Lizenzierung von vergriffenen Werken für Kulturerbe-Einrichtungen auf „nicht-kommerzielle Zwecke“ eingeschränkt hat. Ähnliche Einschränkungen finden sich auch in den Privilegierungen der InfoSoc-RL³⁰ und der Verwaiste-Werke-RL³¹.

Eine entsprechende Einschränkung fehlt in Art. 3 DSM-RL. Es ist anzunehmen, dass es sich hierbei um ein Redaktionsversehen handelt. Denn der ursprüngliche Kommissionsentwurf privilegierte nur Forschungsorganisationen,

²⁹ Dafür: Schricker/Loewenheim-Stieper, Urheberrecht, 6. Aufl. 2020, § 60e Rn. 5; Dreyer/Kotthoff/Meckel/Hentsch, Urheberrecht, 4. Aufl. 2018, § 60e Rn. 5; Dreier/Schulze, UrhG, 6. Aufl. 2018, § 60e Rn. 4 über Verweis auf § 27 Rn. 19; BeckOK UrhR-Hagemeyer, 26. Ed. 15.7.2019, UrhG § 60e Rn. 19. Dagegen: Wandtke/Bullinger-Jani, 5. Aufl. 2019, UrhG §§ 60e, 60f Rn. 8; Fromm/Nordemann-A. Nordemann, Urheberrecht, 12. Aufl. 2018, § 61 Rn. 12; Spindler, NJW 2008, 9, 13.

³⁰ Art. 5 Abs. 2 lit. c: „die keinen unmittelbaren oder mittelbaren wirtschaftlichen oder kommerziellen Zweck verfolgen“.

³¹ Art. 1 Abs. 1: „um die Ziele im Zusammenhang mit ihren im Gemeinwohl liegenden Aufgaben zu erreichen“.

die *per definitionem*³² nicht gewinnorientiert bzw. im öffentlichen Interesse handeln müssen. Eine entsprechende Einschränkung der Schranke war im Kommissionsentwurf deswegen nicht erforderlich. Als die Einrichtungen des Kulturerbes in den endgültigen Entwurf aufgenommen wurden, ist offensichtlich übersehen worden, dass eine entsprechende Einschränkung in deren Definition fehlte. Aus Erwgr. 11 Abs. 3 DSM-RL ergibt sich, dass der Richtliniengeber von einer gemeinwohlorientierten Ausrichtung der Einrichtungen des Kulturerbes ausgegangen ist. Zudem macht eine solche Einschränkung für die Abgrenzung zur allgemeinen Text und Data Mining Schranke in Art. 4 DSM-RL Sinn.

Es bietet sich an, für den Lückenschluss in entsprechender Weise auf die gemeinwohlorientierten Kriterien von § 60d Abs. 2 S. 2 UrhG/Art. 2 Nr. 1 DSM-RL zurückzugreifen. Die Einrichtungen des Kulturerbes dürfen daher entweder nicht gewinnorientiert sein, müssen ihre Gewinne in die weitere Forschung oder ihre hauptsächliche Tätigkeit investieren oder im Rahmen eines staatlich anerkannten Auftrags im öffentlichen Interesse tätig sein. In jedem Fall muss der Zugang bzw. die Entleihe daher nicht unentgeltlich sein, solange die Gebühren lediglich kostendeckend erhoben werden oder zur Aufgabenerfüllung reinvestiert werden.³³ In jedem Fall dürfen die Forschungsergebnisse nicht einem Unternehmen in bevorzugter Weise zugänglich gemacht werden.

d) Organisationsform

Wenn die Einrichtung des Kulturerbes im Sinn von c) gemeinwohlorientiert handelt, ist die konkrete Organisationsform und Trägerschaft unerheblich. Daher können sich auch privatwirtschaftlich organisierte und in privater Trägerschaft befindliche Bibliotheken, Museen und Archive auf die Schranke berufen.³⁴

³² Art. 2 Nr. 1 DSM-RL-E entspricht der verabschiedeten Fassung.

³³ So auch Dreier/Schulze, UrhG, 7. Aufl. 2022, § 60e Rn. 4.

³⁴ Vgl. auch Dreier/Schulze, UrhG, 7. Aufl. 2022, § 60e Rn. 4, § 61 Rn. 16; Schricker/Loewenheim-Spindler, Urheberrecht, 6. Aufl. 2020, § 61 Rn. 20 f.; Fromm/Nordemann-A. Nordemann, Urheberrecht, 12. Aufl. 2018, § 61 Rn. 13.

e) Berechtigte Personen

aa) Berechtigung nach § 60d Abs. 3 Nr. 1 UrhG

Nach dem Wortlaut der Schranke können sich lediglich die Einrichtungen des Kulturerbes auf sie berufen. Aus der Gesetzesbegründung ergibt sich aber, dass selbstverständlich auch die ihnen angehörenden Personen (in der Richtlinie eng.: „*persons attached thereto*“; frz.: „*personnes qui y sont rattachées*“) von der Schranke erfasst werden.³⁵ Wer Angehöriger einer Einrichtung des kulturellen Erbes ist, regelt das UrhG nicht. Es ist daher auf das Binnenrecht der jeweiligen Organisation abzustellen.³⁶ Dazu gehören in jedem Fall die Organe und Angestellten der Einrichtungen, nach der Gesetzesbegründung auch die individuellen Nutzer, wenn sie der Institution angehören.³⁷ Rechtssicherheit für Forscher können die Einrichtungen demnach dadurch herstellen, dass sie diese institutionell anbinden, etwa durch Forschungsmitgliedschaften oder Fellowships.

bb) Berechtigung nach § 60d Abs. 3 Nr. 2 UrhG

Forscher, die weder einer Forschungsorganisation noch einer Einrichtung des Kulturerbes angehören, werden nach § 60d Abs. 3 Nr. 2 UrhG begünstigt, unterliegen aber nach Abs. 5 Einschränkungen bei der Aufbewahrung der Forschungskorpora nach Abschluss ihrer Forschungsarbeiten. Hintergrund ist, dass sie nicht vom institutionenbezogenen Ansatz der europäischen Text und Data Mining-Schranke des Art. 3 DSM-RL erfasst werden. Der deutsche Gesetzgeber konnte deren Nutzungen für die nicht-kommerzielle wissenschaftliche Forschung daher nur auf Art. 5 Abs. 3 lit. a InfoSoc-RL-RL stützen,³⁸ der eine eingeschränkte Reichweite hat. Für Einzelforscher sind die öffentlich zugänglichen Bestände von Einrichtungen des Kulturerbes i.S.v. § 44b Abs. 2 UrhG rechtmäßig zugänglich und können für Text und Data Mining-Aktivitäten genutzt werden.

2. Bibliotheken von Forschungseinrichtungen

Bibliotheken von Forschungseinrichtungen sind als solche Einrichtungen des Kulturerbes, wenn sie öffentlich zugänglich sind.³⁹ Darüber hinaus werden

³⁵ BT-Drs. 19/27426, 96; Erwgr. 14 S. 1 DSM-RL.

³⁶ *Raue*, ZUM 2021, 793, 801.

³⁷ RegE BT-Drs. 19/27426, 97.

³⁸ Erwgr. 15 S. 5, Art. 25 DSM-RL; *Stieper*, GRUR 2020, 1, 4.

³⁹ Erwgr. 13 S. 2 DSM-RL. Dazu oben 1.b.

diese Bibliotheken auch von der Privilegierung ihrer Forschungsorganisation erfasst.⁴⁰ Sie werden daher auch dann privilegiert, wenn sie nicht i.S.v. § 60d Abs. 3 Nr. 1 UrhG öffentlich zugänglich sind. Allerdings muss die Forschungsorganisation dann die weiteren Voraussetzungen des Abs. 2 erfüllen, also nicht gewinnorientiert sein oder im öffentlichen Interesse tätig werden.⁴¹

IV. Freigestellte Handlungen

1. Vervielfältigungen und Entnahmen aus Datenbanken

§ 60d Abs. 1 i.V.m. § 44b Abs. 1, 2 UrhG erlaubt alle Handlungen zum Zweck des Text und Data Mining mit wissenschaftlicher Zielsetzung, mit denen nach dem Urheberrecht geschützte Gegenstände vervielfältigt bzw. aus einer Datenbank entnommen werden. Dazu gehören:

- die Digitalisierung analoger Quellen („digitalen oder digitalisierten Werken“, § 44b Abs. 1 UrhG);⁴²
- die Zusammenstellung eines digitalen Text-, Bild- oder sonstigen Datenkorpus;
- die Normalisierung, Standardisierung, Vorstrukturierung und sonstige Anreicherung des Korpus (vgl. § 23 Abs. 3 UrhG);
- das eigentliche Minen des Korpus, soweit dafür Vervielfältigungshandlungen erforderlich sind;⁴³
- die dauerhafte Speicherung und Aufbewahrung des vervielfältigten Materials (wobei allerdings angemessene Sicherheitsvorkehrungen getroffen werden müssen, dazu unten V.).

2. Handlungen der öffentlichen Wiedergabe

Nach § 60d Abs. 4 UrhG dürfen Datenkorpora auch größeren Forschungsgruppen für die gemeinsame wissenschaftliche Forschung öffentlich zugänglich

⁴⁰ RegE BT-Drs. 19/27426, 96; Art. 2 Nr. 1 DSM-RL sowie Erwgr. 12 S. 3; *Steinhauer*, RuZ 2021, 5, 8 unter einer CC BY-SA Lizenz.

⁴¹ Näheres zu diesen Voraussetzungen *Raue*, ZUM 2021, 793, 800 f.

⁴² Das ist richtlinienkonform *Spindler*, CR 2019, 277, 279; *Raue*, ZUM 2019, 684, 687.

⁴³ Andernfalls sind sie urheberrechtlich irrelevant, Erwgr. 9 DSM-RL. Ferner *Dreier/Schulze*, UrhG, 7. Aufl. 2022, § 60d Rn. 1; *Fromm/Nordemann-A. Nordemann*, Urheberrecht, 12. Aufl. 2018, § 60d Rn. 4; *Dreyer/Kotthoff/Meckel/Hentsch*, Urheberrecht, 4. Aufl. 2018, § 60d Rn. 3.

gemacht werden, wenn diese auf einen bestimmt abgegrenzten Kreis von Personen abgegrenzt sind.⁴⁴

Die Text und Data Mining-Schranke stellt nur die eigentliche Forschungstätigkeit frei, nicht aber deren Präsentation. Die Präsentation der Forschungsergebnisse greift jedoch im Regelfall ohnehin nicht in das Urheberrecht der Ausgangsmaterialien ein, weil die in ihnen enthaltenen Informationen als solche frei sind. Sollte im Einzelfall etwa präsentiert werden, wie und welche Informationen aus einem im Einzelnen dargestellten Bild oder Text extrahiert werden konnten, so muss dafür auf andere Schranken zurückgegriffen werden, etwa auf die Zitat-schranke (§ 51 UrhG).

3. Zum Zweck der wissenschaftlichen Forschung

Die Handlung muss zum Zweck der wissenschaftlichen Forschung vorgenommen werden.⁴⁵ Die Privilegierung von Einrichtungen des Kulturerbes wird damit begründet, dass diese „möglicherweise ebenfalls Forschung im Zusammenhang mit ihrer hauptsächlichen Tätigkeit betreiben könnten“.⁴⁶ So gehört es etwa zum grundlegenden Auftrag von Museen, Zeugnisse zur Gewinnung und Erweiterung von Wissen zu sammeln und das Verständnis des Natur- und Kulturerbes zu fördern.⁴⁷ Auch die übrigen Einrichtungen des Kulturerbes sind keine Verwahranstalten des kulturellen Erbes, sondern auf dessen Erschließung und Fruchtbarmachung ausgerichtet. Daher lässt sich aus der Begründung keine Einschränkung des zulässigen Forschungszwecks ableiten, solange wissenschaftliche Standards gewahrt bleiben.

V. Kein Vorbehalt der Rechteinhaber möglich

Die Nutzung der allgemeinen Text und Data Mining-Schranke § 44b UrhG können Rechteinhaber dadurch verhindern, dass sie einen Vorbehalt erklären (§ 44b Abs. 3 S. 1 UrhG). Dieser beschränkt aber nicht die Freistellung nach § 60d UrhG.⁴⁸ Ganz allgemein dürfen aus § 44b Abs. 2, 3 UrhG keine Einschränkungen für die Wissenschaftsschranke in § 60d UrhG abgeleitet werden.⁴⁹

⁴⁴ Im Regelfall stellen Forschergruppen aber keine Öffentlichkeit i.S.v. Art. 3 InfoSoc-RL = § 19a UrhG dar, BT-Drucks. 18/12329, 41; *Spindler*, CR 2019, 277, 280. Die Freistellung kann nicht auf Art. 3 DSM-RL, sondern nur Art. 5 Abs. 3 lit. a InfoSoc-RL gestützt werden.

⁴⁵ Dazu ausführlicher *Raue*, ZUM 2019, 684, 688 f.; *Spindler*, CR 2019, 277, 278.

⁴⁶ Erwgr. 8 S. 4 DSM-RL.

⁴⁷ ICOM, Ethische Richtlinien für Museen von ICOM (oben Fn. 15), Richtlinie 3 und 4.

⁴⁸ BT-Drs. 19/27426, 88 f.; Erwgr. 18 UAbs. 2 S. 3 DSM-RL.

⁴⁹ Art. 4 Abs. 4 DSM-RL; *Raue*, ZUM 2021, 793, 798.

VI. Vergütung

Im Anschluss an die Vorgaben aus Erwgr. 17 S. 2 DSM-RL ist das Text und Data Mining nun – anders als unter der Vorgängerschranke des § 60d UrhG a.F. – vergütungsfrei ausgestaltet (§ 60h Abs. 2 Nr. 3 UrhG n.F.).⁵⁰ Das ist dadurch gerechtfertigt, weil nur das Minen von solchen Inhalten freigestellt wird, zu denen die Nutzer rechtmäßigen Zugang haben (§ 44b Abs. 2 S. 1 UrhG). Rechteinhaber können darüber ihre Vergütung steuern; sie müssen dann aber hinnehmen, dass auf die urheberrechtlich nicht geschützten Informationen in ihren Schutzgegenständen zugegriffen wird.

VII. Einrichtungen des Kulturerbes als „vertrauenswürdige Stellen“ (Erwgr. 15 DSM-RL)

Nach Abschluss eines Forschungsprojekts besteht folgendes Dilemma:⁵¹ Die Forscher haben ein erhebliches Interesse daran, dass sie die Forschungskorpora samt Zwischenergebnissen weiter aufbewahren dürfen. Zum einen verlangen die Regeln guter wissenschaftlicher Praxis, dass Forscher alle für das Zustandekommen eines Forschungsergebnisses relevanten Informationen nachvollziehbar dokumentieren, archivieren und berechtigten Interessenten zur Überprüfung zugänglich machen müssen.⁵² Zum anderen haben sie die Ursprungsmaterialien mit zum Teil erheblichem Aufwand für die algorithmische Durchsichtung aufbereitet, so dass sie diese für weitere Forschungsvorhaben weiterverwenden wollen. Verlage und andere Rechteinhaber haben dagegen die Sorge, dass durch die dauerhafte Speicherung umfangreicher Ursprungsmaterialien Schattenbibliotheken entstehen, die etwa die Nutzung der von ihnen angebotenen Datenbanken substituieren können.⁵³

⁵⁰ Das begrüßt etwa *Steinhauer*, RuZ 2021, 5, 21 unter einer CC BY-SA Lizenz.

⁵¹ Dazu bereits *Raue*, ZUM 2019, 684, 688.

⁵² Vgl. etwa *DFG*, Leitlinien zur Sicherung guter wissenschaftlicher Praxis, Kodex, 2019, Leitlinie 12, 13, 17 sowie *Max-Planck-Gesellschaft*, Regeln zur Sicherung guter wissenschaftlicher Praxis, abrufbar unter https://www.mpg.de/229457/Regeln_guter_wiss_Praxis__Volltext-Dokument_.pdf, zuletzt abgerufen am 05.08.2022.

⁵³ Vgl. etwa die Stellungnahme zur Umsetzung der DSM-RL der International Association of Scientific, Technical and Medical Publishers (“STM”) vom 06.09.2019, abrufbar unter https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Stellungnahmen/2019/Downloads/090619_Stellungnahme_STM_EU-Richtlinien_Urheberrecht.pdf?__blob=publicationFile&v=5, zuletzt abgerufen am 05.08.2022.

Der deutsche Gesetzgeber hat diesen Zielkonflikt in § 60d Abs. 5 UrhG so gelöst, dass Berechtigte die Datenkorpora aufbewahren dürfen, allerdings zeitlich begrenzt, solange die Korpora noch für Zwecke der wissenschaftlichen Forschung oder zur Überprüfung wissenschaftlicher Erkenntnisse erforderlich sind.⁵⁴

Darüber hinaus sind die nutzenden Institutionen verpflichtet, „angemessene Sicherheitsvorkehrungen“ gegen unbefugte Nutzungen zu treffen. Der deutsche Gesetzgeber hat diese Voraussetzungen selbst nicht präzisiert, sondern in der Gesetzesbegründung weitgehend auf die Vorgaben der DSM-RL verwiesen.⁵⁵ Die Sicherheitsanforderungen dürfen demnach nicht über das hinausgehen, was für die sichere Aufbewahrung der Kopien erforderlich ist, und sie dürfen Text und Data Mining-Aktivitäten nicht unverhältnismäßig beeinträchtigen. Nach Erwgr. 15 S. 2 DSM-RL sollen die digitalen Kopien in einer „sicheren Umgebung“ gespeichert werden.

Diese sichere Umgebung muss nicht notwendig von der berechtigten Stelle selbst angeboten werden. Denn zum einen ermöglicht Erwgr. 15 S. 3 DSM-RL den Mitgliedstaaten, vertrauenswürdige Stellen zu benennen, die eine solche sichere Aufbewahrung sicherstellen können. Zum anderen ergibt sich aus Erwgr. 11 Abs. 3 DSM-RL, dass privilegierte Institutionen für die Durchführung des Text und Data Mining auf die Hilfe Dritter zurückgreifen dürfen; das muss *pars pro toto* auch für die Aufbewahrung der Datenkorpora gelten. Das Angebot zentraler, institutionenübergreifender Repositorien ist insofern sinnvoll, weil so einzelne, insbesondere kleinere Institutionen nicht selbst eine sichere, aber dennoch komfortabel nutzbare Speichermöglichkeit entwickeln und dauerhaft aufrechterhalten müssen. Hierin kann insbesondere für Bibliotheken und Archive eine Chance liegen, etwa im Rahmen der Förderung der Nationalen Forschungsdateninfrastruktur (NFDI).

VIII. Ergänzende Freistellung nach § 44b UrhG

Forschende an Kulturerbe-Einrichtungen können sich weiter auf andere Schranken des UrhG berufen.⁵⁶ Das kann insbesondere für die automatisierte

⁵⁴ Dazu näher *Rauc*, ZUM 2021, 793, 798 f. Die Löschungspflicht kritisieren etwa *Brinkbus*, RuZ 2021, 56, 65 unter einer CC BY-SA Lizenz; *Heesen/Jüngels*, RuZ 2021, 45, 50 f. unter einer CC BY-SA Lizenz.

⁵⁵ RegE BT-Drs. 19/27426, 97.

⁵⁶ Erwgr. 5 S. 6 DSM-RL.

Analyse von Computerprogrammen notwendig sein, die nicht von § 60d UrhG, wohl aber von § 44b UrhG freigestellt wird (§ 69d Abs. 6 UrhG).⁵⁷

IX. Fazit

§ 60d UrhG n.F. erlaubt Gedächtnisinstitutionen nun ausdrücklich, auch ihre urheberrechtlich geschützten Text-, Bild- und andere Bestände systematisch mit algorithmischer Hilfe zu erschließen und ihnen mit Hilfe von Text und Data Mining neue Informationen zu entlocken.

In dem Beitrag wird aufgezeigt, welche Institutionen § 60d Abs. 3 Nr. 1 UrhG als Kulturerbe-Einrichtungen ansieht und welcher Personenkreis die von der Schranke freigestellten Handlungen vornehmen darf (dazu III.). Anschließend werden unter IV. die von § 60d UrhG freigestellten Handlungen und unter V. das Konzept einer „vertrauenswürdige Stelle“ vorgestellt, die von den Mitgliedstaaten für die sichere Aufbewahrung von Datenkorpora benannt werden können.

Mit diesen neuen gesetzlichen Möglichkeiten geht zugleich die Verpflichtung von Kulturerbe-Einrichtungen einher, sie zu nutzen und die digitalen Möglichkeiten der Wissenserschließung zu nutzen.⁵⁸

⁵⁷ Zu den Hintergründen *Raue*, ZUM 2021, 793, 800; *ders.*, ZUM 2019, 684, 689.

⁵⁸ So auch *Döbl*, RuZ 2020, 195, 209 unter einer CC BY-SA Lizenz: „es entsteht erstens eine politische Pflicht/Verantwortung, sich aktiv mit dem Bereich TDM zu beschäftigen, ihn gar strategisch als Priorität auszubauen. Und es entsteht zweitens das Privileg, selbst inhaltlich die Agenda mitzugestalten, wo und wie TDM zum Zwecke wissenschaftlicher Forschung eingesetzt wird.“

C. Das Zitat und dessen Rahmen für Belege bei Textanalysen

Katharina Erler-Fridgen

Version 1.0 (02.06.2022), CC BY-SA 4.0.

Werden Ergebnisse von Textanalysen präsentiert, können Texte oder Textteile als Beleg für deren Qualität oder zur Überprüfung der Plausibilität hilfreich sein.¹ Neben dem Rahmen der Text und Data Mining-Schranken für Vervielfältigungen im Verlauf der Textanalyse² bietet das Zitatrecht einen Freiraum für die geistige Auseinandersetzung mit urheberrechtlich geschützten Texten und Textteilen.³ Im Folgenden soll die Schranke der Zitierfreiheit nach § 51 UrhG und deren Voraussetzungen erläutert werden. Ausgegangen wird von den Voraussetzungen des Zitatrechts, die sich entlang der vom Gesetzgeber in den Blick genommenen klassischen Fällen der Gewährleistung der geistigen Auseinandersetzung mit fremden Gedanken herausgebildet haben.⁴ Ein solcher klassischer Fall aus den Literaturwissenschaften wäre etwa die Entlehnung von Passagen eines Gedichts im Rahmen einer Gedichtinterpretation. Die vorliegende Handreichung will zum einen diese allgemeinen Voraussetzungen des Zitatrechts beschreiben und zum anderen darauf aufbauend insbesondere eine Konstellation beleuchten, die aus Sicht der Digital Humanities ein Desiderat darstellt: Die Nutzung von Texten oder Textteilen als Beleg bei automatisierten Textanalysen, um so die Plausibilität der Analyseergebnisse überprüfbar zu machen. Im Folgenden wird also auch darauf eingegangen, welche Voraussetzungen des Zitatrechts in dieser Konstellation in den Blick zu nehmen sind. Die Handrei-

¹ Zum urheberrechtlichen Schutz von Textteilen und Beispiele für deren Einsatz als Beleg bei Textanalysen siehe *Erler-Fridgen*, Teil 6, A. in diesem Band.

² Zu den Text und Data Mining-Schranken und dazu, dass eine Präsentation von Ausgangstexten an die Allgemeinheit hiervon nicht erfasst ist, siehe *Erler-Fridgen*, Teil 4, A. in diesem Band.

³ Zu den Kriterien der urheberrechtlichen Schutzfähigkeit siehe *Erler-Fridgen*, Teil 2, A. in diesem Band; zum urheberrechtlichen Schutz von Textteilen siehe *Erler-Fridgen*, Teil 6, A. in diesem Band.

⁴ *Schulz*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn. 1.

chung baut für diese Konstellation gedanklich auf einer Handreichung zur Präsentation von Textteilen als Ergänzung von Textanalysen⁵ auf, da das Zitatrecht grundsätzlich nur dann greift, wenn die als Beleg genutzten Textteile auch urheberrechtlich schutzfähig sind.⁶

I. Hintergrund der Belege bei Textanalysen⁷

Um die Qualität von Textanalysen überprüfen zu können, ist es aus Sicht der Digital Humanities hilfreich, Textteile als Nachweis und Quelle des Analyseergebnisses zu präsentieren.

Bei dem Verfahren der Named Entity Recognition können solche Nachweise helfen, die Ausgangsbasis für Analyseergebnisse transparent und nachvollziehbar zu machen.⁸ Beispielsweise können bei der Erkennung von Entitäten wie den Namen der Autorinnen und Autoren durch Abkürzungen oder Umschreibungen der Autorinnen und Autoren Zweifelsfälle entstehen, die auf diese Weise nachvollziehbar gemacht werden können.⁹ Sollen beispielsweise durch das Verfahren der Named Entity Recognition Namen von Autoren und Autorinnen literaturwissenschaftlichen Erläuterungen erkannt werden, können Umschreibungen der Autorin oder des Autors oder Namensabkürzungen zu Zweifelsfällen führen, weshalb es hilfreich sein kann, den zugehörigen Textteil zu hinterlegen.¹⁰ Auch wenn Werktitel und Figurenname eines Werkes übereinstimmen, kann ein Nachweis bereits auf Wortebene das Ergebnis der Analyse nachvollziehbar machen.¹¹ In diesen Fällen können bereits ein Wort oder der direkte Kontext von zwei bis drei Wörtern vor und nach der Entität oder auch ein Satzteil als Beleg genügen.

Auch bei der Analyse von Themenaussagen, der sogenannten Statement Extraction, können Belege aus Sicht der Digital Humanities hilfreich sein. Hier ist bei-

⁵ *Erler-Fridgen*, Teil 6, A. in diesem Band.

⁶ *Dustmann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 13; zur Schutzfähigkeit von Textteilen siehe *Erler-Fridgen*, Teil 6, A. in diesem Band; zu den allgemeinen Voraussetzungen der urheberrechtlichen Schutzfähigkeit siehe *Erler-Fridgen*, Teil 2, A. in diesem Band.

⁷ Dank gilt Dr. Maria Hinzmann, Prof. Dr. Benjamin Raue und Prof. Dr. Christof Schöch (MiMoText, Universität Trier) für die Diskussionen zum Thema.

⁸ *Erler-Fridgen*, Teil 6, A. I. in diesem Band.

⁹ Zum genannten Beispiel siehe *Erler-Fridgen*, Teil 6, A. I. in diesem Band.

¹⁰ *Erler-Fridgen*, Teil 6, A. I. in diesem Band.

¹¹ *Erler-Fridgen*, Teil 6, A. I. in diesem Band: So zum Beispiel bei Voltaires Werk „Candide“, in dem Werktitel und Figurenname „Candide“ übereinstimmen.

spielsweise bei der Extraktion von Themenaussagen im Rahmen des Information Retrieval¹² in der Struktur von Subjekt-Prädikat-Objekt (sog. Resource Description Framework/RDF-Triple) der Nachweis des zugrundeliegenden Satzes ein Desiderat der Digital Humanities.¹³ Denn der zugrundeliegende Satz, aus dem das Triple generiert wurde, macht eine Überprüfung der extrahierten Themenaussage möglich.¹⁴ Ein Beispiel für eine solche Themenaussage wäre ‚Candide‘ à ABOUT à ‚Theodizee‘ im Werk „Candide“ von Voltaire.¹⁵ Der diesem Triple zugrundeliegende Satz „Voltaires Candide ou L’optimisme (1759) lässt sich als pointierte Absage an die Theodizee und den Leibnizschen Optimismus lesen“¹⁶ könnte als Nachweis der ausgelesenen Themenaussage hilfreich sein.¹⁷

Wichtiger Hintergrund jedoch ist, dass solche Textteile eigenständig urheberrechtlich geschützt sein können, sodass deren Vervielfältigung bzw. Präsentation urheberrechtlich relevante Eingriffe darstellen können (hierzu bereits *Erler-Fridgen*, Teil 6, A. in diesem Band).¹⁸ Daher muss in diesem Fall die Zustimmung der Rechteinhaber eingeholt werden oder auf die urheberrechtlichen Schranken in § 44a ff. UrhG zurückgegriffen werden.¹⁹ Die Text und Data Mining Schranke ermöglicht in § 60d Abs. 4 UrhG eine öffentliche Zugänglichmachung von Vervielfältigungen lediglich für einen abgegrenzten Personenkreis für die gemeinsame wissenschaftliche Forschung oder für einzelne Dritte zur Qualitätsüberprüfung.²⁰ Eine Präsentation von Vervielfältigungen über diese Grenze hinaus wird durch die TDM-Schranke jedoch nicht möglich gemacht.²¹ In diesem Spannungsfeld werden die Rahmenbedingungen des Zitatrechts in den Blick genommen, da das Zitatrecht einen gewissen Freiraum für die geistige Auseinandersetzung²² mit urheberrechtlich geschützten Textteilen bietet.

¹² *Ignatow/Mihalcea*, Text Mining, 137 ff.

¹³ *Erler-Fridgen*, Teil 6, A. I. in diesem Band.

¹⁴ *Erler-Fridgen*, Teil 6, A. I. in diesem Band.

¹⁵ Ausführlich hierzu bereits: *Erler-Fridgen* Teil 6, A. I. in diesem Band.

¹⁶ *Schlüter*, Von der Aufklärung bis zur Französischen Revolution, in: Grimm/Hartwig, Französische Literaturgeschichte. 6. Auflage, 2014, 196, 201.

¹⁷ *Erler-Fridgen*, Teil 6, A. I. in diesem Band.

¹⁸ Zum eigenständigen urheberrechtlichen Schutz und daraus resultierenden möglichen urheberrechtlichen Eingriffen siehe *Erler-Fridgen*, Teil 6. A. in diesem Band.

¹⁹ Hierzu: *Erler-Fridgen*, Teil 6, A. IV. in diesem Band.

²⁰ *Erler-Fridgen*, Teil 4, A. IV. in diesem Band zum Begriff der „öffentlichen Zugänglichmachung“; ebenda IV.: Die Zugänglichmachung zur Qualitätsüberprüfung erfasst wohl vornehmlich die Phase des Peer-Review-Verfahrens vor der Veröffentlichung.

²¹ *Erler-Fridgen*, Teil 4, A. IV. in diesem Band.

²² Entwurf eines Zweiten Gesetzes zur Regelung des Urheberrechts in der Informationsgesellschaft, BT-Drs. 16/1828, 25.

Werden Textteile als Nachweise für Textanalysen herangezogen, so stehen insbesondere auch der Zitzweck und der hierdurch gebotene Umfang des Zitats in Frage.²³ Auch der Charakter des zitierenden Gegenstands, die urheberrechtliche Selbstständigkeit kann relevant sein, wenn Analyseergebnisse (digital) gebündelt werden.²⁴ Schließlich kann bei einer digitalen Darstellung zudem die Einbindung des Belegs²⁵ in Rede stehen. Im Folgenden werden die Rahmenbedingungen des Zitatrechts und dabei auch unter anderem die genannten Anforderungen dargestellt.

II. Allgemeine Voraussetzungen des Zitatrechts

§ 51 UrhG legt allgemeine Voraussetzungen der Zitierfreiheit fest. Dabei ist die gemeinsame Voraussetzung aller Zitatfälle (§ 51 S. 2 Nr. 1 - 3 UrhG) ein Zitzweck, der den gebotenen Umfang eines zulässigen Zitats bestimmt.²⁶ Das Zitatrecht in § 51 UrhG erfasst in S. 2 Nr. 1 - 3 drei Fälle zulässiger Zitate, die gesonderte Kriterien voraussetzen.²⁷ Diese Regelbeispiele sind jedoch keine abschließenden Regelungen, sondern ergänzen sich und können auch nebeneinander anwendbar sein.²⁸ Schließlich wurde 2007 § 51 S. 1 UrhG als Generalklausel ausgestaltet.²⁹ Zweck der Generalklausel ist es nach der Gesetzesbegründung, einzelne Lücken zu schließen, die aus der unflexiblen Grenzziehung der früher abschließend geregelten Regelbeispiele folgen.³⁰

1. Charakteristika des zitierten Textes

Die Schranke der Zitierfreiheit nach § 51 UrhG dient der Freiheit der geistigen Auseinandersetzung.³¹ Liegen die Voraussetzungen des § 51 UrhG vor, sind sämtliche Verwertungsrechte, also insbesondere die Vervielfältigung, Verbreitung und öffentliche Wiedergabe – mit Ausnahme des Ausstellungsrechts nach

²³ Siehe zum Zitzweck und der geistigen Auseinandersetzung unten jeweils in II., III. und IV.

²⁴ Siehe zur Selbstständigkeit des zitierenden Werks unten in II.

²⁵ Zu Verlinkungen siehe unten in II.

²⁶ *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 18; *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 3.

²⁷ Siehe dazu unten III. und IV.

²⁸ *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 19.

²⁹ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 22; siehe Entwurf eines Zweiten Gesetzes zur Regelung des Urheberrechts in der Informationsgesellschaft, BT-Drs. 16/1828, 25.

³⁰ Entwurf eines Zweiten Gesetzes zur Regelung des Urheberrechts in der Informationsgesellschaft, BT-Drs. 16/1828, 25.

³¹ *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 10 f.

§ 18 Abs. 1 Nr. 1 UrhG – zulässig.³² Der Schranke bedarf es nur, wenn der zitierte Gegenstand urheberrechtlich schutzfähig nach § 2 Abs. 2 UrhG oder durch ein Leistungsschutzrecht geschützt ist.³³ Das Zitatrecht nach § 51 UrhG ist demnach beispielsweise auch dann einschlägig, wenn Textteile genutzt werden, die dem Leistungsschutzrecht der wissenschaftlichen Ausgabe nach § 70 UrhG unterliegen.³⁴ Auch einzelne Werkteile können urheberrechtsschutzfähig sein und deren Nutzung in diesem Fall auch die Berufung auf das Zitatrecht erfordern.³⁵ Fehlt es hingegen an der Individualität von Werkteilen und damit an der urheberrechtlichen Schutzfähigkeit der zitierten Stelle, liegt durch deren Übernahme keine urheberrechtliche Verletzung vor.³⁶ Auch kleinste Textteile können grundsätzlich schutzfähig sein, jedoch gilt typischerweise: Je kürzer ein Textteil ist, desto höhere Anforderungen sind an seine Originalität zu stellen, um eine eigenschöpferische Prägung und damit Individualität nach § 2 Abs. 2 UrhG festzustellen.³⁷ Für Belege bei Textanalysen können beispielsweise Nachweise mit unterschiedlichen Umfängen hilfreich sein: So könnte bei dem Verfahren der Named Entity Recognition zum Beispiel lediglich die Bezeichnung der in Frage stehenden Entität selbst (etwa ein Autoren- oder Autorinnenname) oder der Kontext der Entität (wenige Worte vor und nach der analysierten Entität) oder schließlich der gesamte Satz, in dem die Entität enthalten ist, als Beleg eingesetzt werden. Im Falle des Beispiels der Statement Extraction wird hingegen aus Sicht der Digital Humanities der gesamte Satz den Mindestumfang eines Beleges darstellen.

³² *Schulz*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn. 7.

³³ *Dustmann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 13; zur Voraussetzung der urheberrechtlichen Schutzfähigkeit siehe *Erler-Fridgen*, Teil 2, A. in diesem Band.

³⁴ *Erler-Fridgen*, Teil 2, C. in diesem Band.

³⁵ *Dustmann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 13; zu den Voraussetzungen der Schutzfähigkeit von Werkteilen siehe *Erler-Fridgen*, Teil 6, A. in diesem Band.

³⁶ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 545b; zur urheberrechtlichen Schutzfähigkeit von Werkteilen siehe *Erler-Fridgen*, Teil 6, A. in diesem Band.

³⁷ Ausführlich hierzu *Erler-Fridgen*, Teil 6, A. II. 1. in diesem Band. mit weiteren Hintergründen zur hinreichenden Individualität von Textteilen sowie zu EuGH C-5/08, ECLI:EU:C:2009:465, GRUR 2009, 1041 Rn. 38 – Infopaq/DDF, in dem der EuGH geurteilt hat, dass es nicht ausgeschlossen ist, dass schon ein Auszug von 11 Wörtern aus einem Pressartikel schutzfähig ist.

Gemeinfreie Werke und Werke oder Werkteile, die nicht urheberrechtlich schutzfähig sind, können frei genutzt und müssen daher nicht erst durch die Zitatschranke freigestellt werden.³⁸ Umgekehrt gilt aber: Wenn urheberrechtlich oder leistungsschutzrechtlich geschützte Werkteile etwa als Nachweise bei Textanalysen präsentiert werden, muss die Nutzung auf eine urheberrechtliche Schranke gestützt werden können (§ 44a ff. UrhG)³⁹ und es sind die Voraussetzungen des Zitatrechts zu prüfen.

Der zitierte Text muss ein nach § 6 Abs. 1 UrhG veröffentlichtes Werk sein. Hierzu muss das Werk mit Zustimmung des Berechtigten in seiner konkreten Fassung nach § 6 Abs. 1 UrhG an die Öffentlichkeit gerichtet und ihr tatsächlich zugänglich sein.⁴⁰

2. Zitatzwecke

Entscheidende Voraussetzung des § 51 UrhG ist der Zitatzweck.⁴¹ Die Nutzung fremder Werke oder Werkteile muss nicht allein dem Zitatzweck dienen, dieser muss jedoch andere Zwecke (etwa die Schmuckwirkung s.u.) überwiegen.⁴² Von diesem Zweck ist der zulässige Umfang des Zitats nach § 51 UrhG abhängig.⁴³ Ausdrücklich erläutert wird dieser Zweck lediglich für das wissenschaftliche Großzitat⁴⁴ nach § 51 S. 2 Nr. 1 UrhG, wonach das zitierte Werk zur Erläuterung des Inhalts des aufnehmenden (zitierenden) Werkes angeführt werden muss.⁴⁵ Der Erläuterungszweck kann auch jedes übrige Zitat, bspw. das

³⁸ *Dustmann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 13.

³⁹ Oder es muss die Zustimmung der Rechteinhaber eingeholt werden.

⁴⁰ *Schulz*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn. 8.

⁴¹ BGH GRUR 1986, 59, 60 – Geistchristentum; *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 27.

⁴² BGH NJW 1968, 1875, 1878 – Kandinsky.

⁴³ *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 27.

⁴⁴ Siehe hierzu III.; das wissenschaftliche Großzitat nach § 51 S. 2 Nr. 1 UrhG gestattet unter gewissen Voraussetzungen die Aufnahme einzelner ganzer Werke in ein selbstständiges wissenschaftliches Werk.

⁴⁵ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 3.

Kleinzitat⁴⁶ rechtfertigen. Es kommen aber darüber hinaus auch weitere Zitzwecke in Betracht.⁴⁷ Für alle anderen Fälle ist der Zitzweck durch das Ergänzen von erkennbar fremden Werken oder Werkteilen charakterisiert.⁴⁸ Die InfoSoc-Richtlinie (2001/29/EG)⁴⁹, in deren Lichte § 51 UrhG auszulegen ist,⁵⁰ spricht in Art. 5 Abs. 3 lit. d von „Zitaten zu Zwecken wie Kritik und Rezension“, wobei diese Aufzählung lediglich Beispielcharakter aufweist und nicht abschließend ist.⁵¹

Nach der Rechtsprechung des BGH erfordert der allgemeine Zitzweck, dass eine innere Verbindung zwischen den verwendeten fremden Werken oder Werkteilen und den eigenen Gedanken des Zitierenden hergestellt wird.⁵² Ein Zitat ist nach dem allgemeinen Zitzweck auch nur dann zulässig, wenn es als Belegstelle oder Erörterungsgrundlage für selbstständige Ausführungen des Zitierenden genutzt wird.⁵³ Bei Textanalysen muss der Zitierende die Textteile also nutzen, um in irgendeiner Form die eigenen Ausführungen zu belegen.⁵⁴ In der Folge des Urteils Spiegel Online/Volker Beck des EuGH (s.u.)⁵⁵ präziserte der BGH diese Kriterien und stellte zunächst fest, dass diese im Einklang mit den unionsrechtlichen Grundlagen des Zitatrechts nach Art. 5 Abs. 3 lit. d InfoSoc-Richtlinie (2001/29/EG) stehen.⁵⁶ Hiernach bestünde das wesentliche Merkmal des Zitats darin, dass ein Werk oder ein Auszug aus einem Werk genutzt wird, um Aussagen zu erläutern, eine Meinung zu verteidigen oder eine geistige

⁴⁶ Siehe hierzu IV.; das Kleinzitat nach § 51 S. 2 Nr. 2 UrhG erlaubt, dass unter gewissen Voraussetzungen Stellen eines Werkes zitiert werden; dies ist auch zu anderen als wissenschaftlichen Zwecken zulässig.

⁴⁷ Lüft, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 3.

⁴⁸ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 3.

⁴⁹ Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft.

⁵⁰ Spindler, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 7.

⁵¹ EuGH C-516/17, ECLI:EU:C:2019:625, Rn. 79 – Spiegel Online/Volker Beck; Schulz, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn. 12; Spindler, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 27.

⁵² BGH GRUR 2016, 368 Rn. 25 – Exklusivinterview; BGH MMR 2010, 475 Rn. 26 – Vorschaubilder; BGH GRUR 1986, 59, 60 – Geistchristentum.

⁵³ BGH GRUR 2016, 368 Rn. 25 – Exklusivinterview; BGH MMR 2008, 536, 539 – TV-Total; BGH GRUR 1986, 59, 60 – Geistchristentum.

⁵⁴ Zur Belegfunktion bei Textteilen als Nachweisen bei Textanalysen siehe unten III. b) (1).

⁵⁵ EuGH C-516/17, ECLI:EU:C:2019:625, Rn. 78 – Spiegel Online/Volker Beck.

⁵⁶ BGH GRUR 2020, 859, Rn. 83 – Reformistischer Aufbruch II.

Auseinandersetzung zwischen dem Werk und den Aussagen des Nutzers zu ermöglichen.⁵⁷ Dabei müsse der Zitierende zwingend eine direkte und enge Verknüpfung zwischen dem zitierten Werk und seinen eigenen Überlegungen herstellen.⁵⁸ Außerdem dürfe das Zitat nach Art. 5 Abs. 5 InfoSoc-Richtlinie (2001/29/EG) nicht so umfangreich sein, dass es die normale Verwertung des Werkes oder eines sonstigen Schutzgegenstands beeinträchtigt oder die berechtigten Interessen des Rechtsinhabers ungebührlich verletzt.⁵⁹

Nach der Rechtsprechung des BGH ist es unzulässig, das Werk oder Werkteile lediglich zur Ausschmückung aufzunehmen: Dies ist beispielsweise der Fall, wenn Abbildungen der Werke eines Künstlers in einem Buch über dessen Leben ohne erkennbare Verbindung (ohne Bezugnahme, Erläuterungen, Verweise) des Textes enthalten sind und diese damit rein als Illustration dienen.⁶⁰ Der BGH setzt daher zum Beispiel eine innere Verbindung derjenigen Textteile, die die Textanalyse nachvollziehbar machen sollen, mit eigenen Ausführungen oder Gedanken des Zitierenden voraus.⁶¹ Auch ist es unzulässig, Zitate als Ersatz für eigene Ausführungen anzuführen.⁶² Schließlich ist auch die Darstellung eines Werks um seiner selbst Willen ohne die vorausgesetzte innere Bezugnahme nicht zulässig: Beispielsweise geht es bei der Darstellung von Vorschau Bildern in der Trefferliste einer Bildersuchmaschine lediglich um die Darbietung dieser Bilder um ihrer selbst Willen und als Hilfsmittel zum Auffinden von Inhalten im Internet.⁶³ Dementsprechend reicht es auch nicht aus, die Werke lediglich deshalb zu nutzen, um Dritten zu diesen leichter Zugang zu verschaffen, etwa zu Informationszwecken.⁶⁴ Werden daher Textteile bei Textanalysen lediglich zur leichteren Auffindbarkeit eingefügt, genügt dies nicht für die Belegfunktion.⁶⁵ Eine untrennbare Einbindung in den zitierenden Gegenstand ist hingegen nicht nötig. Daher genügen beispielsweise Verlinkungen auf zitierte Werke (im PDF-

⁵⁷ BGH GRUR 2020, 859, Rn. 83 – Reformistischer Aufbruch II unter Verweis auf EuGH C-516/17, ECLI:EU:C:2019:625, Rn. 78 – Spiegel Online/Volker Beck.

⁵⁸ BGH GRUR 2020, 859, Rn. 83 – Reformistischer Aufbruch II.

⁵⁹ Daher sei die Nutzung des zitierten Werkes gegenüber den Aussagen des Zitierenden akzessorischer (abhängiger) Natur: BGH GRUR 2020, 859, Rn. 83 – Reformistischer Aufbruch II unter Bezugnahme auf EuGH C-516/17, ECLI:EU:C:2019:625, Rn. 79 – Spiegel Online/Volker Beck.

⁶⁰ BGH NJW 1968, 1875, 1878 – Kandinsky.

⁶¹ Zu dieser Belegfunktion bei Textteilen als Nachweise bei Textanalysen siehe unten III. b) (1).

⁶² *Lüft*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 3.

⁶³ BGH GRUR 2010, 628, Rn. 27 – Vorschau Bilder.

⁶⁴ BGH GRUR 2017, 901 Rn. 25 – Afghanistan Papiere.

⁶⁵ Zur Belegfunktion in diesem Fall allgemein siehe unten III. b) (1).

Format) den Voraussetzungen des § 51 UrhG.⁶⁶ Das ist etwa auf eine Verlinkung von Nachweisen bei der Darstellung der Ergebnisse von Textanalyseverfahren übertragbar. Auch ist es für § 51 UrhG nicht erforderlich, dass sich der Zitierende in erheblichem Umfang mit dem übernommenen Werk auseinandersetzt.⁶⁷ Beispielsweise reichte die Übernahme von Interviewteilen als Beleg für eine These zur interviewten Person in einer Fernsehsendung für die erforderliche innere Verbindung aus, ohne dass eine darüberhinausgehende Auseinandersetzung mit der Interviewten notwendig wäre.⁶⁸

Laut dem EuGH bestehen die wesentlichen Merkmale eines Zitats darin, dass ein Werk oder Auszug aus einem Werk von einem Nutzenden, der nicht dessen Urheber oder Urheberin ist, genutzt wird, um Aussagen zu erläutern, eine Meinung zu verteidigen oder eine geistige Auseinandersetzung zwischen dem Werk und dessen Aussagen des Nutzenden zu ermöglichen.⁶⁹ Für die Schranke „zum Zweck der Zitierung“ bei Gemeinschaftsgeschmacksmustern hat der EuGH in seiner Entscheidung Nintendo/BigBen darüber hinaus einen großzügigeren Ansatz als der BGH verfolgt: Bei der autonomen, europäischen Auslegung des Art. 20 Abs. 1 c Geschmacksmuster-VO Nr. 6/2002 zeigt der EuGH, dass die Darstellung eines geschützten Designs in der Internetwerbung für ein kombinierbares Produkt eine „Zitierung“ im Sinne der Verordnung sein kann.⁷⁰ Damit ist hiernach der Zweck der Zitierung weit auszulegen und kann – sofern dies dem Zweck der Verordnung entspricht – hierbei auch im Sinne von Illustration verstanden werden.⁷¹ Ob diese weite Auslegung des EuGHs auf das urheberrechtliche Zitatrecht in § 51 UrhG übertragen werden kann, ist offen. Die Auslegung des Zitatziels befindet sich noch in der Entwicklung. Der BGH hat bereits für das deutsche DesignG (heute: § 40 Nr. 3 DesignG) umgekehrt auf die Auslegung des Zitatziels in § 51 UrhG zurückgegriffen und die gleichlaufenden Ziele der Vorschriften betont, nämlich die geistige Auseinandersetzung mit fremden Gedanken zu erleichtern.⁷²

⁶⁶ EuGH C-516/17, ECLI:EU:C:2019:625, GRUR 2019, 940 Rn. 80 ff. – Spiegel Online/Volker Beck; BGH GRUR 2017, 1027 Rn. 55 ff. – Reformistischer Aufbruch.

⁶⁷ BGH GRUR 2016, 368 Rn. 31 – Exklusivinterview.

⁶⁸ BGH GRUR 2016, 368 Rn. 31 – Exklusivinterview.

⁶⁹ EuGH C-516/17, ECLI:EU:C:2019:625, Rn. 78 – Spiegel Online/Volker Beck.

⁷⁰ EuGH C-24/16, C-25/16, ECLI:EU:C:2017:724, GRUR 2017, 1120 Rn. 86 – Nintendo/BigBen

⁷¹ Anmerkung *Kur* zu EuGH C-24/16, C-25/16, ECLI:EU:C:2017:724, GRUR 2017, 1120 – Nintendo/BigBen.

⁷² BGH GRUR 2011, 1117 Rn. 45 – ICE.

3. Zitat bei Textanalysen und bei automatisierter Auswertung

a) Ausgangspunkt des BGH-Urteils „Vorschaubilder“

Eine eigene geistige Auseinandersetzung und damit die erforderliche Belegfunktion wurde zum Beispiel bei der automatisierten Aufnahme eines Vorschaubilds in die Trefferliste bei einer Bildersuchmaschine vom BGH verneint.⁷³ Denn zum einen diene der automatisierte Vorgang der Aufnahme der Bilder in die Trefferliste selbst nicht der geistigen Auseinandersetzung mit den übernommenen Werken.⁷⁴ Zum anderen sei die Trefferliste lediglich Hilfsmittel zum Auffinden der Inhalte im Internet und erschöpfe sich im reinen Nachweis der Bilder.⁷⁵ Dass diese die spätere geistige Auseinandersetzung der Nutzenden der Bildersuchmaschine ermöglichen, ändere hieran nichts.⁷⁶ Diese beiden Gesichtspunkte sind beispielsweise auch bei automatisierter Erstellung von Textanalyseergebnissen zu betrachten und zu diskutieren, denen Nachweise angefügt werden. Insbesondere stellt sich auch die Frage, ob dabei eine ausreichende geistige Auseinandersetzung mit den Nachweisen stattfindet oder ob diese lediglich der geistigen Auseinandersetzung der die Analyseergebnisse später Nutzenden dienen. Wichtig ist jedoch in diesem Kontext, dass der BGH in seinem Urteil einen sehr einfachen Fall einer automatisierten Auswertung behandelt hat. Denn der automatisierte Vorgang betraf lediglich die einfache Hilfestellung für Nutzende, dass ein Bild existiert und an einer gewissen Stelle auffindbar ist. Bei einer automatisierten Textanalyse kann jedoch in der automatisierten Auswertung des Textes eine Aussage derjenigen Person enthalten sein, die den Algorithmus einsetzt. Dieser Hintergrund ist bei der Frage danach in den Blick zu nehmen, ob in einer automatisierten Auswertung eine geistige Auseinandersetzung liegen kann (siehe unten).

b) Automatisierte Auswertungen bei Textanalysen – Belegfunktion und geistige Auseinandersetzung

Die oben genannten Beispiele⁷⁷ für Belege bei Textanalysen – im Rahmen der Named Entity Recognition und Statement Extraction – sind beide im We-

⁷³ BGH MMR 2010, 475 Rn. 27 – Vorschaubilder.

⁷⁴ BGH MMR 2010, 475 Rn. 27 – Vorschaubilder.

⁷⁵ BGH MMR 2010, 475 Rn. 27 – Vorschaubilder.

⁷⁶ Anmerkung *Schack*, OLG Jena MMR 2008, 408 Rn 5.

⁷⁷ Siehe oben I.

sentlichen dadurch gekennzeichnet, dass Satz(-teile) in jeweils unterschiedlichem Umfang als Nachweise angefügt werden, die in einer (semi-)automatisierten Textanalyse zugrunde gelegt wurden. Bei der Named Entity Recognition können einzelne Entitäten wie Autoren oder Autorinnennamen, deren analysierte Ausgangsbasis mehrdeutig ausgelegt werden können, durch Belege nachvollziehbar gemacht werden.⁷⁸ Hier können bereits ein Wort oder der direkte Kontext von zwei bis drei Wörtern vor und nach der Entität oder auch ein Satzteil als Beleg genügen. Im Fall der Statement Extraction, dem Auslesen von Themaussagen, ist der zugrundeliegende Satz als Nachweis typischerweise der erforderliche Mindestumfang für eine Plausibilitätsprüfung. In beiden Fällen ist im Zitatrecht zum einen die Voraussetzung der Belegfunktion und zum anderen das Erfordernis einer geistigen Auseinandersetzung des Zitierenden in den Blick zu nehmen.

aa) Belegfunktion bei Nachweisen im Rahmen von Textanalysen

Als Voraussetzung der Belegfunktion (zum Zitatzweck s.o.) ist im Fall von Nachweisen bei Textanalysen fraglich, ob eine innere Verbindung mit eigenen Ausführungen des Zitierenden besteht. Der Vorgang des Text und Data Mining als solcher erfüllt nicht die Voraussetzungen des Zitatrechts, denn grundsätzlich muss ein Zitat der eigenen Gedankenführung des Zitierenden dienen.⁷⁹ Die systematische Auswertung und das Erzielen neuer Erkenntnisse beim Text und Data Mining dient nicht der Unterstützung eines neuen Gedankens, sondern die ausgelesenen Texte sind im ersten Schritt wie reines Recherchematerial zu charakterisieren.⁸⁰ Werden diese ausgewerteten Textteile jedoch in einem zweiten Schritt an Analyseergebnisse als Nachweise angefügt, so stellt sich die Frage, ob dies der Belegfunktion im Zitatrecht genügen kann.

Hier kommt es maßgeblich darauf an, in welcher Weise die Nachweise eingesetzt werden. Die Belegfunktion bedeutet nach dem BGH, dass das Zitat eine Erörterungsgrundlage für eigene Ausführungen des Zitierenden bilden müsste, die die geistige Auseinandersetzung erleichtert.⁸¹ Werden also Textteile, auf deren Basis die Analyseergebnisse erzielt wurden, als Erörterungsgrundlage und damit als Beleg für eigene Aussagen des Zitierenden genutzt, so werden sie grundsätzlich als Hilfsmittel für dessen geistige Auseinandersetzung eingesetzt.

⁷⁸ Siehe oben I.

⁷⁹ *Spindler*, GRUR 2016, 1112, 1116.

⁸⁰ *Spindler*, GRUR 2016, 1112, 1116.

⁸¹ BGH MMR 2010, 475 Rn. 26 – Vorschaubilder; BGH GRUR 1986 59, 60 – Geistchristentum; *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 31.

Werden die Textteile hingegen nur um ihrer selbst Willen dargestellt und zusammenhangslos angehängt, wäre die Belegfunktion eher zu verneinen.⁸² Außerdem fehlt nach dem BGH die Belegfunktion, wenn der Nachweis lediglich als Illustration genutzt wird, weil es an einer inneren Verbindung zwischen dem fremden Werk und den eigenen Ausführungen des Zitierenden fehlt.⁸³ Nach dem BGH fehlt es an einer solchen inneren Verbindung, wenn „sich das zitierende Werk nicht näher mit dem eingefügten fremden Werk auseinandersetzt, sondern es nur zur Illustration verwendet, es in einer bloß äußerlichen, zusammenhanglosen Weise einfügt oder anhängt oder das Zitat ausschließlich eine informierende Berichterstattung bezweckt“.⁸⁴ Wenn also durch den Textabschnitt ein Analyseergebnis belegt werden soll, wird man eine solche innere Verbindung im Regelfall annehmen können. Der Textabschnitt darf aber nicht in erster Linie deswegen dargestellt werden, damit er in erster Linie als solcher und nicht in seiner Eigenschaft als Beleg zur Kenntnis genommen wird.

bb) Geistige Auseinandersetzung bei Nachweisen im Rahmen von Textanalysen

Zusätzlich muss in beiden Fällen der Belege bei Textanalysen die Frage nach dem Vorliegen einer geistigen Auseinandersetzung⁸⁵ des Zitierenden gestellt werden. Fraglich ist dabei insbesondere, ob bei einer rein automatisierten Analyse, an die Belege angefügt werden, eine geistige Auseinandersetzung vorliegen kann. Der BGH hat, wie bereits oben geschildert, im Fall der Google-Bildersuche bei der automatisierten Aufnahme eines Vorschaubilds in die Trefferliste bei einer Bildersuchmaschine das Kriterium der geistigen Auseinandersetzung verneint.⁸⁶ Jedoch handelt es sich in diesem Fall wie bereits geschildert um einen sehr einfachen Fall der automatisierten Auswertung: lediglich die Existenz eines Bildes und dessen Auffindbarkeit wurde automatisiert bereitgestellt. In den oben genannten Beispielen bei Textanalysen in den Digital Humanities könnte der Einsatz von Algorithmen jedoch eine darüberhinausgehende Aussage enthalten. Beispielsweise wäre bei der Named Entity Recognition die Aussage, dass

⁸² BGH GRUR 2012, 819 Rn. 28 – Blühende Landschaften; *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 3a.

⁸³ BGH GRUR 2016, 368 Rn. 25 – Exklusivinterview.

⁸⁴ BGH GRUR 2016, 368 Rn. 25 – Exklusivinterview.

⁸⁵ BGH MMR 2010, 475 Rn. 26 – Vorschaubilder; BGH GRUR 1986 59, 60 – Geistchristentum.

⁸⁶ BGH MMR 2010, 475 Rn. 27 – Vorschaubilder.

eine Entität einem Autoren- oder Autorinnennamen entspricht (etwa Bezeichnung = Autor:innenname). Bei der Statement Extraction könnte beispielsweise eine extrahierte Themaussage (etwa Werk handelt von X) eine – über eine einfache Aussage wie die reine Auffindbarkeit von Bildern im BGH-Urteil – hinausgehende Aussage dessen sein, der den Algorithmus auf den Texten einsetzt. Zusätzlich kann insbesondere bei der Statement Extraction, dem Auslesen von Themaussagen, bei der Optimierung des Algorithmus eine manuelle Überprüfung und Kontrolle der Zwischenergebnisse erfolgen. Diese Zwischenschritte sorgen dafür, dass hier also nicht ausschließlich von einem einzigen, rein automatisierten Analysevorgang ausgegangen werden könnte. Für eine geistige Auseinandersetzung bei automatisierten Textanalysen könnte noch die Tatsache sprechen, dass die geistige Auseinandersetzung in diesen Fällen vorverlagert stattfindet: Die automatisierte Auswertung beruht nämlich auf zuvor ausgestalteten Verfahren und Konzepten, die mit dem Einsatz des Algorithmus einhergehen und eine geistige Befassung u.a. auch mit den zugrundeliegenden Texten erfordern. Man könnte also eine mittelbare geistige Auseinandersetzung derjenigen Person diskutieren, die den Algorithmus auf den Texten einsetzt. Der Einsatz solcher Textanalysen in der Wissenschaft entspricht schließlich auch dem Hintergrund des Zitatrechts, das Zitieren im Sinne des Allgemeininteresses an freier geistiger Auseinandersetzung gerade auch im Bereich der Wissenschaft zu privilegieren.⁸⁷ Der nach dem EuGH nunmehr nicht mehr vorausgesetzte Werkcharakter des zitierenden Gegenstands (s.o.) könnte außerhalb des Kriteriums der geistigen Auseinandersetzung eine solche Berücksichtigung automatisierter Analysen im Zitatrecht grundsätzlich möglich machen.

Im Ergebnis gilt jedoch: Für Belege bei automatisierten Textanalysen existiert keine höchstrichterliche Rechtsprechung. Ob eine geistige Auseinandersetzung bei automatisierten Textanalysen vorliegen kann, ist damit im Ergebnis (noch) ungeklärt. In der rechtswissenschaftlichen Literatur wird eine geistige Auseinandersetzung bei dem Text und Data Mining teilweise abgelehnt.⁸⁸ Zusätzlich gilt es zu bedenken: Wenn die Analyseergebnisse beispielsweise lediglich dargestellt und die Nachweise nur für die später Nutzenden eingefügt werden, könnte in Anlehnung an die Rechtsprechung des BGH auch argumentiert werden, dass

⁸⁷ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 1.

⁸⁸ Eine Auseinandersetzung mit dem Werk bei Text und Data Mining verneinend: *Specht*, *OdW* 2018, 285, 286.

diese möglicherweise eher deren geistigen Auseinandersetzung als der des Zitierenden dienen (s.o.).⁸⁹

4. Erkennbarkeit und Umfang des Zitats

Es ist in jedem Fall Wesen eines Zitats, dass es als fremde Zutat ersichtlich ist und nicht ununterscheidbar in den Text eingebaut wird.⁹⁰ Beispielsweise ist bereits fraglich, ob allein die Kursivsetzung einer übernommenen, fremden Textpassage aus den Werken von Bertolt Brecht für diese erkennbare Abhebung vom eigenen Werk ausreicht.⁹¹ Textteile, die als Belege angefügt werden sollen, müssten demnach als fremde Textteile kenntlich gemacht werden.⁹² Darüber hinaus macht das Weglassen der Quellenangabe das Zitat insgesamt nach § 63 Abs. 1 UrhG unzulässig, sofern die Quellenangabe nicht unmöglich ist.⁹³

Der Zweck des Zitats bestimmt nach § 51 S. 1 UrhG dessen zulässigen Umfang. Dieser vom Zweck gebotene Umfang erschöpft sich nicht im Minimalumfang, sondern es ist ein in der Gesamtheit vernünftiger und sachgerechter Umfang zulässig.⁹⁴ Es muss der für den Zitatzzweck konkret erforderliche qualitative und quantitative Umfang vorliegen.⁹⁵ Dies ergibt sich durch Abwägung der Umstände des Einzelfalls und dabei insbesondere des Zitatzzwecks, der Besonderheiten sowie des Umfangs des zitierenden wie zitierten Werkes sowie der Art der Zugänglichmachung.⁹⁶ Beispielsweise wurde das Abdrucken einer von drei Strophen eines Verkehrskinderliedes als vom Zweck des zitierenden Artikels umfasst angesehen.⁹⁷ Eine von drei Strophen dieses Liedes in einem Pressebericht über Verkehrserziehung abzudrucken, hat der BGH als notwendigen Umfang des Zitats für einen kurzlebigen Pressebericht eingestuft.⁹⁸ Das Lied mit den enthalte-

⁸⁹ BGH MMR 2010, 475 Rn. 27 – Vorschaubilder; siehe auch Anmerkung *Schack*, OLG Jena MMR 2008, 408 Rn 5.

⁹⁰ *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 31.

⁹¹ OLG München NJW 1999, 1975, 1976 – Stimme Brecht.

⁹² Beachte aber die Belegfunktion siehe oben.

⁹³ Siehe Art. 5 Abs. 3 lit. d InfoSoc-RL (2001/29/EG); *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 31.

⁹⁴ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 5.

⁹⁵ *Schulz*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn. 14.

⁹⁶ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 5; siehe auch *Schulz*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn. 14.

⁹⁷ BGH NJW 1959, 336, 338 – Verkehrs-Kinderlied.

⁹⁸ BGH NJW 1959, 336, 338 – Verkehrs-Kinderlied.

nen Verkehrsregeln sei darauf angelegt, von Kindern in seiner Gesamtheit gelernt zu werden.⁹⁹ Der Abdruck einer Strophe dieses Liedes greife daher nicht unzumutbar in die Verwertung des Liedes ein.¹⁰⁰ Werden Textteile als Belege von Textanalysen eingefügt, ist daher auf den zulässigen Umfang zu achten, der durch den Zitatzweck begrenzt wird. Ist der Umfang des Zitats vom Zitatzweck nicht mehr gedeckt, so ist nicht nur der überschießende Teil, sondern das Zitat in seiner Gesamtheit unzulässig.¹⁰¹

Der Umfang des Zitats darf zudem nicht die Verwertung des zitierten Werkes unzumutbar beeinträchtigen.¹⁰² Es darf insbesondere nicht in einem solchen Umfang Kenntnis vom Original oder dessen Kernstücken verschaffen, dass es den Erwerb des zitierten Werks substituieren kann.¹⁰³ Wie bereits geschildert wurde beispielsweise die Zitierung der ersten von drei Strophen eines Verkehrskinderliedes allerdings nicht als Beeinträchtigung der Verwertung des gesamten Liedes gesehen.¹⁰⁴

5. Selbstständigkeit des zitierenden Textes

Weiterhin muss der zitierende Text nach § 51 S. 2 Nr. 1 - 3 UrhG Selbstständigkeit aufweisen. Ursprünglich setzte dies nach deutschem Verständnis Urheberrechtsschutzfähigkeit¹⁰⁵ des zitierenden Textes voraus.¹⁰⁶ Teilweise wird dies in der rechtswissenschaftlichen Literatur weiterhin so vertreten.¹⁰⁷ Nach der Rechtsprechung des EuGH muss es sich jedoch nunmehr – in richtlinienkonformer Auslegung mit Blick auf Art. 5 Abs. 3 d) InfoSoc-RL (2001/29/EG) –

⁹⁹ BGH NJW 1959, 336, 338 – Verkehrs-Kinderlied.

¹⁰⁰ BGH NJW 1959, 336, 338 – Verkehrs-Kinderlied.

¹⁰¹ BGH ZUM 2012, 681 Rn. 29 – Blühende Landschaften.

¹⁰² *Lüft*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 7.

¹⁰³ BGH GRUR 1986, 59, 60 – Geistchristentum.

¹⁰⁴ BGH NJW 1959, 336, 338 – Verkehrs-Kinderlied.

¹⁰⁵ Zu den Voraussetzungen der urheberrechtlichen Schutzfähigkeit siehe *Erler-Fridgen*, Teil 2, A. in diesem Band; zum Urheberrechtsschutz von Datenbankwerken siehe *Erler-Fridgen*, Teil 2, D. in diesem Band.

¹⁰⁶ BGH GRUR 1994, 800, 802 – Museumskatalog; siehe *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 6 mwN.

¹⁰⁷ *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 545 mwN, es müsse Werkcharakter aufweisen: *Schulz*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn. 10 f.

nicht um ein urheberrechtlich geschütztes Werk nach § 2 Abs. 2 UrhG handeln.¹⁰⁸ Der EuGH hat für Art. 5 Abs. 3 d) InfoSoc-RL (2001/29/EG) entschieden, dass der zitierende Gegenstand, etwa ein Presseartikel, kein urheberrechtlich geschütztes Werk sein muss.¹⁰⁹ Diese richtlinienkonforme Auslegung wird durch die Gesetzesbegründung zu § 51 S. 3 UrhG bestätigt:¹¹⁰ Dies lasse sich auf die Generalklausel des § 51 S. 1 UrhG stützen, wonach das Zitat auch in einem selbst nicht schutzfähigen Werk erfolgen kann.¹¹¹ Beispielsweise könnte demnach eine automatisiert erstellte Trefferliste in einer Bildersuchmaschine ohne die zwingende Voraussetzung eines Werkcharakters als zitierendes Werk charakterisiert werden.¹¹² Daher kann auch eine digitale Bündelung von automatisiert erstellter Textanalyseergebnisse ein hinreichend selbstständiger zitierender Gegenstand sein.

Selbstständig ist der zitierende Gegenstand dann, wenn er urheberrechtlich vom zitierten Werk unabhängig ist, insbesondere darf keine Bearbeitung oder Umgestaltung nach § 23 UrhG vorliegen.¹¹³ Vor dem oben genannten anderslautenden Urteil des EuGH¹¹⁴ wurde die Selbstständigkeit abgesprochen, wenn ein-

¹⁰⁸ EuGH C-145/10, ECLI:EU:C:2011:798, GRUR 2012, 166 Rn. 130 ff. – Painer/Standard; *Lüft*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 8; *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 6; eine zwingende Voraussetzung des Werkcharakters verneinend: OLG Jena GRUR-RR 2008, 223, 225 – Thumbnails; offengelassen von BGH GRUR 2010, 628 Rn. 25 – Vorschaubilder; a.A. *Schulz*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn. 10 f.; zulässige Zitatfälle, in denen das zitierende Werk selbstständig, aber nicht urheberrechtsschutzfähig ist dürfen die Ausnahme bleiben: *Dustmann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 21.

¹⁰⁹ EuGH C-145/10, ECLI:EU:C:2011:798, GRUR 2012, 166 Rn. 130 ff. – Painer/Standard.

¹¹⁰ *Lüft*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 8.

¹¹¹ Entwurf eines Gesetzes zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft, BT-Drs. 18/12329, 32.

¹¹² Schon vor der Rechtssprechung des EuGH C-145/10, ECLI:EU:C:2013:138, GRUR 2012, 166 Rn. 130 ff. – Painer/Standard hat das OLG Jena MMR 2008, 408, 410 die Voraussetzung des Werkcharakters des zitierenden Werks im damals neugefassten § 51 UrhG als nicht zwingend angesehen.

¹¹³ *Lüft*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 8; nach § 62 UrhG gilt das allgemeine Änderungsverbot auch bei den urheberrechtlichen Schranken wie § 51 UrhG, also das Verbot der Abweichung etwa von der Form und des Gesamteindrucks der Veröffentlichung, die durch den Urheber bestimmt wurden: *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 62 Rn. 5 ff.

¹¹⁴ EuGH C-145/10, ECLI:EU:C:2013:138, GRUR 2012, 166 Rn. 130 ff. – Painer/Standard.

zelne Zitate ohne besondere eigene geistige Leistung zusammengestellt wurden.¹¹⁵ Beispielsweise wurde deshalb eine Sammlung von Zitaten nicht von § 51 UrhG erfasst, bei der sich die Leistung des Herausgebers auf die Zusammenstellung und Gliederung der Zitate beschränkte.¹¹⁶ Jedoch kann dies nach dem Urteil des EuGH¹¹⁷ nicht mehr vorausgesetzt werden, insbesondere muss die eigene geistige Leistung, nicht mehr die fremde Leistung, überwiegen.¹¹⁸ Es reicht mithin aus, wenn der Zitatzweck durch irgendwelche eigenen Ausführungen auch ohne schöpferische Leistung begründet wird (s.o.).¹¹⁹

III. Das wissenschaftliche Großzitat (§ 51 S. 2 Nr. 1 UrhG)

§ 51 S. 2 Nr. 1 UrhG umfasst das wissenschaftliche Großzitat und privilegiert die Aufnahme einzelner ganzer Werke in ein selbstständiges wissenschaftliches Werk.¹²⁰ Gekennzeichnet ist das Großzitat dadurch, dass nach Nr. 1 ganze Werke und nicht nur Werkteile (wie in Nr. 2 – 3) zitiert werden dürfen und diese sich auch nicht auf einen geringen Umfang beschränken müssen.¹²¹ Jedoch schränkt der Zitatzweck „zur Erläuterung des Inhalts“ (des zitierenden Werkes) den zulässigen Umfang des Zitats ein.¹²² Werden zum Beispiel lediglich Teile einer Abhandlung in einem Werk besprochen, so ist hiernach auch nur das Zitat dieser Teile zulässig.¹²³ Zusätzlich darf die Verwertung des zitierten Werkes nicht durch das Zitat unzumutbar behindert werden.¹²⁴

Das zitierende wissenschaftliche Werk ist nach dem Zweck der Vorschrift, der Förderung der kulturellen Entwicklung im Allgemeinen, weit auszulegen und aus diesem Grund sind auch populärwissenschaftliche Werke hierunter zu fassen.¹²⁵ Grundsätzlich bestimmt sich der wissenschaftliche Charakter nach dem

¹¹⁵ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 7.

¹¹⁶ BGH GRUR 1973, 216, 218 – Handbuch moderner Zitate; Lüft, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 8.

¹¹⁷ EuGH C-145/10, ECLI:EU:C:2013:138, GRUR 2012, 166 Rn. 130 ff. – Painer/Standard.

¹¹⁸ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 7.

¹¹⁹ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 7.

¹²⁰ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 8.

¹²¹ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 10.

¹²² Lüft, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 10.

¹²³ Spindler, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 73.

¹²⁴ Spindler, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 73.

¹²⁵ Spindler, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 64.

Werkinhalt.¹²⁶ Damit ist ein Werk als wissenschaftlich anzusehen, das nach Rahmen, Form und Gehalt durch eine eigene Geistestätigkeit die Wissenschaft durch Vermittlung von Erkenntnis fördern will und der Belehrung dient.¹²⁷ Eine Sammlung oder Erörterung von Ergebnissen aus Textanalysen, die mit wissenschaftlichen Methoden erstellt wurde und die der wissenschaftlichen Auseinandersetzung mit den Daten dient, kann beispielsweise einen wissenschaftlichen Charakter aufweisen.¹²⁸ Zentral ist dabei die ernsthafte, methodisch geordnete Suche nach Erkenntnis.¹²⁹ Werke ohne wissenschaftliche Methode oder ohne Ansprache des Intellekts, sondern mit Ansprache anderer Persönlichkeitsaspekte wie Gefühlen oder Schönheitssinn – etwa Werke der Belletristik, politische Darstellungen oder Werbung – scheiden als wissenschaftliche Werke wohl aus.¹³⁰ Bei einem Überschneiden von wissenschaftlichen Informationen und unterhaltsamer Darstellung ist entscheidend, ob die wissenschaftliche Auseinandersetzung den Unterhaltungszweck überwiegt.¹³¹ Die Werkgattung ist bei der Einordnung als wissenschaftliches Werk nicht wesentlich: Das zitierende wissenschaftliche Werk kann beispielsweise ein Schriftwerk (§ 2 Abs. 1 Nr. 1 UrhG), ein Filmwerk (§ 2 Abs. 1 Nr. 6 UrhG) sowie eine Darstellung wissenschaftlicher Art (§ 2 Abs. 1 Nr. 7 UrhG) sein.¹³² Dies gilt hingegen wohl nicht für einen Gegenstand, bei dem es regelmäßig an der wissenschaftlichen Auseinandersetzung fehlt, weil er etwa rein funktional-technisch ausgerichtet ist wie ein Computerprogramm, oder auch bei Werken aus dem Bereich der Kunst (§ 2 Abs. 1 Nr. 2-4 UrhG).¹³³

Zulässig ist nach § 51 S. 2 Nr. 1 UrhG das Großzitat einzelner Werke. Wie viele Werke diese „einzelnen Werke“ sein können, bestimmt sich nach dem Gesamtwerk des Zitierten (der Anzahl dessen Werke) und nach der Art und dem Umfang des zitierenden wissenschaftlichen Werks.¹³⁴ Je länger damit das zitierende

¹²⁶ *Schulz*, in Ahlberg/Götting, Möhring/Nicolini, Urheberrecht, 4. Aufl. 2018, § 51 Rn. 16.

¹²⁷ LG München ZUM 1989, 529, 530 unter Verweis auf LG Berlin GRUR 1962, 207 – Maifeiern und RGZ 130, 196 – Codex aureus.

¹²⁸ Zum Kriterium der Selbstständigkeit siehe oben I.

¹²⁹ *Schulz*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn. 16; *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 8.

¹³⁰ *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 67; *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 8.

¹³¹ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 8.

¹³² *Dustmann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 26.

¹³³ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 9.

¹³⁴ *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 69.

Werk ist, desto mehr Werke dürfen typischerweise als Zitat aufgenommen werden.¹³⁵ Dabei ist umstritten, ob die Anzahl der „einzelnen“ Werke absolut auf einige wenige Zitate zu beschränken ist¹³⁶ oder ob – wie teilweise in der rechtswissenschaftlichen Literatur vertreten – eine differenzierte Betrachtung angebracht sei,¹³⁷ die je nach Menge der zitierten Urheber auch eine Mehrzahl an Zitaten erlaubt.¹³⁸ Beispielsweise wurde in absoluter Weise die Entlehnung von 69 Werken Kandinskys in einem Buch über den „Blauen Reiter“ vom BGH als diese Grenze von „einzelnen Werken“ überschreitend angesehen.¹³⁹ Ebenso wurde die Übernahme von 24 Comiczeichnungen eines Künstlers in absoluter Hinsicht als hohe Anzahl qualifiziert und nicht in ein Verhältnis zum gesamten Schaffen des Künstlers gestellt.¹⁴⁰

Zulässig ist ein wissenschaftliches Großzitat nach der Veröffentlichung im Sinne von § 6 Abs. 1 UrhG.¹⁴¹ Da hiermit die alte Rechtslage vor 2008 abgelöst wurde, nach der das zitierte Werk nach § 6 Abs. 2 UrhG zuvor erschienen¹⁴² sein musste, sind nunmehr wissenschaftliche Großzitate aus veröffentlichten¹⁴³, aber (noch) nicht erschienenen¹⁴⁴ Werken möglich.¹⁴⁵

Wesentlich ist, dass der Zitatzweck für das wissenschaftliche Großzitat enger ist als der in den Fällen von § 51 S. 2 Nr. 2 oder Nr. 3 UrhG:¹⁴⁶ Das wissenschaftliche Großzitat ist nämlich nur „zur Erläuterung des Inhalts des zitierenden Werkes“ zulässig. Das bedeutet, dass das Zitat den Inhalt des zitierenden Werkes – über die reine Belegfunktion (s.o.) hinaus – erläutern muss.¹⁴⁷ Es wird also der

¹³⁵ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 11.

¹³⁶ BGH GRUR 1968, 607, 611 – Kandinsky I; OLG München ZUM 1989, 529; Lüft, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 10.

¹³⁷ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 11; Spindler, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 69.

¹³⁸ Spindler, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 69 mwN.

¹³⁹ BGH GRUR 1968, 607, 611 – Kandinsky I.

¹⁴⁰ KG ZUM-RD 1997, 135, 137.

¹⁴¹ Nach § 6 Abs. 2 UrhG ist ein Werk veröffentlicht, wenn es mit Zustimmung des Berechtigten der Öffentlichkeit zugänglich gemacht worden ist.

¹⁴² Nach § 6 Abs. 2 UrhG ist ein Werk erschienen, wenn mit Zustimmung des Berechtigten Vervielfältigungsstücke des Werkes nach ihrer Herstellung in genügender Anzahl der Öffentlichkeit angeboten oder in Verkehr gebracht worden sind.

¹⁴³ Veröffentlicht ist ein Werk nach § 6 Abs. 1 UrhG, wenn es mit Zustimmung des Berechtigten der Öffentlichkeit zugänglich gemacht worden ist.

¹⁴⁴ Erschienen ist ein Werk nach § 6 Abs. 2 S. 1 UrhG, wenn mit Zustimmung des Berechtigten Vervielfältigungsstücke des Werkes nach ihrer Herstellung in *genügender Anzahl* der Öffentlichkeit angeboten oder in Verkehr gebracht worden sind.

¹⁴⁵ Spindler, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 74; Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 12.

¹⁴⁶ Spindler, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 75.

¹⁴⁷ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 13.

eigene Gedankengang durch das Zitat gestützt oder in Bezug genommen zur weiteren Auseinandersetzung oder zur Kombination von Aussagen.¹⁴⁸ Beispielsweise kann ein solcher Erläuterungszweck beim Anführen von Textpassagen aus geschichtlichen Werken oder Photographien in Ausstellungskatalogen vorliegen, die als Bildzitat den wissenschaftlichen Begleittext erläutern.¹⁴⁹ Bei Textanalysen könnte etwa die Analyse von ganzen Gedichten oder Haikus ein Beispiel für den Nachweis eines gesamten Werkes bei der Angabe von Analyseergebnissen sein. In diesem Fall müsste für den beim Großzitat vorausgesetzten Zitatzweck eine geistig-kritische Auseinandersetzung mit Erörterungscharakter vorliegen. Wie bereits oben geschildert, kann die Belegfunktion für den Zitatzweck vorliegen, wenn die Textteile, auf deren Basis die Analyseergebnisse erzielt wurden, als Erörterungsgrundlage und damit als Beleg für eigene Aussagen des Zitierenden genutzt werden und damit grundsätzlich als Hilfsmittel für dessen geistige Auseinandersetzung eingesetzt werden.¹⁵⁰ Beim Großzitat müssen sie zusätzlich „zur Erläuterung des Inhalts des zitierenden Werks“ eingesetzt werden (s.o.). Wie allgemein zum Zitatzweck beschrieben, kann im Regelfall die vom BGH vorausgesetzte innere Verbindung zwischen dem fremden und zitierenden Werk angenommen werden, wenn durch den Textausschnitt ein Analyseergebnis belegt werden soll.¹⁵¹ Der Nachweis muss jedoch in seiner Eigenschaft als Beleg eingesetzt werden und nicht rein um die Darstellung seiner selbst.¹⁵² Im Regelfall dürfte bei dem Einsatz von Textteilen als Beleg für Textanalyseergebnisse auch der für das Großzitat zusätzlich erforderliche Erläuterungscharakter des Belegs wohl gegeben sein, denn die Analyseergebnisse sollen als Erläuterung für enthaltene Zweifelsfälle nachvollziehbar belegt werden. Problematisch ist aber die erforderliche geistige Auseinandersetzung bei automatisierten Textanalyseverfahren, die wie bereits oben geschildert, noch ungeklärt ist.¹⁵³

¹⁴⁸ *Schulz*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn 17.

¹⁴⁹ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 13; zum entsprechenden Bildzitat, das im Mittelpunkt eines Textes steht, der ohne das Zitat nicht verständlich wäre: BGH GRUR 1994, 800, 802 – Museumskatalog.

¹⁵⁰ Siehe dazu oben III. b) (1).

¹⁵¹ Siehe dazu oben III. b) (1).

¹⁵² Siehe dazu oben III. b) (1).

¹⁵³ Siehe dazu oben III. b) (2).

IV. Das Kleinzitat (§ 51 S. 2 Nr. 2 UrhG)

Wenn nur Stellen eines Werkes zitiert werden, dann ist dies nach § 51 S. 2 Nr. 2 UrhG auch zu anderen als wissenschaftlichen Zwecken zulässig.¹⁵⁴ Stellen eines Werkes sind kleine Ausschnitte, deren Umfang weder absolut, noch relativ im Verhältnis zum gesamten genutzten Werk, ins Gewicht fällt.¹⁵⁵ Dabei ist also das Verhältnis der Länge des Zitats zum Umfang des zitierten Werkes (relativ) als auch die Länge des Zitats als solches (absolut) wesentlich.¹⁵⁶ Typischerweise sollte das Kleinzitat lediglich ein Bruchstück des gesamten zitierten Werkes ausmachen.¹⁵⁷ Das bedeutet, dass bei größeren zitierten Werken mehr, bei kleineren zitierten Werken weniger zitiert werden kann.¹⁵⁸ Zusätzlich aber soll ein gewisser absoluter Umfang des Zitats nicht überschritten werden.¹⁵⁹ Die Anwendung arithmetischer Maßstäbe zur Bemessung des zulässigen Zitatumfangs wird jedoch vom BGH abgelehnt, denn eine schöpferische geistige Tätigkeit lasse sich nie nach starren rechnerischen Maßstäben ohne Rücksicht auf den Einzelfall abmessen.¹⁶⁰ Es gibt daher keine regelhafte Länge eines zulässigen Kleinzitats: Weder ist nur das Zitat von ein oder zwei Kernsätzen zulässig,¹⁶¹ noch kann als Regel herangezogen werden, dass eine Seite nicht überschritten werden darf.¹⁶² Vielmehr wird die zulässige Länge eines Zitats vom Zitatzweck bestimmt.¹⁶³ Es ist jeweils nach den Besonderheiten des Einzelfalls und zitierten Werks zu fragen.¹⁶⁴ Der Zitatzweck kann es dabei ausnahmsweise erfordern, dass auch längere Abschnitte übernommen werden.¹⁶⁵ Beispielsweise wurde bei einem Liedtext, der im Original elf Zeilen umfasst, das Zitat von vier Zeilen als zulässig angesehen, da es der Zweck des Zitats war, die Wortwahl und Atmosphäre des entsprechenden Liedes darzustellen, u.a. um den Erfolg des Liedschöpfers bei einer

¹⁵⁴ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 14.

¹⁵⁵ Dustmann, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 28; Spindler, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 81.

¹⁵⁶ BGH GRUR 1959, 197, 199 – Verkehrskinderlied; Spindler, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 83; Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 14.

¹⁵⁷ BGH GRUR 1959, 197, 200 – Verkehrskinderlied; Spindler, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 83.

¹⁵⁸ BGH GRUR 1986, 59, 60 – Geistchristentum; Spindler, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 83.

¹⁵⁹ Spindler, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 83.

¹⁶⁰ BGH GRUR 1959, 197, 199 – Verkehrskinderlied.

¹⁶¹ BGH GRUR 1986, 59, 60 – Geistchristentum.

¹⁶² Spindler, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 83 mwN.

¹⁶³ Dustmann, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 29.

¹⁶⁴ BGH GRUR 1986, 59, 60 – Geistchristentum.

¹⁶⁵ Dustmann, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 29.

bestimmten Zielgruppe zu erörtern.¹⁶⁶ Dieser Zweck sei nach Ansicht des Gerichts nicht durch Angabe einer oder zwei Zeilen des Liedes erfüllbar.¹⁶⁷ Für die Abgrenzung bei längeren Abschnitten ist auf den Grundgedanken des Gesetzes und den Interessenskonflikt zurückzugreifen, der § 51 UrhG zugrunde liegt: Um der Freiheit der geistigen Auseinandersetzung mit fremden Gedanken zu fördern, wird dem Schöpfer eines Werkes dann ein verhältnismäßig geringer Eingriff in seine Verwertungsrechte (§ 15 UrhG) zugemutet, wenn das Zitat – auf der Basis des geistigen Schaffens anderer – dem Nutzen der Allgemeinheit dient, der Förderung des kulturellen Lebens.¹⁶⁸

Als Grenze für die Länge eines Zitats wirkt dabei aber auch, dass durch den Umfang des Zitats die Verwertung des zitierten Werkes nicht unzumutbar beeinträchtigt werden darf.¹⁶⁹ Das bedeutet, dass das Zitat nicht in einem solchen Umfang Kenntnis vom Original oder dessen Kernstücken verschaffen darf, dass hierdurch ein Ersatz für den Erwerb des Originals geschaffen wird und damit die Verwertungsmöglichkeiten des Schöpfers geschmälert werden.¹⁷⁰ Zum Beispiel wurde das Zitat einer von drei Strophen eines Kinderlieds, das vor den Gefahren des Verkehrs warnen will, als zulässig eingestuft, da das Lied lediglich als Ganzes seinen erzieherischen Zweck erfüllen könne und damit eine einzelne Strophe nicht die Verwertung des vollständigen Lieds beeinträchtige.¹⁷¹ Beispielsweise könnte dies problematisch werden, wenn im Falle der Sentiment Analyse¹⁷² jeder Satz als Beleg für die Analyseergebnisse ergänzt würde und damit darüber hinaus auch die Satzreihenfolge nachvollziehbar würde.

Der Zitatzweck ist bei dem Kleinzitat nach § 51 S. 2 Nr. 2 UrhG weiter gefasst als beim wissenschaftlichen Großzitat, bei dem sich der umfasste Zweck auf die Erläuterung des Inhalts beschränkt (§ 51 S. 2 Nr. 1 UrhG).¹⁷³ Beispielsweise wurde auch als zulässiges Zitat die Entlehnung in Form einer Devise oder eines Mottos angesehen, die anderen Ausführungen vorangestellt wurden.¹⁷⁴ Auch

¹⁶⁶ OLG Hamburg GRUR 1970, 38, 39 – Heintje.

¹⁶⁷ OLG Hamburg GRUR 1970, 38, 39 – Heintje.

¹⁶⁸ BGH GRUR 1986, 59, 60 – Geistchristentum.

¹⁶⁹ *Spindler*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 83.

¹⁷⁰ BGH GRUR 1959, 197, 200 – Verkehrskinderlied.

¹⁷¹ BGH GRUR 1959, 197, 200 – Verkehrskinderlied.

¹⁷² *Ignatov/Mibalcea*, Text Mining, 148 ff.

¹⁷³ Zum Zitatzweck im Rahmen des wissenschaftlichen Großzitats siehe oben; siehe etwa auch BGH GRUR 1973, 216, 218 – Handbuch moderner Zitate.

¹⁷⁴ *Spindler*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 38; KG GRUR-RR 2002, 313, 315 – Das Leben, dieser Augenblick; oder siehe auch für die Voranstellung als Motto: OLG München ZUM 2009, 970, 971.

ein Zitat zum Beleg einer referierenden¹⁷⁵ Darstellung kann den Zitatzweck des § 51 S. 2 Nr. 2 UrhG erfüllen.¹⁷⁶ Hat jedoch das Zitat lediglich den Zweck, den Text aufzulockern, oder verfolgt es Zwecke der Bewerbung ohne eigene kritische geistige Auseinandersetzung, so ist dies nicht mehr von der Zitierfreiheit nach § 51 S. 2 Nr. 2 UrhG erfasst.¹⁷⁷ Beispielsweise stellt eine Datenbank im Internet, die als Mediathek um einen Film herum aufgebaut ist und neben Filmausschnitten auch weitere Informationen zum Film und Hintergründe und Fakten zum Film und Filmkritiken enthält, keine eigene geistige Auseinandersetzung mit dem Film dar.¹⁷⁸ Eine solche geistige Auseinandersetzung ist beispielsweise auch dann fraglich, wenn Textteile als Nachweise bei Textanalysen genutzt werden. Insbesondere bei automatisiert erstellten Analyseergebnissen ist die geistige Auseinandersetzung zu diskutieren (s.o.).¹⁷⁹

Das Bundesverfassungsgericht räumte Werken der Kunst (Art. 5 Abs. 3 S. 1 GG) einen weiteren Anwendungsbereich des § 51 UrhG ein und erkannte bei dem Kleinzitat nach § 51 S. 2 Nr. 2 UrhG für Kunstwerke auch einen über die Belegfunktion hinausgehenden Zitatzweck an.¹⁸⁰ Als Ausdruck der Kunstfreiheit genüge in diesen Fällen der Einsatz des Zitats als künstlerisches Ausdrucks- und Gestaltungsmittel bei einer inneren Verbindung der zitierten Stellen.¹⁸¹ Ein Auseinandersetzen mit dem fremden Werk wäre nach diesem in seiner dogmatischen Herleitung nicht unumstrittenen Ansatz¹⁸² nicht notwendig, sondern es

¹⁷⁵ Zu einer rein informierenden Berichterstattung: BGH GRUR 2012, 819 Rn. 28 – Blühende Landschaften.

¹⁷⁶ *Spindler*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 34.

¹⁷⁷ *Spindler*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 34.

¹⁷⁸ KG MMR 2003, 110, 111 – Paul und Paula.

¹⁷⁹ Zur Belegfunktion in diesen Fällen siehe oben III. b) (1), zur geistigen Auseinandersetzung siehe III. b) (2).

¹⁸⁰ BVerfG GRUR 2001, 149, 152 – Germania 3.

¹⁸¹ BVerfG GRUR 2001, 149, 152 – Germania 3; siehe auch BGH GRUR 2012, 819 Ls. 1 – Blühende Landschaften; ausführlich zu den Hintergründen siehe *Dustmann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 31; *Spindler*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 38; *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 14; *Schulz*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 51 Rn 13.1; kritisch jedenfalls zur dogmatischen Herleitung dieser Rechtsprechung im Lichte des Schrankensystems des UrhG: *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 537; siehe auch BGH GRUR 2017, 1027 Rn. 31 – Reformistischer Aufbruch; BGH GRUR 2003, 956, 957 – Gies-Adler.

¹⁸² Kritisch zur dogmatischen Herleitung etwa *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 537 unter Hinweis auf die spätere Rechtsprechung des BGH GRUR 2017, 1027 Rn. 31 – Reformistischer Aufbruch; BGH GRUR 2003, 956, 957 – Gies-Adler.

sei allein maßgeblich, ob sich das Zitat funktional in die künstlerische Gestaltung und Intention des Werks einfüge.¹⁸³

Es wird teilweise diskutiert, diese Grundsätze zum Zitat bei Kunstwerken auf Zitate im Rahmen der Wissenschaftsfreiheit und Freiheit der Lehre zu übertragen.¹⁸⁴ Für eine entsprechende verfassungsgemäße Auslegung des Zitatzwecks in § 51 UrhG hat sich das LG München beim Zitat von Texten eines Komikers in einem Vorlesungsskript ausgesprochen.¹⁸⁵ Gegen eine Übertragung dieser Grundsätze wird jedoch vorgebracht, dass zum einen zwar eine wissenschaftsspezifische Auslegung der Schrankenbestimmung möglich sei, jedoch Zitate anders als in der Kunst im wissenschaftlichen Bereich gerade nicht als Stilmittel, sondern als Beleg einer fremden Meinung eingesetzt werden.¹⁸⁶ Zum anderen sei eine großzügigere Auslegung der Schranke der Zitatzfreiheit jedenfalls für den Fall der Lehre nicht vereinbar, dass sonst die Schranken der §§ 60a ff. UrhG ausgehebelt werden würden.¹⁸⁷ Nach diesen Einwänden kann auf dieser Basis wohl auch kein weiterer Anwendungsbereich des Zitatrechts nach § 51 UrhG angenommen werden, wenn wissenschaftliche Textanalysen vorgenommen und dabei Nachweise angefügt werden.

V. Die Generalklausel des § 51 UrhG

Das bis 2007 geltende Prinzip der abschließenden Aufzählung von zulässigen Zitatformen in § 51 UrhG wurde durch das Zweite Gesetz zur Regelung des Urheberrechts in der Informationsgesellschaft¹⁸⁸ aufgegeben.¹⁸⁹ Im Zuge dessen wurde § 51 S. 1 UrhG unter Angleichung an Art. 5 Abs. 3 lit. d InfoSoc-RL (2001/29/EG) ausdrücklich als Generalklausel ausgestaltet.¹⁹⁰ Das bedeutet, dass nach § 51 S. 1 UrhG allgemein die Vervielfältigung, Verbreitung und öffentliche Wiedergabe eines veröffentlichten Werkes zum Zwecke des Zitats zulässig sind, sofern die Nutzung ihrem Umfang nach durch den besonderen

¹⁸³ BVerfG GRUR 2001, 149, 152 – Germania 3; siehe *Dustmann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 31.

¹⁸⁴ *Dustmann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 31.

¹⁸⁵ LG München I GRUR-RR 2006, 7, 8 – Karl Valentin.

¹⁸⁶ *Dustmann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 51 Rn. 31.

¹⁸⁷ *Lüft*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 4.

¹⁸⁸ BGBl. I 2513.

¹⁸⁹ *Spindler*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 93.

¹⁹⁰ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 22; siehe Entwurf eines Zweiten Gesetzes zur Regelung des Urheberrechts in der Informationsgesellschaft, BT-Drs. 16/1828, 25.

Zweck gerechtfertigt ist.¹⁹¹ Die Voraussetzungen des Zitatzwecks, der Einhaltung des gebotenen Umfangs, die Selbstständigkeit des zitierenden Werkes und die Veröffentlichung des zitierten Werks gelten damit auch für die Generalklausel.¹⁹² Das bedeutet, dass auch im Rahmen der Generalklausel bei Belegen für Textanalysen die Voraussetzungen des Zitatzwecks, insbesondere die Belegfunktion und das Kriterium der geistigen Auseinandersetzung, vorliegen müssen.¹⁹³ Die Regelbeispiele der Nr. 1 - 3¹⁹⁴ wurden dennoch beibehalten.¹⁹⁵ Zweck der Generalklausel ist es nach der Gesetzesbegründung, einzelne, aus der unflexiblen Grenzziehung des früheren Rechts (der Regelbeispiele) folgende Lücken zu schließen.¹⁹⁶ Die Wertungen der Regelbeispiele in Nr. 1 - 3 sind jedoch nach wie vor zu beachten, da der Gesetzgeber in diesem Rahmen die Befugnis zur Konkretisierung wahrgenommen hat.¹⁹⁷

Die Generalklausel umfasst daher alle Kleinzitate in jeder Werkart, jedoch nicht Musikwerke (siehe § 51 S. 2 Nr. 3 UrhG).¹⁹⁸ Beispielsweise erfasst § 51 UrhG über die Generalklausel das Filmzitat sowie das Zitat in Multimediawerken¹⁹⁹, wie Homepages oder Videospiele²⁰⁰.²⁰¹ Auch das sogenannte „große Kleinzitat“, das die Übernahme längerer Werkteile oder ganzer Werke zu anderen als wissenschaftlichen Zwecken beschreibt,²⁰² wird unter die Generalklausel gefasst, wenn der Zitatzweck ebensolches zulässt.²⁰³

VI. Ergebnis

Das Zitatrecht in § 51 UrhG bietet einen gewissen Freiraum für die geistige Auseinandersetzung mit urheberrechtlich geschützten Texten. Die Schranke

¹⁹¹ *Lüft*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 1.

¹⁹² *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 93.

¹⁹³ Hier näher bereits oben III. b) (1) und III. b) (2).

¹⁹⁴ Hierzu schon oben allgemein in II. und ausführlich zu Nr. 1 und Nr. 2 in III. und IV.

¹⁹⁵ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 1c.

¹⁹⁶ Entwurf eines Zweiten Gesetzes zur Regelung des Urheberrechts in der Informationsgesellschaft, BT-Drs. 16/1828, 25.

¹⁹⁷ *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 93.

¹⁹⁸ *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 94; schon vor Einführung der Generalklausel wurde § 51 Nr. 2 UrhG entsprechend erweitert ausgelegt: *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 23.

¹⁹⁹ Entwurf eines Zweiten Gesetzes zur Regelung des Urheberrechts in der Informationsgesellschaft, BT-Drs. 16/1828, 25.

²⁰⁰ *Ahlberg*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 2 Rn. 49.

²⁰¹ *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 Rn. 23.

²⁰² Zum wissenschaftlichen Großzitat siehe oben.

²⁰³ *Spindler*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 95.

dient generell der Freiheit der geistigen Auseinandersetzung mit dem Inhalt fremder Text(teilen) als Belege. Liegen die Voraussetzungen des § 51 UrhG vor, sind sämtliche Verwertungsrechte also insbesondere die Vervielfältigung, Verbreitung und öffentliche Wiedergabe der betroffenen Text(teile) – außer das Ausstellungsrecht nach § 18 Abs. 1 Nr. 1 UrhG – zulässig.

§ 51 UrhG setzt allgemeine Voraussetzungen der Zitierfreiheit fest. Zentrale allgemeine Voraussetzung aller Zitate (§ 51 S. 2 Nr. 1 - 3 UrhG) ist ein Zitatzweck, in dessen gebotenen Umfang das Zitat lediglich zulässig ist. Der BGH setzt als Zitatzweck voraus, dass eine innere Verbindung zwischen den verwendeten fremden Werken oder Werkteilen und den eigenen Gedanken des Zitierenden bestehen muss. Ein Zitat ist hiernach nur dann zulässig, wenn es als Beleg für selbstständige Ausführungen des Zitierenden genutzt wird. Der EuGH legt diese Anforderung für das Designrecht etwas weiter aus; ob dieser Ansatz – über das Designrecht hinaus – angewendet kann, ist offen. Der Zitatzweck legt den zulässigen Umfang eines Zitats fest: Es muss der für den Zitatzweck konkret erforderliche qualitative und quantitative Umfang getroffen sein.

Die Schranke des Zitatrechts nach § 51 UrhG ist nur dann erforderlich, wenn der zitierte Gegenstand urheberrechtlich schutzfähig ist (§ 2 Abs. 2 UrhG). Sie greift nur bei veröffentlichten Werken ein.

§ 51 S. 2 Nr. 1 UrhG regelt das wissenschaftliche Großzitat und privilegiert die Aufnahme einzelner ganzer Werke in ein selbstständiges wissenschaftliches Werk. Das wissenschaftliche Großzitat nach § 51 S. 2 Nr. 1 UrhG verlangt ausdrücklich einen bestimmten Zitatzweck: die Aufnahme „zur Erläuterung des Inhalts“ des zitierenden Werkes.

Werden lediglich Stellen eines Werkes zitiert, dann ist dies nach § 51 S. 2 Nr. 2 UrhG auch zu anderen als wissenschaftlichen Zwecken zulässig. Der Zitatzweck ist bei diesem Kleinzitat weiter gefasst als beim wissenschaftlichen Großzitat („zur Erläuterung des Inhalts“ § 51 S. 2 Nr. 1 UrhG).

§ 51 UrhG enthält neben den Regelbeispielen in Nr. 1 - 3 eine Generalklausel, deren Zweck es ist, einzelne, aus der unflexiblen Grenzziehung der Regelbeispiele folgende Lücken zu schließen. Die allgemeinen Voraussetzungen des § 51 UrhG, wie der Zitatzweck, der gebotene Umfang, die Selbstständigkeit des zitierenden Werkes und die Veröffentlichung des zitierten Werks gelten auch hier. Die Zitatzfreiheit in § 51 UrhG stellt eine selbstständige Schranke dar, die die geistige Auseinandersetzung mit Text(teilen) in gewissen Grenzen ermöglicht.

Liegen diese Voraussetzungen nicht vor, muss sich der Nutzer auf eine der übrigen Schranken (§§ 44a ff. UrhG) berufen können oder die Zustimmung der Rechteinhaber einholen.

Leseempfehlungen zur vertiefenden Lektüre: *Spindler*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 51; *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 51 und *Lüft*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51.

D. Die Wissenschaftsschranke in den Digital Humanities

Karolina Benedyk/Katharina Erler-Fridgen

Version 1.0 (05.01.2024), CC BY-SA 4.0.

Die Wissenschaftsschranke in § 60c UrhG ermöglicht es, für die Zwecke der nicht kommerziellen wissenschaftlichen Forschung Teile eines urheberrechtlich geschützten Schutzgegenstands zu vervielfältigen und gegebenenfalls zu verbreiten und öffentlich zugänglich zu machen. Sie dient der Freiheit der wissenschaftlichen Forschung und soll neue, digitale Arbeitsmittel und -methoden in der Wissenschaft unterstützen.¹ Kennen Wissenschaftlerinnen und Wissenschaftler ihre Voraussetzungen, erleichtert dies ihnen, urheberrechtliche Rahmenbedingungen für ihre wissenschaftliche Forschung einzuhalten und zu nutzen. Die Wissenschaftsschranke ist von den Text und Data Mining-Schranken (§ 44b und § 60d UrhG) abzugrenzen, welchen sich eine eigene Handreichung widmet.²

Textanalysen in den Digital Humanities setzen regelmäßig voraus, Texte zu sammeln, aufzubereiten und aus ihnen Informationen zu extrahieren und diese gegebenenfalls zu präsentieren und zu archivieren.³ Werden für diese Verfahrensschritte Kopien der urheberrechtlich geschützten⁴ Texte angefertigt oder Text(teile)⁵ präsentiert, können sie in das Vervielfältigungsrecht (§16 UrhG) und das Recht auf öffentliche Zugänglichmachung (§ 19a UrhG) sowie Entnahmen aus Datenbanken⁶ (§ 87a ff. UrhG) eingreifen.⁷ Die Text und Data Mining-Schranken (§ 44b und § 60d UrhG) ermöglichen zu diesem Zweck unter gewissen Voraussetzungen Vervielfältigungen. Abgegrenzte Personenkreise dürfen diese zur gemeinsamen Forschung (§ 60d Abs. 4 Nr. 1 UrhG) teilen.⁸ Die Wissenschaftsschranke in § 60c UrhG hingegen erlaubt, Kopien eines gewissen

¹ *Stieper*, in Schrickler/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 1.

² S. hierzu *Erler-Fridgen*, Teil 4, A. in diesem Band.

³ Zu den fünf Verfahrensschritten beim Text und Data Mining (Sammlung, Aufbereitung, Extraktion, Präsentation und Archivierung) siehe *Erler-Fridgen*, Teil 1, C. in diesem Band.

⁴ Zu den Kriterien der urheberrechtlichen Schutzfähigkeit siehe *Erler-Fridgen*, Teil 2, A. in diesem Band.

⁵ Zur Präsentation von Textteilen: *Erler-Fridgen*, Teil 6, A. in diesem Band.

⁶ *Erler-Fridgen*, Teil 2, C. in diesem Band.

⁷ *Erler-Fridgen*, Teil 1, C. in diesem Band.

⁸ *Erler-Fridgen*, Teil 4, A. in diesem Band.

Prozentsatzes eines Werkes für die wissenschaftliche Forschung anzufertigen. Zudem gestattet sie, solche Vervielfältigungen mit einem abgegrenzten Personenkreis zur gemeinsamen Forschung zu teilen (§ 60c Abs. 1 Nr. 1 UrhG). Wissenschaftlerinnen und Wissenschaftler dürfen die Kopien verwerten, auch wenn sie diese nicht zum Zwecke des TDM vornehmen.

Die Wissenschaftsschranke in § 60c UrhG beruht auf Art. 5 Abs. 3 lit. a InfoSoc-RL⁹ (RL 2001/29/EG)¹⁰. Hiernach kann es gerechtfertigt sein, die Schranke zum Zwecke der wissenschaftlichen Forschung außerhalb des Text und Data Mining – gegebenenfalls auch in dessen Umfeld¹¹ – zu nutzen.¹²

I. Voraussetzungen für die privilegierte Nutzung

§ 60c UrhG ermöglicht es, Teile eines Werkes für den Zweck der nicht kommerziellen wissenschaftlichen Forschung zu nutzen. Nach § 60c Abs. 1 UrhG dürfen bis zu 15 Prozent für einen spezifischen Kreis von Nutzenden (Nr. 1 und Nr. 2, vgl. S. 5 f.) vervielfältigt (§ 16 UrhG), verbreitet (§ 17 UrhG) und öffentlich zugänglich (§ 19a UrhG) gemacht werden.¹³ Der Umfang von 15 Prozent entspricht nach der Gesetzesbegründung dem Ziel, einen erleichterten Zugang zu Forschung zu verschaffen.¹⁴ § 60c Abs. 2 UrhG stellt es für die *eigene* wissenschaftliche Forschung frei, bis zu 75 Prozent eines Werkes zu nutzen. Absatz 3 erweitert den Umfang der zulässigen Nutzungen. Danach dürfen Abbildungen, einzelne Beiträge aus einer Fachzeitschrift oder wissenschaftlichen Zeitschrift (ebenso mehrere, einzelne Beiträge aus derselben Fachzeitschrift oder wissenschaftlichen Zeitschrift), sonstige Werke geringen Umfangs und vergriffene Werke vollständig genutzt werden.¹⁵ Demgegenüber verbietet Absatz 4 die

⁹ Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft, Rechtsgrundlagen des § 60c UrhG sind außerdem: Artikel 10 Absatz 1 Buchstabe d Vermiet- und Verleih-RL 2006/115/EG, Artikel 6 Absatz 2 Buchstabe b und Artikel 9 Buchstabe b Datenbanken-RL 96/9/EG, siehe BT-Drs. 18/12329, 39.

¹⁰ Lüft, in Wandtke/Bullinger, Urheberrecht, 6. Aufl. 2022, § 60c Rn. 6.

¹¹ Beispielsweise wird vertreten, § 60c UrhG als Möglichkeit zu nutzen, Auszüge von Korpora für Anschlussnutzungen wie Anschlussforschung zu erstellen und herauszugeben: siehe Kleinkopf/Jacke/Gärtner, MMR 2021, 196, 198; auch zur erforderlichen Kombinationsmöglichkeit von Schranken siehe ebenda 198 f.

¹² Vgl. Erwägungsgrund 15 S. 5 DSM-RL zum Verhältnis der TDM Schranken zu Art. 5 Abs. 3 lit. a InfoSoc-RL.

¹³ Lüft, in Wandtke/Bullinger, Urheberrecht, 6. Aufl. 2022, § 60c Rn. 9.

¹⁴ BT-Drs. 18/12329, 35.

¹⁵ Grübler, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 60c Rn. 15.

Liveaufnahme und spätere öffentliche Zugänglichmachung von Veranstaltungen wie öffentlichen Vorträgen vor Ort durch Dritte.¹⁶

1. Erfasste Schutzrechte

§ 60c UrhG umfasst Werke aller Art und damit auch Datenbankwerke^{17,18}. Auf Computerprogramme ist § 60c UrhG im Rahmen von § 69a Abs. 4 UrhG anwendbar.¹⁹ Bei Sammelwerken scheint es schwierig, die 15 Prozent-Regel auf die schutzbegründenden²⁰ Faktoren der Auswahl und Anordnung der Elemente anzuwenden.²¹ Daher wird in der rechtswissenschaftlichen Literatur vertreten, auf den einzelnen Beitrag eines Sammelwerks als eigenständiges Werk im Sinne von § 2 Abs. 1 Nr. 1-7 UrhG abzustellen.²² Anders als bei § 60a UrhG dürfen in § 60c UrhG auch unveröffentlichte Werke genutzt werden. Das erleichtert es etwa, Nachlässe zu erforschen.²³ Auf verwandte Schutzrechte ist § 60c UrhG anwendbar, soweit deren Regelungen auf die Norm verweisen.²⁴

2. Erfasster Personenkreis

Jeder und jede darf die Nutzungen in § 60c UrhG vornehmen, soweit sie für Forschungszwecke erfolgt. Dadurch erfasst die Schranke Forschende und Mitarbeitende an Forschungsinstituten sowie unabhängige Forschende, Studierende und Privatgelehrte.²⁵ Nach der Gesetzesbegründung dürfen insbesondere auch Dritte die Nutzungshandlungen in § 60c UrhG durchführen, die selbst keine Forschungszwecke verfolgen – wenn sie die Nutzungshandlungen für

¹⁶ BT-Drs. 18/12329, 39.

¹⁷ *Erler-Fridgen*, Teil 2, D. in diesem Band.

¹⁸ *Stieper*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 10.

¹⁹ *Stieper*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 10: Ob das auch für Computerprogramme gilt, die den elektronischen Zugang zu den einzelnen Elementen einer Datenbank vermitteln, ist umstritten. Schranken in Bezug auf Computerprogramme ergeben sich aus den gesonderten Regeln in §§ 69a – 69e UrhG, vgl. *Lüft*, in Wandtke/Bullinger, Urheberrechtsgesetz, 6. Auflage 2022, § 60c Rn. 4.

²⁰ Zum urheberrechtlichen Schutz von Sammelwerken siehe: *Erler-Fridgen*, Teil 2, A. in diesem Band.

²¹ *Berger*, GRUR 2017, 953, 960.

²² *Berger*, GRUR 2017, 953, 960; *Lüft*, in Wandtke/Bullinger, Urheberrecht, 6. Aufl. 2022, § 60c Rn. 14.

²³ *Lüft*, in Wandtke/Bullinger, Urheberrecht, 6. Aufl. 2022, § 60c Rn. 10.

²⁴ *Berger*, GRUR 2017, 953, 959; *Lüft*, in Wandtke/Bullinger, Urheberrecht, 6. Aufl. 2022, § 60c Rn. 3.

²⁵ BT-Drs. 18/12329, 39; *Lüft*, in Wandtke/Bullinger, Urheberrecht, 6. Aufl. 2022, § 60c Rn. 9.

Forschende vornehmen.²⁶ § 60c Abs. 1 Nr. 1 und Nr. 2 UrhG spezifizieren dann den Personenkreis (Adressaten), für den die Nutzungen (Vervielfältigung, öffentliche Zugänglichmachung) vorgenommen werden dürfen.²⁷

3. Beschränkung auf nicht kommerzielle wissenschaftliche Forschung

Privilegiert ist nach § 60c UrhG die nicht kommerzielle wissenschaftliche Forschung. Dieser unbestimmte Rechtsbegriff aus Art. 5 Abs. 3 lit. a InfoSoc-RL (2001/29/EG) ist nicht näher definiert. Er muss vor dem Hintergrund von Art. 13 S. 1 GRCh²⁸ ausgelegt werden.²⁹ Hiernach ist Forschung jede methodische und systematische Tätigkeit mit dem Ziel, in nachprüfbarer Weise neue Erkenntnisse zu erlangen.³⁰ Erfasst ist sowohl die eigentliche forschende Tätigkeit als auch die nachträgliche Darstellung der Forschungsergebnisse.³¹ Die bloße Information über den Stand der wissenschaftlichen Forschung ist hingegen nicht von der Wissenschaftsschranke umfasst.³²

Die Vorschrift beschränkt sich weiter auf die nicht kommerzielle wissenschaftliche Forschung. Nach Erwägungsgrund (42) InfoSoc-RL (2001/29/EG) wird dies durch die Tätigkeit als solche bestimmt und nicht etwa durch die organisatorische Struktur und Finanzierung der betreffenden Einrichtung. Das bedeutet, dass mit der Forschung als solcher kein Gewinn erzielt werden darf oder jedenfalls alle Gewinne in die Forschung reinvestiert werden.³³ Auf die Quelle der Finanzierung kommt es dabei nicht an: auch Forschung, die an öffentlichen Hochschulen stattfindet und über private Drittmittel finanziert wird, kann nach der Gesetzesbegründung unter § 60c UrhG fallen.³⁴ Auch die Tatsache, dass Forschende die Ergebnisse bei einem Verlag veröffentlichen und gegebenenfalls ein Honorar erhalten, beeinflusst die Qualifikation von Forschung als nicht kommerziell nicht.³⁵ Forscht jedoch ein Unternehmen, um Waren oder

²⁶ BT-Drs. 18/12329, 39.

²⁷ BT-Drs. 18/12329, 39.

²⁸ EU-Grundrechtecharta.

²⁹ *Stieper*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 5.

³⁰ *Stieper*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 5; *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 60c Rn. 1.

³¹ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 60c Rn. 1.

³² *Stieper*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 5; aA für eine Ausdehnung des Begriffs iSd früheren „wissenschaftlichen Gebrauchs“, s. *Lüft*, in *Wandtke/Bullinger*, Urheberrecht, 6. Aufl. 2022, § 60c Rn. 11.

³³ *Stieper*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 5.

³⁴ BT-Drs. 18/12329, 39.

³⁵ BT-Drs. 18/12329, 39.

Dienstleitungen zu entwickeln und diese zu vermarkten, so liegt eine kommerzielle Forschung vor.³⁶ In übrigen Fällen, beispielsweise bei Ausgründungen aus Hochschulen, kommt es auf die Umstände des Einzelfalls an.³⁷ Abzugrenzen ist auch eine Nutzung zum Zweck der wissenschaftlichen Lehre, die sich nach § 60a UrhG und nicht nach § 60c UrhG richtet. Forschungsergebnisse in Lehrbüchern oder Vorlesungen werden daher nicht von § 60c UrhG erfasst.³⁸ Ein zusätzlicher Lehrzweck neben der Forschung ist unschädlich und sperrt die Anwendung von § 60c UrhG nicht.³⁹

4. Berechtigter Personenkreis nach § 60c Abs. 1 Nr. 1 und Nr. 2 UrhG

§ 60c Abs. 1 Nr. 1 und Nr. 2 UrhG legt fest, für welche Adressaten zu welchen Nutzungen die Werke vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden dürfen. Von Nr. 1 ist ein bestimmt abgegrenzter Personenkreis für dessen *eigene* wissenschaftliche Forschung umfasst. Hierdurch sind Forschungsteams eingeschlossen, deren Teilnehmende selbst neue wissenschaftliche Erkenntnisse erzielen wollen.⁴⁰ Unerheblich ist dabei, ob alle Personen an derselben Forschungseinrichtung forschen,⁴¹ solange sie sich im Inland befinden.⁴² Forschende dürfen daher nach der Gesetzesbegründung Materialien auch innerhalb loser Forschungsverbände nutzen.⁴³ Wie groß eine solche Forschungsgruppe ist, hängt maßgeblich vom Forschungsgegenstand und dem damit verbundenen personellen Aufwand zusammen.⁴⁴ Wichtig ist in jedem Fall, dass der Kreis der Mitglieder einer solchen Forschungsgruppe klar abgegrenzt ist.⁴⁵ Bei der öffentlichen Zugänglichmachung ist außerdem zu beachten, dass der Personenkreis auf diejenigen eingegrenzt werden muss, die das Angebot für die eigene wissenschaftliche Forschung nutzen.⁴⁶ Die Beteiligten müssen dafür die nach dem jeweiligen Stand der Technik wirksamen Vorkehrungen treffen.

³⁶ BT-Drs. 18/12329, 39.

³⁷ Berger, GRUR 2017, 953, 961.

³⁸ Stieper, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 8.

³⁹ Stieper, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 8.

⁴⁰ Lüft, in Wandtke/Bullinger, Urheberrecht, 6. Aufl. 2022, § 60c Rn. 13.

⁴¹ BT-Drs. 18/12329, 39.

⁴² Vgl. dazu Benedyk, Teil 2, E. in diesem Band.

⁴³ BT-Drs. 18/12329, 39.

⁴⁴ Stieper, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 15.

⁴⁵ Berger, GRUR 2017, 953, 960.

⁴⁶ Zu § 51a UrhG a.F.: BT-Drs. 15/38, 20.

Privilegiert ist nach Nr. 1 allein die eigene Forschung. Werden die Vervielfältigungsstücke daher an außenstehende Forschende oder Institutionen weitergegeben, ist Nr. 1 nicht mehr einschlägig.⁴⁷

Nr. 2 berechtigt einzelne Dritte, die Qualität wissenschaftlicher Forschung zu überprüfen. Nach der Gesetzesbegründung soll dadurch Forschung in Peer Review-Verfahren vor Veröffentlichung oder Preisvergaben leichter überprüft werden können.⁴⁸ Daher dürfen auch verwendete Quellen zur Qualitätssicherung gesichtet werden.⁴⁹ Nicht freigestellt ist dabei jedoch eine öffentliche Zugänglichmachung an Dritte wie dies gegebenenfalls auf wissenschaftlichen Konferenzen üblich ist.⁵⁰ In der rechtswissenschaftlichen Literatur werden jedoch Zweifel daran geäußert, ob mit dem Umfang von 15 Prozent beispielsweise eine wirksame Plagiatskontrolle möglich ist.⁵¹

II. Voraussetzungen für den eigenen wissenschaftlichen Nutzen nach § 60c Abs. 2 UrhG

Für den eigenen wissenschaftlichen Nutzen wird der Umfang der zulässigen Nutzung in § 60c Abs. 2 UrhG auf 75 Prozent erweitert. Hierbei baut Absatz 2 auf § 60c Abs. 1 Nr. 1 UrhG auf und weitet den Umfang der freigestellten Nutzung von 15 auf 75 Prozent aus. Dabei muss die Nutzung für die eigene Forschung des Handelnden vorgesehen sein.⁵² Unerheblich ist dabei, ob das Werk ausleihbar oder der Erwerb des Werkes dem Forschenden zumutbar wäre.⁵³ Als Bezugsgröße für den zulässigen Umfang wird bei Büchern verbreitet auf die Gesamtheit des Werkes abgestellt, also einschließlich Inhaltsverzeichnis, Vorwort, Literaturverzeichnis und Register.⁵⁴ Wird ein Werk sowohl für die Lehre (§ 60a UrhG) als auch für die Forschung eingesetzt, so schließt dies eine Berufung auf § 60c Abs. 2 UrhG nicht aus.⁵⁵

⁴⁷ *Stieper*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 16.

⁴⁸ BT-Drs. 18/12329, 39.

⁴⁹ *Berger*, GRUR 2017, 953, 960.

⁵⁰ *Berger*, GRUR 2017, 953, 960; *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 60c Rn. 10.

⁵¹ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 60c Rn. 10; *Grübler*, in *Ahlber/Götting/Lauber-Rönsberg*, BeckOK Urheberrecht, 33. Edition 2022, § 60c Rn. 12.

⁵² *Stieper*, in Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 19.

⁵³ BT-Drs. 18/12329, 40.

⁵⁴ Siehe zu § 54a UrhG a.F.: BGH GRUR 2014, 549 Rn. 24 ff. – Meilensteine der Psychologie; *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 60c Rn. 11.

⁵⁵ *Dreier*, in *Dreier/Schulze*, Urheberrechtsgesetz, 7. Aufl. 2022, § 60c Rn. 11.

1. Privilegierter Personenkreis

Privilegiert sind nach § 60c Abs. 2 UrhG Forschende (natürliche Person) oder auch Einrichtungen, die für die Vervielfältigung verantwortlich sind.⁵⁶ Vorausgesetzt wird jedoch der eigene wissenschaftliche Nutzen. Demnach ist es nicht gestattet, die Vervielfältigungen an sonstige Dritte weiterzugeben.⁵⁷ Vervielfältigungen wie Scans und Kopien können jedoch auch Dritte herstellen, solange sie sich auf technischen Vorgang der Vervielfältigung beschränken.⁵⁸

2. Beschränkte Nutzungshandlungen

Die Nutzungsmöglichkeit von 75 Prozent eines Werks gilt nur für die Vervielfältigung von Werken. Das unterscheidet Absatz 2 von Absatz 1, da letzterer auch die Verbreitung und öffentliche Zugänglichmachung freistellt, dafür aber auf 15 Prozent eines Werkes begrenzt ist. Da Absatz 2 den Umfang der Vervielfältigung aus § 60c Abs. 1 UrhG erweitert, gilt auch für ihn die Beschränkung auf nicht kommerzielle wissenschaftliche Forschung.⁵⁹

III. Übernahme ganzer Werke nach § 60c Abs. 3 UrhG

§ 60c Abs. 3 UrhG gestattet im Gegensatz zu seinen Absätzen 1 (15 Prozent) und 2 (75 Prozent) die Übernahme ganzer Werke. Dies gilt für Abbildungen, einzelne Beiträge aus derselben Fachzeitschrift oder wissenschaftlichen Zeitschrift, sonstige Werke geringen Umfangs und vergriffene Werke. Diese dürfen nach § 60c Abs. 3 UrhG vollständig genutzt werden. Mit Abbildungen meint Absatz 3 auch Fotografien.⁶⁰ Nicht erfasst als wissenschaftliche Zeitschriften sind Tageszeitungen oder Publikumszeitschriften.⁶¹ Für diese gilt daher die 15 Prozent-Grenze des § 60c Abs. 1 UrhG.⁶² Beispiele für relevante Vervielfältigungshandlungen sind das Kopieren oder Scannen von Aufsätzen für eine eigene Forschungsarbeit, der Download wissenschaftlicher Texte geringen Um-

⁵⁶ *Grübler*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 60c Rn. 13.

⁵⁷ BT-Drs. 18/12329, 40; *Grübler*, in Ahlberg/Götting/Lauber-Rönsberg, BeckOK Urheberrecht, 33. Edition 2022, § 60c Rn. 13.

⁵⁸ *Stieper*, in Schrickler/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 19.

⁵⁹ *Dreier*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 60c Rn. 14.

⁶⁰ BT-Drs. 18/12329, 35.

⁶¹ *Lüft*, in Wandtke/Bullinger, Urheberrecht, 6. Aufl. 2022, § 60c Rn. 15.

⁶² *Lüft*, in Wandtke/Bullinger, Urheberrecht, 6. Aufl. 2022, § 60c Rn. 15.

fangs aus dem Internet, aber auch die Digitalisierung von (nicht übermäßig langen) Briefwechseln oder – für den Bereich der Bild-Text-Medien – von Weinetiketten.⁶³ Für den Maßstab sonstiger Werke geringen Umfangs kann nach der Gesetzesbegründung auf die Vereinbarung in Gesamtverträgen zwischen Verwertungsgesellschaften und Nutzenden zurückgegriffen werden.⁶⁴ Geringfügig sind etwa 25 Seiten bei Druckwerken, 5 Minuten bei Filmen und Musik.⁶⁵ Vergriffene⁶⁶ Werke dürfen vollständig genutzt werden, unabhängig davon, wie lange sie vergriffen sind.⁶⁷

IV. Anwendungsbereich weiterer Schranken

Von der Wissenschaftsschranke unberührt bleibt die Freiheit der Privatkopie des einzelnen Forschenden nach § 53 UrhG.⁶⁸ Auch sperrt § 60c UrhG nicht andere wissenschaftsrelevante Schranken wie das Zitatrecht nach § 51 UrhG.⁶⁹ Diese kann neben § 60c UrhG Anwendung finden.⁷⁰

V. Ergebnis

§ 60c UrhG macht es möglich, Teile eines Werkes zur nicht kommerziellen wissenschaftlichen Forschung zu nutzen. Grundsätzlich dürfen bis zu 15 Prozent eines Werkes vervielfältigt (§ 16 UrhG), verbreitet (§ 17 UrhG) und öffentlich zugänglich (§ 19a UrhG) gemacht werden (§ 60c Abs. 1 UrhG). § 60c Abs. 1 Nr. 1 und Nr. 2 UrhG begrenzen den Personenkreis, für den die Nutzungen vorgenommen werden dürfen. Nr. 1 erlaubt die Nutzungshandlungen zugunsten eines bestimmt abgegrenzten Personenkreises für deren eigene wissenschaftliche Forschung, Nr. 2 für einzelne Dritte, soweit dies der Überprüfung der Qualität wissenschaftlicher Forschung dient.

⁶³ *Stieper*, in Schrickler/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c Rn. 20.

⁶⁴ BT-Drs. 18/12329, 35.

⁶⁵ BT-Drs. 18/12329, 35.

⁶⁶ Vergriffen sind Printwerke, wenn sie „nicht mehr lieferbar“ sind, vgl. BT-Drs. 17/13423, 22.

⁶⁷ BT-Drs. 18/12329, 35.

⁶⁸ *Schack*, ZUM 2017, 802, 805; *Dreier*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 53 Rn. 3b.

⁶⁹ *Berger*, GRUR 2017, 953, 960; Das Zitatrecht verfolgt auch einen anderen Zweck als die Wissenschaftsschranke, denn sie fördert die geistige Auseinandersetzung insbesondere unabhängig davon, ob eine kommerzielle Nutzung gegeben ist: *Schack*, ZUM 2016, 266, 275.

⁷⁰ *Berger*, GRUR 2017, 953, 960.

Für die eigene wissenschaftliche Forschung darf man bis zu 75 Prozent eines Werkes nutzen (§ 60c Abs. 2 UrhG). Abbildungen, einzelne Beiträge aus derselben Fachzeitschrift oder wissenschaftlichen Zeitschrift, sonstige Werke geringen Umfangs und vergriffene Werke dürfen sogar vollständig genutzt werden (§ 60c Abs. 3 UrhG). § 60c Abs. 4 UrhG verbietet jedoch die Liveaufnahme und spätere öffentliche Zugänglichmachung von Veranstaltungen wie öffentlichen Vorträgen vor Ort.

Die Nutzungen in § 60c UrhG darf jede forschend tätige Person vornehmen (beispielsweise erfasst werden Forschende und Mitarbeiterinnen und Mitarbeiter an Forschungsinstituten sowie Studierende und Privatgelehrte). Wer Vielfältigkeitsstücke außerhalb der beschriebenen Freistellungen der Wissenschaftsschranke hinaus nutzen möchte, muss die Zustimmung der Rechteinhaber einholen oder sich auf eine der übrigen Schranken der §§ 44a ff. UrhG berufen können.

Leseempfehlungen zur vertiefenden Lektüre: *Stieper*, in Schrickler/Loewenheim, Urheberrecht, 6. Auflage 2020, § 60c; *Dreier*, in Dreier/Schulze, Urheberrechtsgesetz, 7. Aufl. 2022, § 60c.

Teil 5: Vertraglich erlaubte Handlungen

A. Vertragsrechtliche Lösungen

Karolina Benedyk

Version 1.0 (05.01.2024), CC BY-SA 4.0.

Das Urhebervertragsrecht befasst sich mit dem Verhältnis des Urhebers und dem in die Verwertung seines Werks einbezogenen Dritten.¹ In den seltensten Fällen verwerten Urheber heutzutage ihre Rechte selbst. Vielmehr erledigen das die Verwertungsunternehmen für die Schaffenden.² Das Urheberrecht und einige Leistungsschutzrechte sind grundsätzlich unübertragbar (vgl. § 29 I UrhG)³, sodass die Rechteverschaffung in Form von Lizenzen erfolgt.⁴ Verträge, die darauf ausgerichtet sind, Nutzungsrechte (also Lizenzen) einzuräumen, werden als Lizenzverträge bezeichnet.⁵

Der Urheber kann anderen gestatten, Handlungen vorzunehmen, die in den Schutzbereich seiner Rechte fallen, vgl. § 29 II UrhG. Es gibt unterschiedliche Gestaltungsformen, die sich nach dem Intensitätsgrad der erlaubten Handlung unterscheiden und deswegen auch als Stufenleiter bezeichnet werden.⁶ Die vollständige Übertragung unter Lebenden ist nur in engen Ausnahmefällen möglich, vgl. § 29 I UrhG. Räumt der Urheber ein ausschließliches Nutzungsrecht ein, berechtigt es den Inhaber des Nutzungsrechts, das Werk unter Ausschluss aller anderen Personen auf die ihm erlaubte Art zu nutzen und Nutzungsrechte einzuräumen, vgl. § 31 III 1 UrhG. Das bedeutet, dass der Lizenzinhaber das Werk alleinig nutzen darf. Der Urheber kann sich sein Nutzungsrecht nach

¹ *Soppe*, NJW 2018, 729, 730.

² *Loewenheim/J.B. Nordemann/Obly*, in: Loewenheim, Handbuch des Urheberrechts, 3. Auflage 2021, § 24 Rn. 1.

³ Vgl. hierzu *Benedyk*, Teil 1, B. in diesem Band.

⁴ *Loewenheim/J.B. Nordemann/Obly*, in: Loewenheim, Handbuch des Urheberrechts, 3. Auflage 2021, § 24 Rn. 1.

⁵ *Loewenheim/J.B. Nordemann/Obly*, in: Loewenheim, Handbuch des Urheberrechts, 3. Auflage 2021, § 24 Rn. 1; *Schulze*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 31 Rn. 4.

⁶ *Obly*, in: Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 29 Rn. 3.

§ 31 III 2 UrhG vorbehalten.⁷ Eine schwächere Form stellt das einfache Nutzungsrecht dar, welches den Inhaber berechtigt, das Werk auf die gestattete Art zu nutzen, ohne dass eine Nutzung durch andere ausgeschlossen ist, § 31 II UrhG. Der Sukzessionsschutz schützt das einfache und das ausschließliche Nutzungsrecht.⁸ Es bedeutet, dass die eingeräumten Nutzungsrechte gegenüber später eingeräumten wirksam bleiben, vgl. § 33 S. 1 UrhG. Ist der Sukzessionsschutz ausgeschlossen, handelt es sich bei dem eingeräumten Recht um eine rein schuldvertragliche Gestattung, die keine Wirkung gegenüber Dritten aufweist.⁹ Dem eingeräumten gestattet der Vertrag zu, sein Recht zu nutzen, er kann Dritte allerdings nicht davon abhalten, es ebenso zu tun. Die schwächste Form stellt die schlichte, einseitige Einwilligung dar. Sie ist jederzeit widerruflich und kann auch konkludent, also durch eine eindeutige Handlung ohne mündliche Abmachung, erfolgen.

Die gesetzlichen Regelungen des Urhebervertragsrecht haben immer den angemessenen Schutz des Urhebers im Sinn, der häufig durch eine hinreichende Kontrolle und eine faire Vergütung erreicht wird.¹⁰ „Kreative müssen fair vergütet werden“ war mithin das Motto zahlreicher Reformvorschläge.¹¹ Ausgangspunkt der Überlegungen war stets die strukturelle Unterlegenheit der Urheber gegenüber den Verlegern.¹² Die vertragsrechtliche Position des Urhebers ist im Jahre 2016 durch die §§ 32, 32a, 36 und 36a UrhG verbessert worden.

I. Grundzüge des Urhebervertragsrechts

Das Urhebervertragsrecht regelt in Grundzügen, wie Nutzungsrechte einzuräumen sind, vgl. §§ 31 ff. UrhG.¹³ Wie bereits beschrieben, gibt es kein einheitliches Urhebervertragsrecht. Das „Herzstück“ ist § 31 UrhG.¹⁴

⁷ *Obly*, in: Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 29 Rn. 3.

⁸ *Obly*, in: Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 29 Rn. 3.

⁹ *Obly*, in: Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 29 Rn. 3.

¹⁰ *Dreier*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Einl. Rn. 19.

¹¹ *Lucas-Schloetter*, GRUR 2017, 235.

¹² *Lucas-Schloetter*, GRUR 2017, 235, 236.

¹³ *Schulze*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 31 Rn. 3.

¹⁴ *Schulze*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 31 Rn. 1.

1. Verwertungsrechte

Ausgangspunkt des Urhebervertragsrechts sind die Verwertungsrechte. Diese ergeben sich aus §§ 15 ff. UrhG. Sie umschreiben die Befugnisse des Urhebers an seinem Werk.¹⁵ Sie sind gesetzlich eingeräumte ausschließliche Rechte, die Urhebern die wirtschaftliche Verwertung ihrer Werke ermöglichen.¹⁶ Inhaber sämtlicher Rechte ist der Urheber, § 7 UrhG. Verwerter können nur abgeleitete Rechte in Anspruch nehmen, sogenannte positive Nutzungsrechte.¹⁷ Aus den Verwertungsrechten kann der Urheber allerdings auch ein Verbotrecht ableiten und somit anderen Personen die Nutzung des Werkes verbieten.¹⁸

2. Nutzungsrechte, § 31 UrhG

Nach § 31 I UrhG kann der Urheber einem anderen das Recht einräumen, das Werk auf einzelne oder alle Nutzungsarten zu nutzen (Nutzungsrecht).

a) Einräumung von Nutzungsrechten

§ 31 I UrhG stellt einige Grundsätze für den Erwerb von Nutzungsrechten auf. Hiernach bedarf es einer Einräumung.¹⁹ Mit dem Einräumen von Nutzungsrechten ist eine Einigung im Sinne eines Vertrages gemeint.²⁰ Im allgemeinen Sprachgebrauch spricht man auch von Lizenzen.²¹ Eine vollständige Übertragung des Urheberrechts ist nicht möglich, der Kern verbleibt immer beim Urheber.²² Soweit ein Vertrag die Übertragung des Urheberrechts bezweckt, ist er insofern auszulegen, dass dem Vertragspartner Nutzungsrechte eingeräumt werden sollen.²³

¹⁵ *Wandtke*, in: *Wandtke/Bullinger, Urheberrecht*, 6. Auflage 2022, Vor §§ 31 ff. Rn. 23.

¹⁶ *Loewenheim/J.B. Nordemann/Obly*, in: *Loewenheim, Handbuch des Urheberrechts*, 3. Auflage 2021, § 24 Rn. 3.

¹⁷ *Schulze*, in: *Dreier/Schulze, Urheberrechtsgesetz*, 7. Auflage 2022, § 15 Rn. 2.

¹⁸ *Schulze*, in: *Dreier/Schulze, Urheberrechtsgesetz*, 7. Auflage 2022, § 15 Rn. 5.

¹⁹ *Loewenheim/J.B. Nordemann/Obly*, in: *Loewenheim, Handbuch des Urheberrechts*, 3. Auflage 2021, § 24 Rn. 4.

²⁰ *Wandtke*, in: *Wandtke/Bullinger, Urheberrecht*, 6. Auflage 2022, Vor §§ 31 ff. Rn. 22.

²¹ *Loewenheim/J.B. Nordemann/Obly*, in: *Loewenheim, Handbuch des Urheberrechts*, 3. Auflage 2021, § 24 Rn. 4.

²² Vgl. hierzu *Benedyk*, Teil 1, B. in diesem Band.

²³ *Schulze*, in: *Dreier/Schulze, Urheberrechtsgesetz*, 7. Auflage 2022, § 29 Rn. 15.

b) Übertragungszweckregel, § 31 V UrhG

Soweit der Urheber Nutzungsrechte überträgt und diese nicht ausdrücklich benannt sind, kommt neben den allgemeinen Auslegungsregeln die Übertragungszweckregel zum Einsatz, vgl. § 31 V UrhG. Diese besagt, dass der Urheber seinem Verwerter im Zweifel nur so viele Rechte überträgt, wie dieser für die Zwecke der in Aussicht genommenen Verwertung benötigt.²⁴ Dieser Grundsatz wird insbesondere dann angewandt, wenn die Nutzungsarten nicht besonders aufgelistet sind, vgl. § 31 V UrhG. An die Übertragungszweckregel ist gekoppelt, dass die Einräumung von Nutzungsarten, die zum Zeitpunkt des Vertragschlusses neu und unbekannt sind, von einer Schriftform abhängen, vgl. §§ 31a, 32 UrhG.²⁵ Positiv formuliert: Will sich ein Auftrag- oder Arbeitgeber mehr Nutzungsrechte einräumen lassen, als er für die Durchführung des Vertrags benötigt, muss er diese konkret benennen. Eine allgemeine Formulierung wie „es werden alle weltweiten Rechte für die Dauer des Urheberrechts eingeräumt“, reicht dafür etwa nicht aus.²⁶

c) Natur der Nutzungsrechte

Das Nutzungsrecht kann als einfaches oder ausschließliches Recht sowie räumlich, zeitlich und/oder inhaltlich beschränkt eingeräumt werden. Ein einfaches Nutzungsrecht berechtigt den Inhaber nach § 31 II UrhG, das Werk auf die erlaubte Art zu nutzen, ohne dass eine Nutzung durch andere ausgeschlossen ist. Der Urheber darf das Werk selbst weiter nutzen und anderen ebenfalls Nutzungsrechte einräumen. Das ausschließliche Nutzungsrecht berechtigt den Inhaber nach § 31 III UrhG, das Werk unter Ausschluss aller anderen Personen auf die ihm erlaubte Art zu nutzen und Nutzungsrechte einzuräumen. Das schließt auch die weitere Nutzung durch den Urheber ein. Es kann bestimmt werden, dass die Nutzung durch den Urheber vorbehalten bleibt.

3. Verwertungsgesellschaften

Verwertungsgesellschaften nehmen die Rechte an geschützten Werken und Leistungen treuhänderisch wahr.²⁷ Dem liegt die Prämisse zu Grunde, dass

²⁴ BGH GRUR 2019, 1191 – Online-Heftarchiv.

²⁵ Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Einl. Rn. 19.

²⁶ Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Einl. Rn. 19.

²⁷ Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Einl. Rn. 20.

Rechteinhaber ihre Rechte einzeln nur schwer bzw. unter hohen Kosten wahrnehmen können.²⁸ Die kollektive Verwertung der Rechte geht mit einer Pauschalisierung einher in Bezug auf die Bemessung und Verteilung von Einnahmen und Tarifen.²⁹ In Zeiten der Digitalisierung sind immer mehr Rechteinhaber bestrebt, aus der Solidargemeinschaft auszutreten und ihre Rechte selbst wahrzunehmen.³⁰ Hierdurch können sie ihre Rechte auch im Wege von Creative-Commons-Lizenzen zugänglich machen.³¹

II. Kodierte Regeln des Urhebervertragsrechts

Die wenigen Vorschriften im Urheberrechtsgesetz zum Urhebervertragsrecht sind besonders bedeutungsvoll. Die folgenden Ausführungen bieten einen Überblick über alle relevanten Vorschriften.

1. Angemessene Vergütung, § 32 UrhG

Nach § 32 UrhG hat der Urheber für die Einräumung von Nutzungsrechten und die Erlaubnis zur Werknutzung Anspruch auf die vertraglich vereinbarte Vergütung. Obwohl auch im Urhebervertragsrecht Vertragsfreiheit gilt, hat der Gesetzgeber sie bei der Berechnung der Urhebervergütung zugunsten des Schaffenden eingeschränkt.³²

Ist die Höhe der Vergütung nicht bestimmt, gilt die angemessene Vergütung als vereinbart. Eine angemessene Vergütung wird von Gesetzes wegen fingiert, sie gilt dann also als vereinbart.³³ Das bedeutet, dass ohne Vereinbarung eine in den Kreisen übliche Vergütung anzulegen ist. Soweit die vereinbarte Vergütung nicht angemessen ist, kann der Urheber von seinem Vertragspartner die Einwilligung in die Änderung des Vertrages verlangen, durch die dem Urheber die angemessene Vergütung gewährt wird, § 32 I 3 UrhG. Der Gesetzgeber hat es unterlassen, festzustellen, was mit einer angemessenen Vergütung gemeint ist.³⁴ Zudem gibt es keine festgelegten Beträge für bestimmte Bereiche.³⁵ Vielmehr ist

²⁸ *Ahlberg/Lauber-Rönsberg*, in: BeckOK Urheberrecht, Ahlberg/Götting/Lauber-Rönsberg, 36. Ed. 15.10.2022, UrhG Einführung zum UrhG Rn. 105.

²⁹ *Dreier*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Einl. Rn. 20.

³⁰ *Dreier*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Einl. Rn. 20.

³¹ Vgl. hierzu *Benedyk*, Teil 1, B. in diesem Band.

³² *Schulze*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 32 Rn. 1.

³³ *Soppe*, in: BeckOK Urheberrecht, Ahlberg/Götting/Lauber-Rönsberg, 36. Ed. 15.10.2022, UrhG § 32 Rn. 2.

³⁴ *Soppe*, NJW 2018, 729, 730.

³⁵ *Soppe*, NJW 2018, 729, 730.

der Begriff offen konzipiert, der Gesetzgeber verweist auf branchenspezifische Kollektivvereinbarungen.³⁶

§§ 32 ff. UrhG erfasst mehrere Anspruchsgrundlagen. § 32 I UrhG erhält einen Anspruch auf angemessene Vergütung. § 32a I 1 UrhG stellt einen Anspruch auf weitere angemessene Beteiligung, soweit sich im Nachhinein herausstellt, dass zwischen der gezahlten Vergütung und dem Vorteil, den der Vertragspartner erzielte, ein Missverhältnis besteht (sog. Bestseller-Paragraf).³⁷ § 32a II 1 UrhG erstreckt den Anspruch auf weitere angemessene Beteiligung in der Lieferkette, soweit sich der Vorteil daraus ergibt, was nachgelagerte Rechteinhaber in der Lieferkette erwirtschaften.³⁸

Ein Äquivalent zu § 32a UrhG ist § 32c UrhG für Werknutzungen, die im Zeitpunkt des Vertragsschlusses unbekannt sind. Nach § 32c I 1 UrhG hat der Urheber einen Anspruch auf eine gesonderte angemessene Vergütung, soweit der Vertragspartner eine neue Art der Werknutzung nach § 31a aufnimmt, die im Zeitpunkt des Vertragsschlusses vereinbart, aber noch unbekannt war. Absatz 2 bezieht die angemessene Vergütung dann auf spätere bekannte Nutzungsarten in der Lieferkette.

2. Anspruch auf Auskunft und Rechenschaft, §§ 32d, 32e UrhG

Dem Urheber steht ein gesondert geregelter Anspruch auf Auskunft und Rechenschaft über den Umfang der Nutzung und die hieraus gezogenen Erträge zu, vgl. § 32d I 1 UrhG. § 32e I 1 UrhG erstreckt den Auskunftsanspruch des Urhebers auf weitere Verwerter in der Lizenzkette, die nicht primäre Vertragspartner des Urhebers sind. Diese Ansprüche dienen der Transparenz.³⁹ Auf dieser Grundlage hat der Urheber die Möglichkeit zu erfahren, was mit seinen Werken passiert und wie hoch die Erträge aus der Verwertung sind.

3. Weiterübertragung der Nutzungsrechte, §§ 34, 35 UrhG

Nutzungsrechte dürfen grundsätzlich in gleicher Weise übertragen werden, vgl. §§ 34, 35 UrhG. Bei der Übertragung (§ 34 UrhG) geht die vollständige

³⁶ *Soppe*, NJW 2018, 729, 730.

³⁷ *Soppe*, NJW 2018, 729, 732.

³⁸ *Soppe*, NJW 2018, 729, 732.

³⁹ *Schulze*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 32d Rn. 1.

Rechtsposition auf den Erwerber über, sodass zwischen den beiden keine Beziehung mehr besteht.⁴⁰ Soweit A dem B ein Nutzungsrecht einräumt, B dieses Nutzungsrecht an C weitergibt, besteht zwischen A und B sowie B und C keine Rechtsbeziehung mehr, sondern nur zwischen A und C. Bei der Rechtseinräumung (§ 35 UrhG) gewährt der Rechteinhaber dem Erwerber Rechte, wodurch sie in einem Vertragsverhältnis verbleiben.⁴¹ In beiden Fällen sind ideelle und materielle Interessen des Urhebers betroffen, wodurch § 34 I 1 UrhG die Weiterübertragung von der Zustimmung des Urhebers abhängig macht.

4. Bearbeitung und Änderungen, §§ 37 I, 39 I UrhG

Für die Übertragung von Nutzungsrechten gilt allgemein die Zweckübertragungsregel, § 31 V UrhG. Hiernach verbleiben die einzelnen Nutzungsrechte im Zweifel beim Urheber.⁴² Dieser Gedanke wird durch die §§ 37 ff. UrhG verstärkt und bei einzelnen Nutzungsrechten herausgestellt. Das führt dazu, dass Urheber der Übertragung der Nutzungsrechte explizit zustimmen müssen. Der Nutzer darf das Werk grundsätzlich nicht bearbeiten, vgl. § 37 I UrhG. Diese Rechte verbleiben beim Urheber, wodurch er keine Enthaltungspflicht hat.⁴³ Der Urheber darf beispielsweise eine Bearbeitung des Werkes herstellen und verwerten und das Werk auf Bild- und Tonträger übertragen.

Auch das Änderungsrecht verbleibt nach § 39 I UrhG bei dem Urheber. Dies dient insbesondere zum Schutz seiner Werkintegrität, vgl. § 14 UrhG.⁴⁴ Werke sind nach dem Grundgedanken des Gesetzes persönliche geistige Schöpfungen, in welchen die Persönlichkeit der Urheber mit einfließt.⁴⁵ Schaffende entscheiden sich genau, wie sie die Werke in die Welt hinauslassen und so sollen diese auch verbleiben.

5. Form

Das Urhebervertragsrecht ist grundsätzlich formfrei. Nur in besonderen Fällen ist ein Schriftformerfordernis vorgesehen. Diese sind vorgesehen für Nut-

⁴⁰ *Soppe*, in: BeckOK Urheberrecht, Ahlberg/Götting/Lauber-Rönsberg, 36. Ed. 15.10.2022, UrhG § 34 Rn. 4 ff.

⁴¹ *Wandtke*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 35 Rn. 5.

⁴² *Schulze*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 37 Rn. 1.

⁴³ *Schulze*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 37 Rn. 2.

⁴⁴ *Schulze*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 39 Rn. 1.

⁴⁵ Zu den Kriterien der urheberrechtlichen Schutzfähigkeit siehe *Erler-Fridgen*, Teil 2, A. in diesem Band.

zungsrechte unbekannter Nutzungsarten. Nach § 31a I 1 UrhG bedarf ein Vertrag, durch den der Urheber Rechte für unbekannte Nutzungsarten einräumt oder sich dazu verpflichtet, der Schriftform. Sie ist nicht erforderlich, wenn der Urheber unentgeltlich ein einfaches Nutzungsrecht für alle Nutzenenden einräumt, § 31a I 2 UrhG. Auch Verträge über künftige Werke, die überhaupt nicht näher oder nur der Gattung nach bestimmt sind, ist eine Schriftform vorgegeben, vgl. § 40 I UrhG.

6. Das Rückrufsrecht des Urhebers wegen Nichtausübung, § 40a UrhG

Der Urheber hat bei der Einräumung eines ausschließlichen Nutzungsrechts gegen eine pauschale Gebühr grundsätzlich ein Recht auf anderweitige Verwertung nach 10 Jahren, § 40a I 1 UrhG.⁴⁶ Die Regelung wurde eingeführt, da ausschließliche Nutzungsrechte oft über die gesamte urheberrechtliche Schutzdauer eingeräumt wurden und der Urheber hierfür keine laufende Beteiligung an den Erlösen erhielt.⁴⁷

7. Rückruf der Nutzungsrechte, §§ 41, 42 UrhG

Zuletzt hat der Urheber das Recht die Nutzungsrechte zurückzurufen, soweit Nutzende von ihnen kein Gebrauch machen, vgl. §§ 41, 42 UrhG. Die Vorschrift ist weniger dem wirtschaftlichen als vielmehr dem Publikationsinteresse des Urhebers geschuldet.⁴⁸ Dadurch kann der Urheber verhindern, dass Verwerter seine Rechte vom Markt kaufen, aber diese nicht ausüben.⁴⁹

III. Der angestellte Urheber

Ob in der Wissenschaft oder der freien Wirtschaft, der angestellte Urheber ist heutzutage der Regelfall. Das Arbeitsverhältnis sieht vor, dass der Arbeitgeber den Arbeitnehmer dafür entlohnt, das Werk zu schaffen. In der amerikanischen Rechtsordnung gilt die „work made for hire“-Doktrin.⁵⁰ Hiernach ist der Arbeitgeber originärer Inhaber des urheberrechtlich geschützten Werkes. Anders ist es dagegen in der deutschen Rechtsordnung. Hier gilt das Schöpferprinzip, vgl. § 7 UrhG. Das liegt der monistischen Theorie zugrunde, die den Urheber in seinen geistigen und persönlichen Beziehungen zum Werk und in der

⁴⁶ Vgl. hierzu *Benedyk*, Teil 1, B. in diesem Band.

⁴⁷ *Lucas-Schloetter*, GRUR 2017, 235, 238.

⁴⁸ *Soppe*, NJW 2018, 729, 732.

⁴⁹ *Soppe*, NJW 2018, 729, 732.

⁵⁰ *Dreier*, in: *Dreier/Schulze*, Urheberrechtsgesetz, 7. Auflage 2022, § 43 Rn. 1.

Nutzung des Werkes schützt, vgl. § 11 UrhG. Die Theorie besagt, dass die persönlichkeitsrechtlichen und verwertungsrechtlichen Befugnisse des Urhebers untrennbar miteinander verbunden sind. Diese gilt auch im Rahmen von angestellten Urhebern.⁵¹

§ 43 UrhG hat in diesem Kontext eine klarstellende Wirkung. Hiernach gelten die Vorschriften dieses Unterabschnitts auch, wenn der Urheber das Werk in Erfüllung seiner Verpflichtungen aus einem Arbeits- oder Dienstverhältnis geschaffen hat, soweit sich aus dem Inhalt oder dem Wesen des Arbeits- oder Dienstverhältnisses nichts anderes ergibt. Angesichts der Unveräußerlichkeit des Urheberrechts kann der Arbeitgeber die Rechte am Werk nur erhalten, soweit ihm ein Nutzungsrecht eingeräumt wird, § 31 UrhG. Ein Arbeitsverhältnis umfasst alle abhängigen Beschäftigten.⁵² Abzugrenzen ist mithin von selbstständiger Arbeit. Abgrenzungskriterien ist die weisungsgebundene, fremdbestimmte Arbeit des Arbeitnehmers in persönlicher Abhängigkeit.⁵³ Das Dienstverhältnis umfasst öffentlich-rechtliche Dienstverhältnisse wie die der Beamten, ausgeschlossen sind privatrechtliche Dienstverhältnisse nach § 611 BGB.⁵⁴

Die Regelung erleichtert angestellten Urhebern den Rechteerwerb dahingehend, dass sich die Einräumung der Nutzungsrechte aus dem Inhalt oder dem Wesen des Arbeits- oder Dienstverhältnisses ergibt.⁵⁵ Das gilt auch dann, wenn der Arbeits- oder Dienstvertrag keine Regelungen dazu trifft.⁵⁶ Die Grenze ist hierbei die Wissenschaftsfreiheit nach Art. 5 III GG. In der Wissenschaft unterscheidet man zwischen freier wissenschaftlicher Tätigkeit und angestellten Forschern.⁵⁷ Letzteren stehen die Nutzungsrechte für die Dienstarbeit nicht zu. Bei Hochschullehrern ist eine Abgrenzung zu weisungsabhängigen Arbeitnehmern zu treffen. Bei den Ergebnissen ist auf den Charakter der Forschungsmaterialien, auf ihre Zweckbestimmung und die konkret gegebene Interessenlage zwischen Hochschullehrer und Universität abzustellen.⁵⁸ Im Vergleich zu weisungsabhängigen Arbeitnehmern wird die Arbeit eines Hochschullehrers über die rein dienstrechtliche Stellung hinaus wesentlich dadurch charakterisiert, dass er seine

⁵¹ BGH GRUR 2011, 59, 60 – Lärmschutzwand.

⁵² *Wandtke*, in: *Wandtke/Bullinger, Urheberrecht*, 6. Auflage 2022, § 43 Rn. 13.

⁵³ *Dreier*, in: *Dreier/Schulze, Urheberrechtsgesetz*, 7. Auflage 2022, § 43 Rn. 6.

⁵⁴ *Wandtke*, in: *Wandtke/Bullinger, Urheberrecht*, 6. Auflage 2022, § 43 Rn. 14.

⁵⁵ *Dreier*, in: *Dreier/Schulze, Urheberrechtsgesetz*, 7. Auflage 2022, § 43 Rn. 1.

⁵⁶ *Dreier*, in: *Dreier/Schulze, Urheberrechtsgesetz*, 7. Auflage 2022, § 43 Rn. 1.

⁵⁷ Vgl. hierzu *Benedyk*, Teil 1, B. in diesem Band.

⁵⁸ BGH GRUR 1991, 523, 525 – Grabungsmaterialien.

wissenschaftlichen Aufgaben durch Forschung und Lehre in seinem Fach selbständig wahrnimmt.⁵⁹ Hochschullehrer forschen frei und eigenverantwortlich und haben keine Pflicht, ihre Ergebnisse zu veröffentlichen.⁶⁰ Dadurch werden die Ergebnisse ihnen und nicht dem Dienstherrn zugeordnet.⁶¹ Genauso behandelt die Literatur Hochschuldozenten sowie Honorar- und Gastprofessoren.⁶² Bei wissenschaftlichen Assistenten stellt die Rechtsprechung darauf ab, ob es sich um eigene wissenschaftliche Arbeit handelt, oder solche die im Rahmen einer weisungsgebundenen Tätigkeit erfolgt.⁶³

Sonderregelungen gelten für Computerprogramme. Die speziellere Regelung ist hierbei § 69b UrhG. Hiernach stehen dem Arbeitgeber alle vermögensrechtlichen Befugnisse zu.

Da der Schutz wissenschaftlicher Ausgaben nach § 70 UrhG an die kreative menschliche Tätigkeit anknüpft,⁶⁴ kommt auch hier § 43 UrhG zur Anwendung. Knüpft der Schutz verwandter Schutzrechte an die kaufmännisch-organisatorische Investition an, sind keine Sonderregelungen für Arbeitgeber vorgesehen.⁶⁵

1. Werke in Erfüllung der arbeits- und dienstrechtlichen Aufgabe

§ 43 UrhG umfasst nicht alle Werke, sondern nur diejenigen, die in Erfüllung der arbeits- und dienstvertraglichen Verpflichtung entstehen. Ausgeschlossen sind also Werke, die vor oder nach Beendigung des Arbeitsverhältnisses entstanden sind.⁶⁶ Eine Abgrenzung ist jedoch zu den außervertraglichen bzw. nicht-dienstlichen Werken zu finden, die auch als freie Werke bezeichnet werden.⁶⁷ Dabei sind nur objektive Kriterien heranzuziehen und nicht etwa die subjektive Vorstellung des Arbeitgebers oder -nehmers bzw. die Tatsache, dass die Werke während der Arbeitszeit und in den Räumlichkeiten entstanden.⁶⁸ Entscheidend ist der innere Zusammenhang zwischen arbeitsvertraglicher Verpflichtung

⁵⁹ BGH GRUR 1991, 523, 525 – Grabungsmaterialien.

⁶⁰ Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 43 Rn. 12.

⁶¹ Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 43 Rn. 12.

⁶² Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 43 Rn. 12.

⁶³ LG Köln ZUM 2000, 579.

⁶⁴ Zum Leistungsschutzrecht für wissenschaftliche Ausgaben siehe: *Erler-Fridgen*, Teil 2, C. in diesem Band.

⁶⁵ Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 43 Rn. 4.

⁶⁶ Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 43 Rn. 9.

⁶⁷ Wandtke, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 43 Rn. 22.

⁶⁸ Wandtke, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 43 Rn. 17, 20 f.

und der Werkschöpfung.⁶⁹ Liegt der Arbeitsauftrag beispielsweise darin begründet, Handreichungen in der Schnittstelle von Recht und Digital Humanities zu schreiben, endet die arbeitsvertragliche Verpflichtung bei Artikeln zu anderer Thematik. Gleiches gilt entsprechend für dienstvertragliche Verpflichtungen.⁷⁰ Es gilt jedoch nicht die dienstvertragliche Vereinbarung, sondern der festgelegte Aufgabenbereich.⁷¹ Damit entfallen aus dem Aufgabenbereich beispielsweise Gutachtertätigkeiten.⁷²

2. Umfrage der Nutzungsrechtseinräumung

Die Frage, wie viele Nutzungsrechte der Arbeitnehmer dem Arbeitgeber einräumt, beantwortet der Arbeitsvertrag. Jedoch ist auch hierbei die Übertragungszweckregel nach § 31 V UrhG anzuwenden.⁷³ Hiernach verbleiben die Rechte beim Urheber, soweit der Arbeitgeber sie nicht benötigt. Die Auslegungsregel findet wiederum ihre Grenze, soweit klar aufgeführt ist, welche Nutzungsrechte übertragen werden sollen.

IV. Ergebnis

Das Urhebervertragsrecht regelt, inwiefern die aus den Verwertungsrechten abgespaltenen Nutzungsrechte lizenziert werden dürfen. Ihre Einräumung ist von einem Vertragsschluss abhängig. Bei der Auslegung dieses Vertrages kann auf die allgemeinen zivilrechtlichen Auslegungsregeln zurückgegriffen werden. Diese ergänzt das Gesetz durch die Übertragungszweckregel nach § 31 V UrhG. Hiernach ist die Einräumung der Rechte auf die für den Vertrag notwendigen beschränkt.

Des Weiteren ist es für Urheber oftmals schwierig, den Überblick über ihre Rechte und deren Verwertung zu erhalten. Hierbei helfen ihnen Verwertungsgesellschaften. Diese nehmen die Rechte an geschützten Werken und Leistungen treuhänderisch wahr. Durch die kollektive Lizenz- und Beitragseinräumung können sie Rechte im (digitalen) Massenverkehr sinnvoll nachverfolgen.

Obwohl nur wenige Bereiche des Vertragsrechts geregelt sind, greift das Urhebervertragsrecht in einigen Bereichen stark in die Vertragsfreiheit ein. Das lässt

⁶⁹ Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 43 Rn. 10.

⁷⁰ BGH GRUR 2011, 59 – Lärmschutzwand.

⁷¹ Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, § 43 Rn. 11.

⁷² Rojahn/Frank, in: Schricker/Loewenheim, Urheberrecht, 6. Auflage 2020, § 43 Rn. 28.

⁷³ BGH GRUR 2011, 56, 60 – Lärmschutzwand.

sich mit dem Schutz des Urhebers, der vermeintlich schwächeren Partei, begründen. Dieser Schutz ist auf unterschiedliche Art und Weise ausgestaltet. Einerseits steht die angemessene Vergütung des Urhebers im Vordergrund, sodass seine wirtschaftlichen Interessen bedient sind. Auf der anderen Seite ist dessen Urheberpersönlichkeitsrecht maßgeblich. Somit können sich Anwenderinnen und Anwender bei der Verwendung der Vorschriften auf den Schutz zurückbesinnen.

Leseempfehlung zur vertiefenden Lektüre: *Schulze*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor § 31.

B. Kollektive Lizenzen mit erweiterter Wirkung

Karolina Benedyk

Version 1.0 (05.01.2024), CC BY-SA 4.0.

Die Forschung in den Digital Humanities ist zunehmend daran interessiert, Forschungsfragen auf der Grundlage umfassender, relevanter Datensätze zu bearbeiten. Nur dies erlaubt es, Muster und Verteilungen von Phänomenen oder auch Entwicklungen von Phänomenen über die Zeit auf der Ebene größerer historischer, literarischer oder künstlerischer Teilsysteme zu betrachten, statt wie bislang häufig der Fall nur anhand ausgewählter, als repräsentativ gesetzter Beispiele. Entsprechend sind beispielsweise die in der digitalen Literaturwissenschaft genutzten Textsammlungen in den letzten 10 Jahren stetig größer geworden. Galt ein Textkorpus von mehreren Hundert Romanen oder Theaterstücken vor 10 Jahren vielleicht noch als umfangreich, ist es inzwischen nicht unüblich, in der Dimension Tausender, Zehntausender oder noch mehr Werken zu arbeiten.

Eine solche Forschungspraxis erfordert jedoch in unterschiedlichen Werkbereichen eine flächendeckende Verfügbarkeit der relevanten Werke. Soweit die Werke Leistungsschutzrechte bzw. urheberrechtlichen Schutz genießen, müssen die Betroffenen Nutzungsrechte von sämtlichen Rechteinhabern einholen, wenn und soweit sie sich nicht auf Schranken berufen können.¹ Das kann insbesondere bei der öffentlichen Zugänglichmachung von größeren Korpora oder von Forschungsergebnissen der Fall sein, wenn diese urheberrechtlich geschütztes Material von Dritten enthalten. Die individuelle Lizenzierung ist insbesondere bei digitaler Massennutzung zeit- und kostenintensiv – wenn nicht sogar unmöglich. Aushilfe verschaffen Verwertungsgesellschaften, die oftmals eine marktmäßige Stellung einnehmen.² Mit der neu eingeführten kollektiven Lizenz mit erweiterter Wirkung steht ihnen ein weiteres Instrument zur Verfügung, dass die Lizenzierung von großen Werkbereichen erleichtert.

Kollektive Lizenzen mit erweiterter Wirkung ermöglichen es, die Lizenzierung von Nutzung durch Verwertungsgesellschaften zu vereinfachen, indem auch

¹ Vgl. zum Werkschutz *Erler-Fridgen*, Teil 2, A. in diesem Band.

² *Staats*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 2.

Werke und Leistungsschutzrechte von Rechteinhabern eingeräumt werden können, die der Verwertungsgesellschaft keine Rechte einräumen (Außenstehende nach § 7a VGG).³ Dadurch erlaubt es das Gesetz Verwertungsgesellschaften, für bestimmte Werkkategorien flächendeckende Lizenzen zu vergeben.⁴

Eine grundlegende Voraussetzung dabei ist, dass die Verwertungsgesellschaft repräsentativ ist, vgl. § 51b I VGG. Hiernach liegt die Voraussetzung vor, wenn sie für eine ausreichend große Zahl von Rechteinhabern Rechte, die Gegenstand der kollektiven Lizenz sein sollen, auf vertraglicher Grundlage wahrnimmt. Die VGG muss eine ausreichend große Zahl von Rechteinhabern auf vertraglicher Grundlage vertreten.⁵ Es ist nicht erforderlich, dass sie die weit überwiegende Zahl oder auch nur die Mehrheit der Rechteinhaber vertritt.⁶ Eine genaue Abgrenzung ist von Gesetzes wegen nicht vorgegeben.

Will eine DH-Forschungsgruppe beispielsweise eine Textsammlung verarbeiten und hat die Verwertungsgesellschaft Wort für 70 Prozent der Texte Wahrnehmungsverträge mit den Rechteinhabern, dann kann sie für die verbleibenden 30 Prozent der Texte kollektive Lizenzen vergeben.

Kollektive Lizenzen mit erweiterter Wirkung beruhen auf optionalen Vorgaben der Art. 12 DSM-RL.^{7, 8} Die Regelungen sind seit dem 07. Juni 2021 unter den Voraussetzungen der §§ 51 ff. VGG im deutschen Gesetz kodifiziert. Sie entstanden nach dem Vorbild der *extended collective licensing* der nordischen Länder.⁹ Es handelt sich um ein neues Instrument des deutschen und des europäischen Unionsurheberrechts. Hiernach tritt die Lösung neben die gesetzlichen Schranken.¹⁰

I. Sinn und Zweck

Verwertungsgesellschaften zeichnen sich dadurch aus, dass sie eine Vielzahl von Werken und Leistungsschutzrechten bündeln und diese vermarkten.¹¹ Hierdurch ermöglichen sie rechtmäßige Nutzungen, die eine umfassende An-

³ Staats, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 2.

⁴ Staats, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 2.

⁵ Raue, in Dreier/Schulze, § 51b VGG, Rn. 3.

⁶ Raue, in Dreier/Schulze, § 51b VGG, Rn. 3.

⁷ Art. 12 der Richtlinie über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinie 96/9/EG und 2001/29/EG v. 17.4.2019.

⁸ Raue, in Dreier/Schulze, Vor §§ 51 – 52e VGG, Rn. 1.

⁹ Sutterer, GRUR 2021, 662; Raue, in Dreier/Schulze, Vor §§ 51 – 52e VGG Rn. 1.

¹⁰ Raue, in Dreier/Schulze, Vor §§ 51 – 52e VGG, Rn. 1.

¹¹ Sutterer, GRUR 2021, 662, 663.

zahl an Werken oder Leistungsschutzrechten erfordert. Allerdings schließen außerhalb der verwertungsgesellschaftspflichtigen Rechte nicht alle Rechteinhaber Wahrnehmungsverträge mit einer Verwertungsgesellschaft.¹² Hintergrund der kollektiven Lizenzen mit erweiterter Wirkung ist, dass Verwertungsgesellschaften nicht immer alle Rechteinhaber einer Kategorie von Werken und Leistungsschutzrechten vertreten.¹³ Dadurch können sie bei Nutzungshandlungen, die eine massenhafte Verwertung erfordern, oftmals nicht alle Rechte einräumen.¹⁴ Eine individuelle Lizenzierung scheitert an zu hohen Transaktionskosten.¹⁵ Die Verwertung bleibt mithin aus oder ist rechtswidrig. Dadurch bleiben oftmals ältere oder im Handel nicht mehr erhältliche Werke ungenutzt. Ihr kommerzieller Wert mag zwar schwach sein, allerdings können sie aus kultureller bzw. wissenschaftlicher Sicht umso bedeutsamer sein.¹⁶

Eine Anpassung der Schranken ist oft langwierig und scheitert an zu großen Widerständen. Deswegen ermöglichen es die kollektiven Lizenzen mit erweiterter Wirkung Verwertungsgesellschaften, Lizenzen für Außenstehende nach § 7a VGG zu erteilen.¹⁷ Nach § 7a VGG ist Außenstehender ein Rechteinhaber, der im Hinblick auf die betreffende Nutzung nicht in einem vertraglichen Wahrnehmungsverhältnis mit einer Verwertungsgesellschaft steht. Kollektive Lizenzen mit erweiterter Wirkung kehren die Voraussetzungen eines Wahrnehmungsvertrages um.¹⁸ Anstelle, dass Rechteinhaber einen Wahrnehmungsvertrag abschließen müssen (opt-in), um die Rechte zu vermarkten, müssen sie nunmehr widersprechen (opt-out), soweit Verwertungsgesellschaften ihre Rechte nicht einräumen sollen.¹⁹ Dadurch können Schutzgegenstände von Unentschlossenen mitlizenzieren werden.²⁰ Kollektive Lizenzen mit erweiterter Wirkung setzen mithin einen Grundstein für die Funktionsfähigkeit von Lizenzmechanismen bei Massendigitalisierungen oder im Massenverkehr der Onlinenutzungen über Plattformen.²¹

Kollektive Lizenzen mit erweiterter Wirkung sind nicht als Schranken einzuordnen. Es handelt sich vielmehr um eine vertragliche Einräumung der Rechte.²²

¹² Sutterer, GRUR 2021, 662, 663.

¹³ Sutterer, GRUR 2021, 662, 663; Raue, in Dreier/Schulze, Vor §§ 51 – 52e VGG, Rn. 1.

¹⁴ Raue, in Dreier/Schulze, Vor §§ 51 – 52e VGG, Rn. 1.

¹⁵ ErwG 45 DSM-RL; de la Durantaye, GRUR 2020, 7.

¹⁶ de la Durantaye, GRUR 2020, 7.

¹⁷ Raue, in Dreier/Schulze, Vor §§ 51 – 52e VGG, Rn. 1.

¹⁸ Sutterer, GRUR 2021, 662, 663.

¹⁹ de la Durantaye, GRUR 2020, 7; Sutterer, GRUR 2021, 662, 663.

²⁰ Sutterer, GRUR 2021, 662, 663.

²¹ Sutterer, GRUR 2021, 662, 663.

²² Raue, in Dreier/Schulze, Vor §§ 51 – 52e VGG, Rn. 3.

Der Rechteinhaber hat die Möglichkeit, der Lizenz zu widersprechen, vgl. § 51 II bzw. § 52 II VGG.

II. Kollektive Lizenz mit erweiterter Wirkung, § 51 VGG

Soweit eine Verwertungsgesellschaft über ein großes, aber nicht allumfassendes Repertoire an Schutzrechten verfügt, ermöglicht es § 51 VGG der Verwertungsgesellschaft, Nutzenden Lizenzen für Werke Außenstehender einzuräumen.²³ § 51 VGG ist dabei die Eingangsnorm, die ihre Wirkung im Zusammenspiel mit den weiteren Regelungen entfaltet.²⁴ Die Formulierung „kann“ suggeriert, dass die Lizenzvergabe im Ermessen der Verwertungsgesellschaft liegt.

1. Erweiterte Wirkung

Abs. 1 ermöglicht Verwertungsgesellschaften, Rechte von Außenstehenden (§ 7a VGG) zu lizenzieren. Dabei beschränkt die Vorschrift die Adressaten auf Verwertungsgesellschaften und abhängige Verwertungseinrichtungen. Die vertraglichen Nutzungsrechte sind auf keinen Vertragspartner beschränkt.²⁵

Grundlage kollektiver Lizenzen ist, dass die Verwertungsgesellschaft mit einem Nutzenden einen Vertrag über die kollektive Nutzung ihres Repertoires schließt.²⁶ Sie kann kollektive Lizenzen mit erweiterter Nutzung nur dort ermöglichen, wo ihr bereits Berechtigte Nutzungsrechte eingeräumt haben.²⁷ Beispielsweise könnte die VG Wort eine kollektive Lizenz im Bereich von Belletristik oder wissenschaftlicher Fachliteratur einräumen, soweit sie bereits mit vielen Rechteinhabern zusammenarbeitet. Sie kann das nun eben auch für Werke tun, deren Verlage kein Vertragspartner der VG Wort sind, es sei denn diese widersprechen dieser Praxis.

Die kollektive Lizenz ist eine kollektive, vertragliche Rechteeinräumung über wahrgenommene Rechte der Verwertungsgesellschaft.²⁸ Mit erweiterter Wirkung ist gemeint, dass sich die kollektive Lizenz auf die Rechte Außenstehender erstreckt.²⁹

²³ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 1.

²⁴ *Staats*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 13.

²⁵ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 4.

²⁶ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 6.

²⁷ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 6.

²⁸ *Staats*, ZUM 2019, 703, 707.

²⁹ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 7.

2. Internationaler Lizenzumfang

Die Verwertungsgesellschaft darf kollektive Lizenzen mit erweiterter Nutzung nur für die Verwertung im Inland einräumen.³⁰ §§ 51 ff. VGG beschränken die Vergabe von Lizenzen auf das nationale Hoheitsgebiet.³¹ Das hat insbesondere Folgen für die Online-Nutzung. Die Lizenznehmer und Lizenznehmerinnen müssen den Zugang der Nutzung auf Deutschland beschränken.³² Das ist im Kontext zunehmend international agierender und rezipierter Forschung eine gewichtige Einschränkung.

3. Widerspruchsrecht, Abs. 2

Der Außenstehende hat das Recht, der Rechteinräumung zu widersprechen, § 51 II VGG.³³ Der Widerspruch ist an die Verwertungsgesellschaft zu richten.³⁴ Dieser ist an keine Form gebunden und kann jederzeit erfolgen. Hierbei unterscheidet man den ex ante- von dem ex post-Widerspruch.³⁵ Erfolgt der Widerspruch innerhalb der dreimonatigen Informationspflicht (§ 51a II Nr. 4 VGG), ist die Rechteinräumung von vornherein unwirksam, vgl. § 51a II Nr. 5 VGG (ex ante-Widerspruch).³⁶ Der ex post-Widerspruch entwickelt Wirkung für die Zukunft, auch wenn die Verwertungsgesellschaft bereits eine kollektive Lizenz mit erweiterter Wirkung ausgestellt hat.³⁷ Lizenznehmer und Lizenznehmerinnen müssen die Nutzung aber erst nach einer angemessenen Frist beenden.³⁸ Ob dem Nutzenden eine Aufbrauchfrist zusteht, ergibt sich nicht aus der Gesetzesbegründung.³⁹

4. Rechte und Pflichten der Außenstehenden, Abs. 3

Außenstehende haben die gleichen Rechte und Pflichten wie vertraglich Berechtigte.⁴⁰ Sie partizipieren an den Erlösen. Verwertungsgesellschaften müssen sie auf dieselbe Weise informieren wie Berechtigte.⁴¹

³⁰ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 8.

³¹ *Staats*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 4.

³² Vgl. hierzu *Benedyk*, Teil 2, E. in diesem Band.

³³ *Staats*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 20.

³⁴ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 11.

³⁵ *Staats*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 21.

³⁶ *Staats*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 21.

³⁷ *Staats*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 21.

³⁸ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 15.

³⁹ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 15.

⁴⁰ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 16.

⁴¹ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 16.

III. Wirksamkeit der Rechtseinräumung, § 51a VGG

§ 51a VGG stellt kumulative Voraussetzungen für die Wirksamkeit kollektiver Lizenzen mit erweiterter Wirkung. Insbesondere enthält sie Schutzbestimmungen zugunsten der Außenstehenden (§ 7a VGG).

1. Voraussetzungen nach Abs. 1

§ 51a I Nr. 1 VGG fordert, dass nur repräsentative Verwertungsgesellschaften kollektive Lizenzen mit erweiterter Wirkung ausstellen dürfen. Damit verweist die Norm auf die Definition nach § 51b VGG, wonach eine Verwertungsgesellschaft repräsentativ ist, wenn sie die einzuräumenden Rechte für eine ausreichend große Zahl von Rechteinhabern auf vertraglicher Grundlage wahrnimmt.

Eine solche Lizenz dürfen Verwertungsgesellschaften nur vergeben, soweit es unzumutbar ist, individuelle Lizenzen von allen Außenstehenden einzuholen. Das ist anzunehmen, soweit die Transaktionskosten zu hoch sind.⁴² Unerheblich bleibt, ob die Einzellizenzierung mit einigen Außenstehenden möglich ist. Vielmehr dürfen Verwertungsgesellschaften auf die Art der Nutzung oder den Typ der betroffenen Schutzgegenstände abstellen.⁴³ Das ist beispielsweise bei Massennutzungen im Onlinebereich denkbar.

Kollektive Lizenzen mit erweiterter Wirkung sind auf das Inland beschränkt, § 51a I Nr. 3 VGG (s.o.).

Die Verwertungsgesellschaft muss Außenstehende über die geplante Lizenzierung informieren. Damit soll ihnen ermöglicht werden, der Rechtseinräumung zu widersprechen.⁴⁴ Wie die Informationspflicht ausgestaltet sein muss, ist im Gesetz aufgeführt (lit. a – d).

2. Dauerhafte Informationspflicht, § 51a II VGG

Die Verwertungsgesellschaft ist nach § 51a II VGG verpflichtet, die Informationen nach § 51a I Nr. 4 VGG dauerhaft auf ihrer Internetseite zu veröffentlichen.⁴⁵ Diese hat auch nach Beendigung der Lizenzierung zu erfolgen. Das gibt den Außenstehenden die Möglichkeit, ihre Vergütungsansprüche geltend zu machen.

⁴² *Raue*, in Dreier/Schulze, § 51a VGG, Rn. 4.

⁴³ *Raue*, in Dreier/Schulze, § 51a VGG, Rn. 4.

⁴⁴ *Raue*, in Dreier/Schulze, § 51a VGG, Rn. 6.

⁴⁵ *Raue*, in Dreier/Schulze, § 51a VGG, Rn. 9.

IV. Kollektive Lizenzen mit erweiterter Wirkung für nicht verfügbare Werke

Kollektive Lizenzen mit erweiterter Wirkung für nicht verfügbare Werke beruhen auf Art. 8 – 11 DSM-RL^{46, 47}. Die Regelungen enthalten besondere Erfordernisse im Vergleich zu Art. 12 DSM-RL⁴⁸ und haben deswegen Vorrang.⁴⁹ § 52 VGG ermöglicht es, kollektive Lizenzen mit erweiterter Wirkung für nicht verfügbare Werke einzugehen. Nach § 52b I VGG handelt es sich bei nicht verfügbaren Werken um solche, die der Allgemeinheit auf keinem üblichen Vertriebsweg in einer vollständigen Fassung angeboten werden.

Dies ist besonders wichtig, da ansonsten Kulturerbe in Vergessenheit geraten könnte.⁵⁰ Wollen Kulturerbe-Einrichtungen nämlich ihnen verfügbare Werke digitalisieren, handelt es sich um urheberrechtliche Verwertungen. Um rechtmäßig zu handeln, müssen sie von den Rechteinhabern individuelle Einwilligungen einholen.⁵¹ Das kann mit großem Aufwand verbunden sein. Haben auch Verwertungsgesellschaften keine Wahrnehmungsverträge mit den Urhebern, droht, dass die Werke in Zukunft nicht mehr verfügbar sind.⁵² Diese Vorgehensweise soll eine „digitale Renaissance“ ermöglichen.⁵³

Die Voraussetzungen für nicht verfügbare Werke gleichen den allgemeinen Bestimmungen aus §§ 51 ff. VGG. Auch hier liegt eine verschachtelte Regelungssystematik vor.⁵⁴

Eine Ausnahme ergibt sich aus § 52 I 1 Nr. 2 VGG. Hiernach ist die Wirksamkeit auf die Vervielfältigung, Verbreitung, öffentliche Zugänglichmachung und sonstige öffentliche Wiedergabe zu nicht kommerziellen Zwecken beschränkt. Des Weiteren muss sich das betreffende Werk im Bestand der Kulturerbe-Einrichtung befinden. Zudem muss die Verwertungsgesellschaft ihre Informationspflicht erfüllen, indem sie sechs Monate vor Beginn der Rechtseinräumung im Online-Portal des Amtes der Europäischen Union für geistiges Eigentum

⁴⁶ Art. 8 – 11 der Richtlinie über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinie 96/9/EG und 2001/29/EG v. 17.4.2019.

⁴⁷ *Staats*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 9.

⁴⁸ Art. 12 der Richtlinie über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinie 96/9/EG und 2001/29/EG v. 17.4.2019.

⁴⁹ *Staats*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 9.

⁵⁰ *Raue*, in Dreier/Schulze, § 52 VGG, Rn. 1.

⁵¹ *Raue*, in Dreier/Schulze, § 52 VGG, Rn. 1.

⁵² *Raue*, in Dreier/Schulze, § 52 VGG, Rn. 1.

⁵³ *Geiger/Frosio/Bulayenko*, JIPITEC 2018, 240.

⁵⁴ *Raue*, in Dreier/Schulze, § 52 VGG, Rn. 2.

(EUIPO) über das betreffende Werk, die Vertragspartei, die betroffenen Nutzungsrechte, deren Geltungsbereich und das Widerspruchsrecht des Außenstehenden informieren muss, vgl. § 52a I 1 Nr. 4 VGG. Widersprechen müssen Rechteinhaber gem. § 52 II VGG vor dem EUIPO.⁵⁵ Zudem ist die Rechteeinräumung im Fall von vergriffenen Werken nicht freiwillig. Die Verwertungsgesellschaft muss diese anbieten.

V. Ergebnis

Kollektive Lizenzen mit erweiterter Wirkung erlauben es Verwertungsgesellschaften, Lizenzen auch für Werke und Leistungsschutzrechte von Außenstehenden zu vergeben, die mit ihnen keinen Wahrnehmungsvertrag geschlossen haben (§ 7a VGG). Verwertungsgesellschaften müssen Rechteinhaber nicht fragen, ob sie ihre Werke lizenzieren dürfen, vielmehr müssen die Betroffenen widersprechen, soweit sie der Lizenzierung nicht zustimmen.

Hiermit ermöglicht das Gesetz Werknutzern einen erleichterten Zugang zu einer Vielzahl von Daten. Das ist insbesondere im Onlinebereich hilfreich. Sie kann kollektive Lizenzen mit erweiterter Nutzung jedoch nur dort ermöglichen, wo ihr bereits Rechteinhaber Nutzungsrechte eingeräumt haben.⁵⁶ Zudem dürfen nur repräsentative Verwertungsgesellschaften die Rechte vergeben und sind dabei auf das deutsche Hoheitsgebiet beschränkt. Die Voraussetzungen der Rechteeinräumung sind in § 51a VGG gebündelt aufgezählt.

Für Forscherinnen und Forscher in den Digital Humanities bedeutet das eine Erleichterung, soweit sie beispielsweise eine flächendeckende Textsammlung aufbauen. Ohne die vertragliche Möglichkeit von kollektiven Lizenzen mit erweiterter Wirkung, müssten sie individuelle Lizenzen mit jedem Rechteinhaber einholen, der keinen Wahrnehmungsvertrag mit der Verwertungsgesellschaft abgeschlossen hat. Das ist nicht nur kostenintensiv, sondern in manchen Fällen sogar unmöglich. Viele Werke würden dadurch untergehen. Die kollektiven Lizenzen mit erweiterter Wirkung ermöglichen es den Verwertungsgesellschaften nun, für all diejenigen Rechteinhaber Lizenzen zu erteilen, die noch keinen Vertrag mit der Verwertungsgesellschaft haben. Die Forschungsgruppen müssen sich mithin nicht selbst um die Vertragsabstimmungen kümmern.

§§ 52 ff. VGG regelt kollektive Lizenzen mit erweiterter Wirkung für nicht verfügbare Werke. Nach § 52b I VGG handelt es sich bei nicht verfügbaren Werken

⁵⁵ *Staats*, in: Wandtke/Bullinger, Urheberrecht, 6. Auflage 2022, § 51 VGG Rn. 20.

⁵⁶ *Raue*, in Dreier/Schulze, § 51 VGG, Rn. 6.

um solche, die der Allgemeinheit auf keinem üblichen Vertriebsweg in einer vollständigen Fassung angeboten werden. Verwertungsgesellschaften stehen dabei nicht vor der Wahl, Rechte einzuräumen, da die Vorschrift bezweckt, Kulturerbe vor dem Vergessen zu bewahren.

Leseempfehlung zur vertiefenden Lektüre: *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, 7. Auflage 2022, Vor §§ 51 – 52e VGG.

Teil 6: Weitere Aspekte

A. Die Präsentation von Textteilen als Ergänzung von Textanalysen

Katharina Erler-Fridgen

Version 1.0 (19.12.2021), CC BY-SA 4.0.

Werden kurze Teile analysierter Ausgangstexte präsentiert, kann dies Dritten möglich machen, Analyseergebnisse nachzuvollziehen und zu überprüfen – insbesondere im Kontext von Text und Data Mining. Solche kurzen Teile eines Werkes können eigenständig urheberrechtlich geschützt sein, sodass ihre Nutzung urheberrechtlichen Restriktionen unterliegt. So können Kopiervorgänge rechtswidrige Vervielfältigungshandlungen nach § 16 UrhG darstellen. Auch kann in diesen Fällen in das Recht auf öffentliche Zugänglichmachung nach § 19a UrhG eingegriffen werden, wenn die Textteile gemeinsam mit den Analyseergebnissen im Netz oder für weitere Forschende präsentiert werden. Im Folgenden wird ausgeführt, in welchen Fällen Textteile für sich alleinstehend urheberrechtlich geschützt sind und welchen zusätzlichen Schutz Teile von Presseveröffentlichungen erfahren können.

I. Hintergrund der Präsentation von Textteilen in den Digital Humanities¹

Hintergrund der Präsentation von Textteilen bei Textanalysen ist, dass dies Dritten möglich machen kann, die Ergebnisse der Analyse nachzuvollziehen. Denn manche Textanalyseverfahren stellen eine vermeintliche Eindeutigkeit in ihren Ergebnissen dar (bei der Erkennung von Entitäten), die in gewissen Fällen belegt werden sollte. Sollen zum Beispiel durch das Verfahren der *Named Entity*

¹ Die genannten Beispiele wurden gemeinsam mit Dr. Maria Hinzmann, Koordinatorin des Projekts MiMoText, erarbeitet.

*Recognition*² Autorennamen in literaturwissenschaftlichen Abhandlungen erkannt werden, so können Umschreibungen des Autors oder Namensabkürzungen zu Zweifelsfällen führen.³ Werden diese Analyseergebnisse mittels eines hinterlegten Textteils belegt, so wird das Ergebnis überprüfbar. Wird der Autor beispielsweise mittels Personalpronomen erwähnt, so können etwa der vorangegangene Satz bzw. Teile dieses Satzes als Beleg der Referenz hilfreich sein.

Auch bei Übereinstimmung von Werktitel und Figurennamen⁴ können im Rahmen der Named Entity Recognition Klarheit der jeweiligen Zuordnung und entsprechende Abgrenzung durch hinterlegte Textteile hergestellt werden. Hier wäre das Ergebnis der Textanalyse überprüfbar, wenn der Textteil hinterlegt würde, der die Entität repräsentiert (hier könnte bereits die Wortebene ausreichen). In beiden Beispielen aus der Anwendung der Named Entity Recognition können *kurze Textteile* als Beleg genügen.

Längere Textpassagen können beispielsweise als Beleg bei der Analyse von *Themenaussagen* nützlich sein. Die Extraktion von Themenaussagen im Rahmen des *Information Retrieval*⁵ aus literarischen Quellen in der Struktur von Subjekt-Prädikat-Objekt (sog. Resource Description Framework/RDF-Triple) kann durch Textteile belegt werden. Denn es ist für die Überprüfung der extrahierten Themenaussage wesentlich, den Satz oder Satzteil zu präsentieren, aus dem das zugrundeliegende sog. Triple generiert wurde. Beispielsweise wäre eine solche Themenaussage ‚Candide‘ → ABOUT → ‚Theodizee‘ im Werk „Candide“ von Voltaire, die über ein RDF-Triple (Subjekt-Prädikat-Objekt) ausgelesen werden kann. Der dem ausgelesenen Triple zugrundeliegende Satz „Voltaires Candide ou L’optimisme (1759) lässt sich als pointierte Absage an die Theodizee und den Leibnizschen Optimismus lesen“⁶ könnte als ein Beleg für diese Themenaussage hilfreich sein.

² Neumann, 5.3 Text-basiertes Informationsmanagement, in K.-U. Carstensen, Ch. Ebert, C. Ebert, S. Jekat, R. Klabunde, H. Langer (Hrsg.), Computerlinguistik und Sprachtechnologie, 3. Aufl., 2010, 596 ff.

³ Beispielsweise wird der Autor umschrieben oder eine Kurzform des Autorennamens oder ein Pseudonym genutzt.

⁴ So zum Beispiel bei Voltaires Werk „Candide“, in dem Werktitel und Figurenname „Candide“ übereinstimmen.

⁵ Ignatow/Mihalcea, Text Mining, 137 ff.

⁶ Schlüter, Von der Aufklärung bis zur Französischen Revolution, in: Grimm/Hartwig, Französische Literaturgeschichte. 6. Auflage, 2014, 196, 201.

II. Urheberrechtlicher Schutz von Textteilen

Textteile wie Kapitel, Textabschnitte oder auch einzelne Sätze und Satzteile können für sich allein stehend persönliche geistige Schöpfungen nach § 2 Abs. 2 UrhG darstellen. Deshalb kann auch die Übernahme nur einzelner Teile eines Textes Urheberrechtsverletzungen begründen.⁷

1. Schutzvoraussetzungen

Dafür müssen die Textteile für sich allein genommen die Schutzvoraussetzungen des § 2 Abs. 2 UrhG erfüllen.⁸ § 2 Abs. 2 UrhG setzt voraus, dass eine persönliche geistige Schöpfung in wahrnehmbarer Form vorliegt, die einen geistigen Gehalt und hinreichende Individualität aufweisen muss.⁹ Im Besonderen steht dabei für kurze Textteile im Fokus, ob der jeweilige Teil *ausreichende Individualität*¹⁰ beanspruchen kann.¹¹ Nach dem EuGH muss der Werkteil an der Originalität des Gesamtwerkes teilnehmen.¹² Auch kleinste Teile eines Werkes können für sich genommen schutzfähig sein.¹³ Je kürzer ein Text jedoch ist, desto höhere Anforderungen sind an seine Originalität zu stellen, um noch eine eigenschöpferische Prägung nach § 2 Abs. 1 Nr. 1 UrhG annehmen zu können.¹⁴ (Wenige) einzelne Wörter¹⁵ oder Zahlen¹⁶ stellen für sich keine geistige Schöpfung des Urhebers dar.

⁷ A. Nordemann, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 2 Rn. 51.

⁸ Loewenheim/Leistner, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 87.

⁹ Ausführlich zu den Kriterien der Schutzfähigkeit Jotzo, Teil 2, B. in diesem Band, 61, 64ff und Schack, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 180 ff.; siehe auch Erler-Fridgen, Teil 2, A. in diesem Band.

¹⁰ Hinreichende Individualität verlangt, dass das Werk vom individuellen Geist des Urhebers geprägt sein muss und das Ergebnis eines individuellen Schaffens darstellt, siehe Loewenheim/Leistner, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 2 Rn. 50 f.

¹¹ Schulze, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 2 Rn. 76.

¹² EuGH C-5/08, ECLI:EU:C:2009:465, GRUR 2009, 1041 Rn. 38 – Infopaq/DDF; damit meint der EuGH selbige Voraussetzung, denn er betont auch, dass nichts darauf hindeute, dass Teile eines Werkes einer anderen Regelung unterliegen als das Gesamtwerk, so A. Nordemann in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 2 Rn. 51.

¹³ A. Nordemann, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 2 Rn. 51; Schulze, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 2 Rn. 76 mwN.

¹⁴ OLG Köln ZUM-RD 2016, 470, 471 – Wenn das Haus nasse Füße hat.

¹⁵ BGH GRUR 2011, 134 Rn. 54 f. – Perlentaucher; OLG Köln ZUM-RD 2016, 470, 471 – Wenn das Haus nasse Füße hat.

¹⁶ Für den Fall eines Computerprogramms: EuGH C-406/10, ECLI:EU:C:2012:259, GRUR 2012, 814 Rn. 66 – SAS Institute Inc./World Programming Ltd.

Der EuGH hat in seinem Urteil *Infopaq/DDF* entschieden, dass es nicht ausgeschlossen ist, dass schon ein Auszug von 11 Wörtern aus einem Presseartikel eine eigene geistige Schöpfung des Urhebers zum Ausdruck bringt und damit schutzfähig ist.¹⁷ Die *Infopaq*-Entscheidung des EuGH wird wiederkehrend bei der Beurteilung der Urheberrechtsschutzfähigkeit kurzer Textteile herangezogen.¹⁸ Der BGH hat sich in seiner Entscheidung *Perlentaucher*¹⁹ den Ausführungen des EuGHs in *Infopaq* zwar angeschlossen, betonte aber, dass „(...) bei sehr kleinen Teilen eines Sprachwerks – wie einzelnen Wörtern oder Wortfolgen – Urheberrechtsschutz meist daran scheitern wird, dass diese für sich genommen nicht hinreichend individuell sind (...)“.²⁰

2. Teile von Gebrauchstexten und informativen Dokumenten

Bei *Gebrauchstexten* wie Gebrauchsanleitungen, Merkblättern oder Lexika werden für die urheberrechtliche Schutzfähigkeit, so auch im Fall kleiner Textteile, strengere Maßstäbe als bei literarischen Texten angesetzt.²¹ Deswegen sind Gebrauchstexte eher ungeschützt als Teile literarischer Texte. Werden in Texten Tatsachen wiedergegeben oder dienen diese Texte allein Gebrauchszwecken, so sind diese nur dann urheberrechtlich geschützt, sofern das Alltägliche, das Handwerksmäßige oder die mechanisch-technische Aneinanderreihung des Materials deutlich überragt wird.²² Der EuGH hat im Zusammenhang mit militärischen Lageberichten klargestellt, dass rein *informative Dokumente*, die vorhandene Informationen bloß wiedergeben, keinen Raum für schöpferische Originalität lassen.²³ Auch bei solchen Texten gilt: Je länger ein Textteil ist, desto

¹⁷ EuGH C-5/08, ECLI:EU:C:2009:465, GRUR 2009, 1041 Rn. 47 f. – *Infopaq/DDF*.

¹⁸ *Schippan*, ZUM 2013, 358, 363 unter Verweis auf die Urteile des LG München I in ZUM 2013, 230, 235; ZUM-RD 2011, 562, 565, die für 30-60 Wörter die Urheberrechtsfähigkeit von Textauszügen „erst recht“ bejahen; BGH GRUR 2011, 134 Rn. 54 – *Perlentaucher*; LG Köln ZUM-RD 2015, 279, 282 – Ausschreibungsunterlagen.

¹⁹ Der BGH vergibt Namen wie „*Perlentaucher*“, um seine Urteile zu kennzeichnen. Der EuGH benennt seine Urteile nach den jeweiligen Parteien des Rechtsstreits, siehe etwa „*Infopaq/DDF*“.

²⁰ BGH GRUR 2011, 134 Rn. 54 – *Perlentaucher* unter Verweis auf BGHZ 9, 262, 266 ff. – *Lied der Wildbahn I*.

²¹ BGH GRUR 1993, 34, 36 – *Bedienungsanweisung*; BGH GRUR 1987, 704, 706 – *Warenzeichenlexika*; siehe dazu auch *A. Nordemann*, in *Fromm/Nordemann, Urheberrecht*, 12. Aufl. 2018, § 2 Rn. 61 mwN.

²² OLG Köln NJW-RR 2016, 165 Rn. 26 – *Afghanistan Papiere*.

²³ EuGH C-469/17, ECLI:EU:C:2019:623, GRUR 2019, 934 Rn. 24 – *Funke Medien NRW GmbH*; der BGH hat diese unionsrechtlichen Vorgaben in seinem Urteil *Afghanistan Papiere II*, NJW 2020, 2547 Rn. 13 übernommen und angewandt.

eher kann Raum für Gestaltungsmöglichkeiten verbleiben, die eine entsprechend streng geprüfte eigenschöpferische Prägung auslösen können.²⁴ Bei kurzen Teilen von Gebrauchstexten sind damit besonders strenge Maßstäbe an die erforderliche Individualität zu stellen.

3. Weitere Beispiele

Werbeslogans erfüllen im Regelfall allein auf Grund ihrer *Länge* nicht den erforderlichen Grad an Schöpfungshöhe.²⁵ Anderes kann dann gelten, wenn der Verfasser besondere sprachliche Stilmittel wie die Reimform nutzt oder sich sehr pointiert ausdrückt.²⁶ Bei kurzen *wissenschaftlichen Texten* muss die Schutzfähigkeit immer unter dem Gesichtspunkt der freibleibenden wissenschaftlichen Lehre sowie Forschung beurteilt werden.²⁷ Denn grundsätzlich sind bei wissenschaftlichen Texten die Grenzen der Schutzfähigkeit enger, weil wissenschaftliche Erkenntnisse und Ideen urheberrechtlich frei bleiben sollen.²⁸ In diesem Fall ist für den urheberrechtlichen Schutz ein verbleibender Raum für eigenschöpferische Darstellung, Reihenfolge der Darstellung oder Verknüpfung von Fakten notwendig.²⁹

III. Leistungsschutzrecht der Presseverleger

Textteile ohne ausreichende Individualität nach § 2 Abs. 2 UrhG können aber durch das Leistungsschutzrecht des Presseverlegers geschützt sein. Vor der jüngst vorgenommenen Umsetzung³⁰ des europäischen Leistungsschutzrechts aus Art. 15 der Richtlinie über das Urheberrecht im digitalen Binnenmarkt (DSM-RL)³¹ in den §§ 87f ff. UrhG regelten §§ 87f ff. UrhG alte Fassung (a.F.) bereits ein deutsches Leistungsschutzrecht der Presseverleger.

²⁴ OLG Köln NJW-RR 2016, 165 Rn. 26 – Afghanistan Papiere.

²⁵ OLG Frankfurt am Main GRUR 1987, 44, 45 – „für das aufregendste Ereignis des Jahres“; OLG Düsseldorf GRUR 1978, 640, 641 – „fahr’n auf der Autobahn“.

²⁶ *Schippan*, ZUM 2013, 358, 365 mwN.

²⁷ *Schulze*, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, § 2 Rn. 26.

²⁸ *A. Nordemann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 2 Rn. 118; zum urheberrechtlichen Schutz wissenschaftlicher Texte siehe auch *Erler-Fridgen*, Teil 2, A. in diesem Band.

²⁹ *Schippan*, ZUM 2013, 358, 366.

³⁰ Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, in der Fassung der Bekanntmachung vom 4. Juni 2021, BGBl. I S. 1204 ff.

³¹ Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG.

1. Deutsches Leistungsschutzrecht als Vorgängervorschrift

Bereits das Leistungsschutzrecht in §§ 87f und g UrhG a.F.³² räumte Presseverlegern das ausschließliche Recht ein, *Presseerzeugnisse* zu gewerblichen Zwecken öffentlich zugänglich zu machen, es sei denn, es handelte sich um *einzelne Wörter oder kleinste Textausschnitte*. Geschützt werden sollte die verlegerische Leistung als solche.³³ Presseerzeugnisse waren nach § 87f Abs. 2 S. 1 UrhG a.F. redaktionell-technische Festlegungen journalistischer Beiträge im Rahmen einer Sammlung, die bei Würdigung der Gesamtumstände als überwiegend verlagstypisch anzusehen sind und die nicht überwiegend der Eigenwerbung dienen.³⁴ Der EuGH hatte jedoch einen Verstoß der Bundesregierung gegen die Notifizierungspflicht nach Art. 8 Transparenz-RL³⁵ im zugrundeliegenden Gesetzgebungsverfahren festgestellt.³⁶ Deswegen war diese deutsche Vorschrift daraufhin unanwendbar.³⁷

2. Europäisches Leistungsschutzrecht und Umsetzung ins deutsche UrhG

Das Unionsrecht sieht nunmehr in der DSM-RL in Artikel 15 ein entsprechendes europäisches Leistungsschutzrecht für Presseverleger vor. Dieser jüngst in den §§ 87f ff. UrhG umgesetzte unionsrechtliche Rahmen gibt für das europäische Leistungsschutzrecht von Presseverlagen³⁸ eine längere Schutzdauer von zwei Jahren³⁹ (Art. 15 Abs. 4 DSM-RL/§ 87j UrhG) vor.⁴⁰ Der Schutzgegenstand *Presseveröffentlichung* ist in Art. 2 Nr. 4 DSM-RL (§ 87f Abs. 1 UrhG)

³² Das im deutschen Urheberrecht im Jahr 2013 eingeführte Leistungsschutzrecht des Presseverlegers bezweckte, dass Presseverlage im Onlinebereich nicht schlechter gestellt werden als andere Vermittler von Werken und sollte den Schutz von Presseerzeugnissen im Internet verbessern: Amtliche Begründung zum Entwurf eines Siebenten Gesetzes zur Änderung des Urheberrechtsgesetzes, BT-Drs. 17/11470, 6.

³³ Dreier, in Dreier/Schulze, Urheberrechtsgesetz, 6. Aufl. 2018, Vor § 87f Rn. 3.

³⁴ Zum Begriff des Presseerzeugnisses und Ausfüllung der enthaltenen unbestimmten Rechtsbegriffe siehe *Czychowski/J.B. Nordemann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 87f Rn. 15 ff.

³⁵ Richtlinie 98/34/EG des Europäischen Parlaments und des Rates vom 22. Juni 1998 über ein Informationsverfahren auf dem Gebiet der Normen und technischen Vorschriften; siehe zur Notifizierungspflicht und dem Leistungsschutz des Presseverlegers auch *Jani*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 87f Rn. 8.

³⁶ EuGH C-299/17, ECLI:EU:C:2019:716, ZUM 2019, 838 Rn. 40 – VG Media/Google.

³⁷ *Stieper*, GRUR 2019, 1264, 1266.

³⁸ Zum Begriff des Presseverlages in Art. 15 DSM-RL siehe *Ackermann*, ZUM 2019, 375, 376.

³⁹ § 87 g Abs. 2 UrhG a.F. sah eine Schutzdauer von einem Jahr vor.

⁴⁰ *Stieper*, in Schrickler/Loewenheim, Urheberrecht, 6. Aufl. 2020, Vor. § 87 Rn. 14a.

definiert⁴¹ und erfasst ausdrücklich keine Periodika, die für wissenschaftliche oder akademische Zwecke veröffentlicht werden, etwa Wissenschaftsjournale. Das europäische Leistungsschutzrecht erfasst nicht lediglich Suchmaschinenbetreiber, sondern alle Anbieter von Diensten der Informationsgesellschaft (Art. 15 UAbs. 1 DSM-RL/§ 87g Abs. 1 UrhG).⁴² Ausgeschlossen vom Leistungsschutzrecht ist jedoch die Nutzung *einzelner Worte oder sehr kurzer Auszüge* nach Art. 15 Abs. 1 UAbs. 4 DSM-RL (§ 87g Abs. 2 Nr. 4 UrhG). Über seinen Vorgänger, das ehemalige deutsche Leistungsschutzrecht, hinausgehend eröffnet Art. 15 DSM-RL neben einem ausschließlichen Recht zur öffentlichen Zugänglichmachung auch ein ausschließliches Vervielfältigungsrecht der Presseverlage (Art. 15 UAbs. 1 DSM-RL/§ 87 Abs. 1 Alt. 2 UrhG).⁴³

IV. Ergebnis

Textteile können für sich genommen urheberrechtlichen Schutz erfahren, wenn sie die Schutzvoraussetzungen des § 2 Abs. 2 UrhG erfüllen. Insbesondere muss der Werkteil hinreichend individuell geprägt sein. Dies hält der EuGH bereits bei einem Auszug von 11 Wörtern für nicht ausgeschlossen. Je länger ein Textteil ist, desto mehr Gestaltungsspielraum bleibt für eine individuelle Prägung. Bei Gebrauchstexten werden strengere Maßstäbe an die erforderliche Individualität gestellt.

Bei Presseveröffentlichungen kann der Nutzung von Textteilen – auch unter der Schwelle der nach § 2 Abs. 2 UrhG notwendigen Individualität – das Leistungsschutzrecht des Presseverlegers entgegenstehen. Dem Leistungsschutzrecht unterliegen jedoch nicht die Nutzung einzelner Worte oder sehr kurzer Auszüge.

⁴¹ Art 2 Nr. 4 DSM-RL definiert eine Presseveröffentlichung als eine Sammlung, die hauptsächlich aus literarischen Werken journalistischer Art besteht, aber auch sonstige Werke oder sonstige Schutzgegenstände enthalten kann, und die a) in einer unter einem einheitlichen Titel periodisch erscheinenden oder regelmäßig aktualisierten Veröffentlichung, etwa Zeitungen oder Magazinen von allgemeinem oder besonderem Interesse, eine Einzelausgabe darstellt; b) dem Zweck dient, die Öffentlichkeit über Nachrichten oder andere Themen zu informieren und c) unabhängig vom Medium auf Initiative eines Diensteanbieters unter seiner redaktionellen Verantwortung und Aufsicht veröffentlicht wird; ausführlich hierzu *Ackermann*, ZUM 2019, 375, 376.

⁴² Hierin liegt ein wesentlicher Unterschied zu § 87g Abs. 4 UrhG a.F., siehe *Stieper*, in *Schricker/Loewenheim*, Urheberrecht, 6. Aufl. 2020, Vor. § 87 Rn. 14a.

⁴³ *Stieper*, in *Schricker/Loewenheim*, Urheberrecht, 6. Aufl. 2020, Vor. § 87 Rn. 14a.

Wer urheberrechtlich geschützte Textteile als Ergänzung zu Textanalysen einsetzen möchte und dabei urheberrechtlich relevante Nutzungshandlungen vornimmt, muss entweder die Zustimmung des Rechtsinhabers einholen oder auf urheberrechtliche Schranken (§§ 44a ff. UrhG) zurückgreifen. Die Schranke der Zitierfreiheit nach § 51 UrhG etwa erfordert insbesondere eine Belegfunktion⁴⁴ des Zitats, eine innere Verbindung zwischen dem zitierenden und zitierten Werk,⁴⁵ und ermöglicht nur einen dadurch gerechtfertigten Umfang des Zitats.⁴⁶

Leseempfehlungen zur vertiefenden Lektüre: *A. Nordemann*, in Fromm/Nordemann, Urheberrecht, 12. Aufl. 2018, § 2 Rn. 51, *Schippan*, Der Schutz von kurzen Textwerken im digitalen Zeitalter, ZUM 2013, 358 sowie *Stieper*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, Vor. § 87 und § 87f und *Lüft*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51.

⁴⁴ Einen bei dem Kleinzitat nach § 51 S. 2 Nr. 2 UrhG für Werke der Kunst vertretenen Zitat zweck über die Belegfunktion hinaus (BVerfG GRUR 2001, 149, 152 – Germania 3) und dessen Diskussion sind nicht Gegenstand der vorliegenden Ausführungen.

⁴⁵ Eine innere Verbindung zwischen dem zitierenden und zitierten Werk setzt voraus, dass das fremde Werk zum Gegenstand der geistigen Auseinandersetzung geworden ist: *Schack*, Urheber- und Urhebervertragsrecht, 9. Aufl. 2019, Rn. 545 mwN; insbesondere auch OLG Hamburg NJW 1972, 2304, 2305 – Handbuch moderner Zitate; BGH 2008, 693 Rn. 42 ff. – TV-Total; BGH GRUR 2017, 1027 Rn. 14 – Reformistischer Aufbruch.

⁴⁶ *Spindler*, in Schricker/Loewenheim, Urheberrecht, 6. Aufl. 2020, § 51 Rn. 27 ff.; *Lüft*, in Wandtke/Bullinger, Urheberrecht, 5. Aufl. 2019, § 51 Rn. 6 f.

B. Datenschutz bei den Digital Humanities – ein Überblick

Karolina Benedyk

Version 1.0 (05.01.2024), CC BY-SA 4.0.

Der Datenschutz in Europa hat zwei Schutzziele, zum einen den Schutz der Privatsphäre und zum anderen den weitreichenden Schutz des allgemeinen Persönlichkeitsrechts.¹ Dabei leitet das BVerfG den grundrechtlichen Schutz aus Art. 1 und 2 GG ab: „Recht auf informationelle Selbstbestimmung“.² Die Rechtsprechung des BVerfG hatte maßgeblichen Einfluss auf den Grundrechtsschutz bei der Verarbeitung personenbezogener Daten im europäischen Raum, Art. 8 I GrCh.³

Damit ist das Recht geschützt, die Kontrolle über seine Daten zu haben.⁴ Der europäische Gesetzgeber reformierte das Datenschutzrecht zum 25. Mai 2018 grundlegend. Personenbezogene Daten schützt die europäische Datenschutz-Grundverordnung (DS-GVO). Europäische Verordnungen gelten unmittelbar und verbindlich, Art. 288 II AEUV. Sie müssen nicht in nationale Gesetze umgesetzt werden.

Auch die Digital Humanities waren betroffen von dieser europäischen Kodifikation.⁵ Forschungsprojekte können unter Umständen mit Materialien arbeiten, die vom Datenschutzrecht umfasst sind. Eine gängige Arbeitsweise bei den Digitalen Humanities ist die Visualisierung komplexer Datenstrukturen. Das Projekt Mining and Modeling Text befasst sich mit der automatischen Extraktion, Strukturierung und Vernetzung von Fachinformationen aus Text- und Datensammlungen sowie mit der Nutzung solcher Informationsnetzwerke für die Beantwortung geisteswissenschaftlicher Fragestellungen. Dreh- und Angelpunkt der Arbeit ist mithin die Verarbeitung von Daten. Deswegen ist es für Forschende unumgänglich, die Gesetzeslage zu berücksichtigen. Das gilt primär,

¹ *Gola*, in: *Gola*, DS-GVO, Einl. Rn. 1.

² BVerfG NJW 1985, 419 – Volkszählungsurteil; *Gola*, in: *Gola*, DS-GVO, Einl. Rn. 2.

³ EuGH Urt. v. 13.5.2015 – C-131/12; *Gola*, in: *Gola*, DS-GVO, Einl. Rn. 3.

⁴ *Kamocki*, in: *Hawkins*, Access and control in Digital Humanities, 249, 250.

⁵ *Kamocki*, in: *Hawkins*, Access and control in Digital Humanities, 249, 250.

wenn sie mit personenbezogenen Daten arbeiten. Kann in dem Datensatz der Urheber identifiziert werden, liegt Personenbezug vor.⁶ Die Anwendung von Verfahren der Digital Humanities stellt gleichzeitig eine Verarbeitung dar, für die eine Einwilligung oder eine gesetzliche Rechtfertigung vorliegen muss.⁷ Es ergeben sich einige Abweichungen, soweit Daten für einen wissenschaftlichen Zweck verarbeitet werden. Das Datenschutzrecht privilegiert die Datenverarbeitung zu Forschungszwecken.⁸ Im Lichte der Forschungs- und Wissenschaftsfreiheit aus Art. 13 GRCh begründet die Verarbeitung ein besonderes Interesse an der Verarbeitung personenbezogener Daten.⁹ Diese sind allerdings nur anwendbar, soweit die Verarbeitung geeignete Garantien für die Rechte und Freiheiten der betroffenen Person gemäß dieser Verordnung unterliegt, vgl. Art. 89 I DS-GVO. Ob die Datenverarbeitung rechtmäßig ist, bestimmt sich oftmals an einer Interessenabwägung im Einzelfall.

I. Anwendungsbereich der DS-GVO bei den Digital Humanities

Art. 2 DS-GVO bestimmt den sachlichen Anwendungsbereich der Verordnung. Sie gilt für die ganz oder teilweise automatisierte Verarbeitung personenbezogener Daten sowie für die nichtautomatisierte Verarbeitung personenbezogener Daten, die in einem Dateisystem gespeichert sind oder gespeichert werden sollen. Hierzu beschreibt die Regelung die technischen Umstände der jeweils relevanten Datenverarbeitungsvorgänge.¹⁰

1. Darstellung des iterativen Verfahrens in den Digital Humanities

Die Digital Humanities sind ein Forschungsbereich an der Schnittstelle zwischen Geisteswissenschaften und Informatik. Die Wissenschaftlerinnen und Wissenschaftler entwickeln und nutzen Verfahren zur Identifikation, Extraktion, Analyse und Vernetzung von Informationen unterschiedlicher Textquellen.¹¹

Der Ablauf dieses iterativen Verfahrens kann in fünf Arbeitsschritte unterteilt werden: Sammeln von Texten und Daten (1), Aufarbeitung der gesammelten

⁶ *Spindler*, GRUR 2016, 1112, 1117.

⁷ *Spindler*, GRUR 2016, 1112, 1117.

⁸ *Golla/Hofmann/Bäcker*, DuD 2018, 89, 90.

⁹ *Golla/Hofmann/Bäcker*, DuD 2018, 89, 90.

¹⁰ *Ernst*, in: Paal/Pauly, DS-GVO BDSG, 3. Auflage 2021, Art. 2 Rn. 1.

¹¹ *Erler-Fridgen*, Teil 1, C. in diesem Band, 31f; für die einzelnen Verfahrensschritt vgl. 31ff.

Werke (2), Informationsextraktion (3), Präsentation der Analyseergebnisse und Text(teile) (4) sowie die Archivierung der Forschungsdaten (5).¹²

Nach diesen fünf Schritten behandelt werden insbesondere Primärquellen, bibliografische Daten und Fachliteratur.¹³

2. Räumlicher Anwendungsbereich der DS-GVO

Der räumliche Anwendungsbereich der DS-GVO ist in Art. 3 DS-GVO beschrieben. Diese Verordnung findet Anwendung auf die Verarbeitung personenbezogener Daten, soweit diese im Rahmen der Tätigkeiten einer Niederlassung eines Verantwortlichen oder eines Auftragsverarbeiters in der Union erfolgt, unabhängig davon, ob die Verarbeitung in der Union stattfindet. Da die Datenverarbeitung regelmäßig von Forschungseinrichtungen übernommen wird, die im Inland der EU arbeiten, ist die DS-GVO anwendbar.

3. Sachlicher Anwendungsbereich der DS-GVO

Der sachliche Anwendungsbereich der Datenschutz-Grundverordnung umfasst nach Art. 2 Abs. 1 DS-GVO die ganz oder teilweise automatisierte Verarbeitung personenbezogener Daten sowie die nichtautomatisierte Verarbeitung personenbezogener Daten, die in einem Dateisystem gespeichert sind oder gespeichert werden sollen. Ausschlussgründe finden sich in Art. 2 Abs. 2 DS-GVO.

a) Verarbeitung

Um den Anwendungsbereich zu eröffnen, müssen personenbezogene Daten entweder ganz oder teilweise automatisiert verarbeitet werden. In Art. 4 Nr. 2 DS-GVO ist „Verarbeitung“ definiert, als ein mit oder ohne Hilfe automatisierter Verfahren ausgeführter Vorgang oder jede solche Vorgangsreihe im Zusammenhang mit personenbezogenen Daten wie das Erheben, das Erfassen, die Organisation, das Ordnen, die Speicherung, die Anpassung oder Veränderung, das Auslesen, das Abfragen, die Verwendung, die Offenlegung durch Übermittlung, Verbreitung oder eine andere Form der Bereitstellung, den Abgleich oder die Verknüpfung, die Einschränkung, das Löschen oder die Vernichtung. Die

¹² *Erler-Fridgen*, Teil 1, C. in diesem Band, 33; *de la Durantaye/Raue*, Teil 1, A. in diesem Band, 7, 15f.

¹³ *Erler-Fridgen*, Teil 1, C. in diesem Band, 34f.

Aufzählung ist nicht abschließend, sodass jeglicher Umgang mit personenbezogenen Daten eine Verarbeitung im Sinne dieser Vorschrift darstellt.¹⁴ Die Technik der Verarbeitung ist für den Anwendungsbereich unabhängig. Soweit die Daten nicht verarbeitet werden, müssen sie in einem Dateisystem gespeichert sein oder gespeichert werden sollen.

b) Personenbezogenheit des Datums

Der zentrale Begriff der DS-GVO ist der des personenbezogenen Datums, vgl. Art. 4 Nr. 1.¹⁵ Danach sind „personenbezogene Daten“ alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann. Der sachliche Anwendungsbereich greift mithin, soweit es sich um eine identifizierte oder eine identifizierbare natürliche Person handelt.

Nach Erwägungsgrund 27 ist die Verordnung nicht auf personenbezogene Daten Verstorbener anzuwenden. Nach Satz 2 können die Mitgliedstaaten können Vorschriften für die Verarbeitung der personenbezogenen Daten Verstorbener vorsehen, was Deutschland allerdings nicht gemacht hat.

Personenbezogene Daten liegen zunächst vor, soweit sie Informationen über eine identifizierte, d.h. konkret benannte Person enthalten.¹⁶ Art. 4 Nr. 1 prägt ein sehr weites Verständnis personenbezogener Daten und umfasst alle Informationen mit Personenbezug.¹⁷ Das ist zu bejahen, soweit sich die natürliche Person ohne weiteres von einer anderen Person oder einer Gruppe unterscheiden lässt.¹⁸ Personenbezug beinhaltet Informationen, die beispielsweise den Doktorgrad einer Person wiedergeben.

Personenbezogene Daten sind nicht gegeben, soweit Daten anonymisiert sind. Anonymisieren bedeutet, dass Daten so verarbeitet sind, dass das Subjekt nicht

¹⁴ Eßer, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 4 Nr. 32.

¹⁵ Ernst, in: Paal/Pauly, DS-GVO BDSG, 3. Auflage 2021, Art. 4 Rn. 3.

¹⁶ Gola, in: Gola/Heckmann, Datenschutz-Grundverordnung – Bundesdatenschutzgesetz, 3. Auflage 2022, Art. 4 Rn. 7.

¹⁷ Eßer, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 4 Nr. 7.

¹⁸ Artikel-29-Datenschutzgruppe, WP 136, 14.

mehr identifiziert werden kann.¹⁹ Der Standard ist hierbei sehr hoch. Die Anonymisierung muss irreversibel sein, sodass sie in den meisten Fällen nicht in Betracht kommt.

II. Abweichungen zu wissenschaftlichen Zwecken (Öffnungsklauseln)

Für die Datenverarbeitung zu wissenschaftlichen Zwecken bietet die DSGVO Öffnungsklauseln. Öffnungsklauseln sind Klauseln, die eine Abweichung von Grundsätzen im Gesetz unter bestimmten Zwecken ermöglichen.

In der DS-GVO ergeben sich zwei Öffnungsklauseln zu wissenschaftlichen Zwecken. Nach Art. 85 II DS-GVO sind für die Verarbeitung von wissenschaftlichen Zwecken Abweichungen und Ausnahmen vorzusehen. Eine solche findet sich auch in Art. 89 II DS-GVO zu wissenschaftlichen Forschungszwecken. Das Verhältnis der beiden Normen zueinander ist unklar.²⁰ Erstere bietet bis auf Sanktionen in Kapiteln VIII. der Datenschutz-Grundverordnung Abweichungen von allen verbleibenden Kapiteln, während Art. 89 Abs. 2 DS-GVO dem Wortlaut nach nur Abweichungen von Art. 15, 16, 18 und 21 DS-GVO zulässt.²¹

Zur Abgrenzung der beiden Normen kann nicht der deutsche Wortlaut herangezogen werden. Vergleicht man den Wortlaut anderer Sprachfassungen, geht daraus hervor, dass sich der Anwendungsbereich des Art. 85 II DS-GVO auf die Wissenschaftskommunikation, d.h. die Publikation von Forschungsergebnissen beschränkt.²²

Die DS-GVO enthält keine Definition des Begriffs „wissenschaftliche Forschung“. Nach Erwägungsgrund 159 S. 2 und 3 ist der Begriff weit auszulegen. Die Verarbeitung personenbezogener Daten zu wissenschaftlichen Forschungszwecken im Sinne dieser Verordnung sollte weit ausgelegt werden und die Verarbeitung für beispielsweise die technologische Entwicklung und die Demonst-

¹⁹ *Gola*, in: *Gola*, DS-GVO, Art. 4 Rn. 41.

²⁰ *Pötters*, in: *Gola/Heckmann*, Datenschutz-Grundverordnung – Bundesdatenschutzgesetz, 3. Auflage 2022, Art. 85 Rn. 13.

²¹ *Lauber-Rönsberg*, in: *BeckOK Datenschutzrecht*, Wolff/Brink, 42. Edition, Stand: 01.11.2022, Art. 85 Rn. 24.

²² *Lauber-Rönsberg*, in: *BeckOK Datenschutzrecht*, Wolff/Brink, 42. Edition, Stand: 01.11.2022, Art. 85 Rn. 25.

ration, die Grundlagenforschung, die angewandte Forschung und die privat finanzierte Forschung einschließen.²³ Satz 3 des Erwägungsgrundes verweist darüber hinaus auf die in Artikel 179 Absatz 1 AEUV festgeschriebenen Ziel, einen europäischen Raum der Forschung zu schaffen. Eine Verarbeitung zu Forschungszwecken ist von der Unterrichtung des aktuellen Standes der Wissenschaft bis zur Anwendung der Forschung anzunehmen, wenn sie auf die Gewinnung neuartiger Erkenntnisse ausgerichtet ist.²⁴

III. Öffnungsklausel aus dem Landesdatenschutzgesetz Rheinland-Pfalz

Die DS-GVO enthält Öffnungsklauseln, die ergänzende oder einschränkende Regelungen zum Datenschutz treffen. Jene Öffnungsklauseln erlauben dem nationalen Bundes- und Landesgesetzgeber, im Rahmen der vorgegebenen Möglichkeiten Sonderregelungen zu treffen. Ein Beispiel dafür ist die Datenverarbeitung zur Erfüllung gesetzlicher Pflichten und öffentlicher Aufgaben (Art. 6 II und III DSGVO) und im Zusammenhang mit der wissenschaftlichen Forschung (vgl. etwa Art. 9 II lit. j DSGVO und Art. 89 II DSGVO). Solche Regelungen finden sich in den Bundes- und Landesdatenschutzgesetzen. Das BDSG gilt grundsätzlich für öffentliche Stellen und für nicht-öffentliche Stellen des Bundes (§ 1 II BDSG). Die Landesdatenschutzgesetze gelten für öffentliche Stellen der Länder. Allerdings sind die Landesdatenschutzgesetze an die Erlaubnisatbestände der Bundesgesetze angelehnt.²⁵

Soweit Forschung und Lehre an Einrichtungen und sonstigen öffentlichen Stellen des Landes betrieben werden, ist das Landesdatenschutzgesetz des jeweiligen Landes zu beachten. Für Rheinland-Pfalz beispielsweise ist die Öffnungsklausel aus § 2 I Nr. 3 LDSG Rheinland-Pfalz maßgeblich. Im weiteren Verlauf konzentrieren sich die Ausführungen auf die DS-GVO. Soweit einschlägig, verweist die Handreichung auf die Norm aus dem rheinland-pfälzischen Landesdatenschutzgesetz.

²³ Golla, in: Specht/Mantz, Handbuch europäisches und deutsches Datenschutzrecht, § 23 Rn. 14.

²⁴ Golla, in: Specht/Mantz, Handbuch europäisches und deutsches Datenschutzrecht, § 23 Rn. 15.

²⁵ Golla, in: Specht/Mantz, Handbuch europäisches und deutsches Datenschutzrecht, § 23 Rn. 12.

IV. Grundsätze für die Verarbeitung personenbezogener Daten, Art. 5 DS-GVO

Art. 5 und Art. 6 der DS-GVO bilden sie „Magna Charta der zulässigen Datenverarbeitung“.²⁶ Dabei unterscheidet sich Art. 5 DS-GVO von Art. 6 DS-GVO, da dieser allgemeine Prinzipien zulässiger Datenverarbeitung aufstellt. Letzterer bietet Erlaubnistatbestände, da der Datenschutz in Form eines Verbotsprinzips mit Erlaubnisvorbehalt ausgestaltet ist.²⁷

Die Prinzipien sind wie folgt aufgeführt:

- (1) Rechtmäßigkeit, Art. 5 I a) DS-GVO
- (2) Verarbeitung nach Treu und Glauben, Art. 5 I a) DS-GVO
- (3) Transparenz, Art. 5 I a) DS-GVO
- (4) Zweckbindung, Art. 5 I b) DS-GVO
- (5) Datenminimierung, Art. 5 I c) DS-GVO
- (6) Richtigkeit, Art. 5 I d) DS-GVO
- (7) Speicherbegrenzung, Art. 5 I e) DS-GVO
- (8) Integrität und Vertraulichkeit, Art. 5 I f) DS-GVO
- (9) Rechenschaftspflicht, Art. 5 II DS-GVO

Im weiteren Verlauf werden die Zweckbindung und die Datenminimierung dargestellt.

1. Zweckbindung, Art. 5 I b) DS-GVO

a) Allgemein

Die Zweckbindung wird auch als „Grundstein des Datenschutzes“ bezeichnet.²⁸ Sie hat zwei Ausprägungen. Einerseits die Zweckfestlegung, wonach eine Verarbeitung nur für festgelegte, eindeutige und legitime Zwecke erfolgen darf.²⁹ Andererseits die Zweckbindung im engeren Sinne, die eine Verarbeitung

²⁶ *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 5 Rn. 1.

²⁷ *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 6 Rn. 1.

²⁸ *Schantz*, in: BeckOK Datenschutzrecht, Wolff/Brink, 42. Edition, Stand: 01.11.2021, Art. 5 Rn. 12.

²⁹ *Schantz*, in: BeckOK Datenschutzrecht, Wolff/Brink, 42. Edition, Stand: 01.11.2021, Art. 5 Rn. 12.

personenbezogener Daten verbietet, soweit sie mit dem Erhebungszweck unvereinbar sind.³⁰

b) Datenschutz durch Voreinstellung, Art. 25 II DS-GVO

Art. 25 II DS-GVO steht in einem abgestuften Verhältnis zu seinem Absatz 1. Sie ermöglicht den Datenschutz durch Voreinstellung (data protection by default). Die Möglichkeit Voreinstellungen zu treffen, ist abhängig von dem Programm bzw. der Dienstleitung.³¹ Hiernach sollen die Voreinstellung im technischen Verfahren nur die personenbezogenen Daten verarbeiten, deren Verarbeitung für den jeweiligen bestimmten Verarbeitungszweck erforderlich sind. Die Auswahl der Voreinstellung ist an den jeweiligen Verarbeitungszweck gebunden.³²

2. Datenminimierung

a) Allgemeiner Grundsatz, Art. 5 I c) DS-GVO

Die Datenminimierung ist kein absolutes Gebot, sondern vielmehr ein allgemeiner Grundsatz.³³ Hiernach muss jede Datenverarbeitung angemessen im Verhältnis zum Zweck (Angemessenheit) und auf das notwendige Maß beschränkt sein (Erforderlichkeit).³⁴ Damit bedient sich die Abwägung dem Verhältnismäßigkeitsgrundsatz.

b) Datenschutz durch Technikgestaltung, Art. 25 I DS-GVO

Nach Art. 24 I DS-GVO hat der Verantwortliche unter Berücksichtigung der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung sowie der unterschiedlichen Eintrittswahrscheinlichkeit und Schwere der Risiken für die Rechte und Freiheiten der natürlichen Person geeignete technische und organisatorische Maßnahmen umzusetzen. Art. 25 DS-GVO formuliert zwei Möglichkeiten und ergänzt so den Grundsatz der Datenminimierung.

³⁰ *Schantz*, in: BeckOK Datenschutzrecht, Wolff/Brink, 42. Edition, Stand: 01.11.2021, Art. 5 Rn. 12.

³¹ *Brüggemann*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 25 Rn. 23.

³² *Brüggemann*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 25 Rn. 24.

³³ *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 5 Rn. 25.

³⁴ *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 5 Rn. 25.

Hiernach kann der Datenschutz auch durch Technikgestaltung (data protection by design) oder durch datenschutzfreundliche Voreinstellungen (data protection by default) erreicht werden.³⁵

Nach Art. 25 I DS-GVO trifft der Verantwortliche unter Berücksichtigung des Standes der Technik geeignete technische und organisatorische Maßnahmen – wie z. B. Pseudonymisierung –, die dafür ausgelegt sind, die Datenschutzgrundsätze wie etwa Datenminimierung wirksam umzusetzen und die notwendigen Garantien in die Verarbeitung aufzunehmen, um den Anforderungen dieser Verordnung zu genügen und die Rechte der betroffenen Personen zu schützen. Dieser Regelung liegt die Vorstellung zugrunde, dass Technik und Datenschutz komplementär zueinander sind.³⁶ Sie ermöglicht es, eine Abwägung zu treffen.³⁷ Ein Mehr an technischen Sicherheitsvorkehrungen, beeinflusst den rechtlichen Spielraum bei der Beurteilung der Zulässigkeit der Datenverarbeitung.³⁸ Die Grenze dieses Wechselspiels bildet immer das Grundrecht auf informationelle Selbstbestimmung der Betroffenen.

3. Sonderregelungen für die wissenschaftliche Forschung, Art. 5 I DS-GVO

Art. 5 I lit. b und lit. e DS-GVO regeln Einschränkungen zu den Grundsätzen für die Verarbeitung personenbezogener Daten, soweit sich die Verarbeitung zum Zwecke der wissenschaftlichen Forschung ergibt. Art. 5 I lit. b DS-GVO schränkt den Grundsatz der Zweckvereinbarung ein. Damit dürfen Daten weiterverarbeitet werden, auch wenn sie ursprünglich nicht für wissenschaftliche oder historische Forschungszwecke gespeichert wurden. Ihre Wirkung wird als Fiktion beschrieben.³⁹ Der Datenverarbeiter darf sich im Ergebnis auf den gleichen Erlaubnistatbestand stützen, für den die ursprüngliche Verarbeitung erfolgte.⁴⁰

³⁵ *Brüggemann*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 25 Rn. 1.

³⁶ *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 6 Rn. 1; *Hornung*, ZD 2011, 51 f.

³⁷ *Paulus*, in: BeckOK Datenschutzrecht, Wolff/Brink, 42. Edition, Stand: 01.11.2021, Art. 25 Rn. 4.

³⁸ *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 6 Rn. 1.

³⁹ *Golla*, in: Specht/Mantz, Handbuch europäisches und deutsches Datenschutzrecht, § 23 Rn. 56.

⁴⁰ *Golla*, in: Specht/Mantz, Handbuch europäisches und deutsches Datenschutzrecht, § 23 Rn. 56.

Art. 5 I lit. e Hs. 2 DS-GVO ermöglicht eine Ausnahme vom Grundsatz der Speicherbegrenzung. Es gilt jedoch weiterhin das Erfordernis nach Art. 89 I DS-GVO, wonach bei der Datenverarbeitung zu Forschungszwecken geeignete technische und organisatorische Maßnahmen zur Datenminimierung ergriffen werden müssen.⁴¹ Die Ausnahme ermöglicht lediglich einen Mehrwert, wenn zu dem ursprünglich erfolgten Zweck eine Datenverarbeitung zu Forschungszwecken hinzukommt, die eine längere Speicherung erfordert.⁴²

V. Rechtmäßige Datenverarbeitung personenbezogener Daten, Art. 6 DS-GVO

Art. 6 I DS-GVO nennt Erlaubnistatbestände, die eine Verarbeitung von personenbezogenen Daten zulassen. Für Daten von Kindern und sensitive Daten führen Art. 8 f. DS-GVO höhere Anforderungen heran. Jede Datenverarbeitung setzt mithin eine der folgenden Erlaubnistatbestände voraus und setzt damit der Privatautonomie Grenzen.⁴³

- (1) Einwilligung, Art. 6 I 1 lit. a DS-GVO
- (2) durch Vertrag, Art. 6 I 1 lit. b DS-GVO
- (3) zur Erfüllung einer rechtlichen Verpflichtung, Art. 6 I 1 lit. c DS-GVO
- (4) Schutz lebenswichtiger Interessen, Art. 6 I 1 lit. d DS-GVO
- (5) erforderlich für die Wahrnehmung einer öffentlichen Aufgabe, Art. 6 I 1 lit. e DS-GVO
- (6) Interessenabwägung, Art. 6 I 1 lit. f DS-GVO

Im Rahmen der Verarbeitung von Daten in Projekten der Digital Humanities kommen die Einwilligung oder die Interessenabwägung in Frage.

1. Einwilligung, Art. 6 I 1 lit. A DS-GVO

Die Einwilligung ist Ausdruck der datenschutzrechtlichen Selbstbestimmung jedes Einzelnen.⁴⁴ Die Einwilligung ist in Art. 4 Nr. 11 DS-GVO als jede freiwillige, für den bestimmten Fall, in informierter Weise und unmissverständ-

⁴¹ *Golla*, in: Specht/Mantz, Handbuch europäisches und deutsches Datenschutzrecht, § 23 Rn. 59.

⁴² *Golla*, in: Specht/Mantz, Handbuch europäisches und deutsches Datenschutzrecht, § 23 Rn. 59.

⁴³ *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 6 Rn. 1.

⁴⁴ *Schulz*, in: Gola, DS-GVO, Art. 6 Rn. 21.

lich abgegebene Willensbekundung in Form einer Erklärung oder einer sonstigen eindeutigen bestätigenden Handlung, definiert. Hiernach gibt die betroffene Person zu verstehen, dass sie mit der Verarbeitung der sie betreffenden personenbezogenen Daten einverstanden ist. Art. 7 DS-GVO konkretisiert die aufgestellten Voraussetzungen.⁴⁵

Eine Einwilligung kommt in der Praxis allerdings selten vor.⁴⁶ Die betroffene Person müsste den Zweck der Datenverarbeitung kennen und darin eingewilligt haben. Die Einwilligung muss freiwillig erfolgen, die betroffene Person muss also eine echte und freie Wahl haben. Des Weiteren ist in Art. 7 IV DS-GVO ein Kopplungsverbot verankert. Hiernach ist eine Einwilligung nur wirksam, wenn sie ohne jeden Druck oder Zwang abgegeben wurde.

Zudem kann die Einwilligung jederzeit widerrufen werden. Danach kann der Datenverarbeiter nicht einfach den Erlaubnistatbestand wechseln, sondern muss die betroffene Person informieren.⁴⁷ Da für die Verarbeitung oftmals eine Vielzahl an Daten erforderlich ist, erscheint die Einwilligung nicht praxistauglich.

In Art. 5 I lit. b DS-GVO ist eine Öffnungsklausel im Sinne des Art. 89 I DS-GVO normiert. Die Verarbeitung von personenbezogenen Daten bedarf entweder einer Einwilligung oder eines gesetzlichen Rechtsfertigungstatbestand, vgl. Art. 6 I 1 DS-GVO. Die Einwilligung hat hohe Hürden und ist zweckgebunden, vgl. Art. 6 I 1 lit. a DS-GVO iVm Art. 5 I lit. b DS-GVO. Der Zweckbindungsgrundsatz erschwert die Verarbeitung von Big Data. Um das für Forschungsvorhaben abzumildern, durchbricht Art. 5 I lit. b DS-GVO den Zweckbindungsgrundsatz bei Weiterverarbeitungen zu wissenschaftlichen Zwecken.⁴⁸ Soweit also eine Weiterverarbeitung wissenschaftlichen Zwecken dient, dürfen die Daten zu weiterverarbeitet werden. Daraus folgt, dass die Datenverarbeitung auch rechtmäßig ist, soweit eine Einwilligung vorliegt, die für einen anderen Zweck erhoben wurden.

2. Wahrnehmung einer öffentlichen Aufgabe, Art. 6 I 1 lit. e DS-GVO

Nach Art. 6 I 1 lit. e DS-GVO ist die Verarbeitung rechtmäßig, soweit sie für die Wahrnehmung einer Aufgabe erforderlich ist, die im öffentlichen Interesse liegt oder in Ausübung öffentlicher Gewalt erfolgt, die dem Verantwortlichen

⁴⁵ *Schulz*, in: Gola, DS-GVO, Art. 7 Rn. 1.

⁴⁶ *Spindler*, GRUR 2016, 1112, 1117.

⁴⁷ *Kamocki*, in: Hawkins, Access and control in Digital Humanities, 249, 260.

⁴⁸ *Spindler*, GRUR 2016, 1112, 1117.

übertragen wurde. Wie aus Erwägungsgrund 45 S. 1 hervorgeht, begründet die Vorschrift keine Ausnahmeregelung für sich alleine gesehen, sondern bietet eine Öffnungsklausel. Mit § 3 BDSG hat der deutsche Gesetzgeber und mit § 3 LDSG der Landesgesetzgeber eine solche Regelung geschaffen. Nach § 3 BDSG ist die Verarbeitung personenbezogener Daten durch eine öffentliche Stelle zulässig, wenn sie zur Erfüllung der in der Zuständigkeit des Verantwortlichen liegenden Aufgabe oder in Ausübung öffentlicher Gewalt, die dem Verantwortlichen übertragen wurde, erforderlich ist. Nach § 3 LDSG ist fast wortlautgenau die Verarbeitung personenbezogener Daten durch eine öffentliche Stelle unbeschadet anderer Rechtsgrundlagen zulässig, wenn sie zur Erfüllung einer im öffentlichen Interesse liegenden Aufgabe oder in Ausübung öffentlicher Gewalt, die dem Verantwortlichen übertragen wurde, erforderlich ist. Die Forschung kann, muss jedoch nicht im öffentlichen Interesse liegen.⁴⁹ Das hängt davon ab, ob der Europäische Gerichtshof wissenschaftliche Forschung und wissenschaftlicher Fortschritt nach Art. 3 III UA 1 3 EUV und Art. 179 I AEUV als Allgemeinwohlziele anerkennt.⁵⁰ Damit wäre Art. 6 Abs. 1 lit. e DS-GVO i.V.m. § 3 LDSG eine Öffnungsklausel für die Datenverarbeitung zu Forschungszwecken durch öffentlich-rechtliche Forschungseinrichtungen wie Universitäten.⁵¹

3. Interessenabwägung, Art. 6 I 1 lit. f DS-GVO

Die Interessenabwägung nach Art. 6 I 1 lit. f DS-GVO ist die zentrale Erlaubnisnorm, um personenbezogene Daten zu verarbeiten.⁵² Hiernach ist die Verarbeitung rechtmäßig, wenn sie zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich ist. Das gilt, sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz personenbezogener Daten erfordern, überwiegen, insbesondere dann, wenn es sich bei der betroffenen Person um ein Kind handelt. In die Abwägung fließt also das berechnete Interesse des Verantwortlichen und der betroffenen Person hinein. Es genügt jedes wirtschaftliche oder ideelle Interesse.⁵³ Berechtigt ist jedes Interesse, welches nicht gegen die Rechtsordnung verstößt.⁵⁴ Die wissenschaftliche Forschung ist so ein berechtigtes Interesse.⁵⁵

⁴⁹ *Beyvers/Gärtner/Kipker*, PinG 2015, 241, 244.

⁵⁰ *Golla/Hofmann/Bäcker*, DuD 2018, 89, 93.

⁵¹ *Golla/Hofmann/Bäcker*, DuD 2018, 89, 93.

⁵² *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 6 Rn. 41.

⁵³ *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 6 Rn. 46.

⁵⁴ *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 6 Rn. 49.

⁵⁵ *Golla/Hofmann/Bäcker*, DuD 2018, 89, 93.

Da eine breite Vorschrift für Rechtsunsicherheit sorgt, muss der Verantwortliche für jedes Datenverarbeitungsverfahren das angenommene berechtigte Interesse dokumentieren, Art. 13 I lit. d, 14 II lit. b DS-GVO.⁵⁶

4. Abgrenzung zwischen öffentlichen Stellen und der Interessenabwägung

Die Abgrenzung zwischen den beiden Tatbeständen erfolgt im Schwerpunkt aus der Organisation der jeweiligen Forschungseinrichtung.⁵⁷ Denn Behörden im Sinne der DS-GVO können sich nicht auf die Interessenabwägung aus lit. f stützen. Daraus ist zu folgern, dass öffentlich-rechtlich organisierte Hochschulen und andere Forschungseinrichtungen sich für die Datenverarbeitung zu Forschungszwecken auf Art. 6 Abs. lit. e DS-GVO i.V.m. § 3 BDSG stützen müssen.⁵⁸ Soweit es privat organisierte Forscherinnen und Forscher sind, können sie weiterhin ihren Erlaubnistatbestand auf die Interessenabwägung aus Art. 6 Abs. lit. f DS-GVO beziehen.

VI. Öffnungsklausel zu Forschungszwecken bei sensiblen Daten, Art. 9 II lit. j DS-GVO iVm § 22 I LDSG (§ 27 I 1 BDSG)

Die Verarbeitung personenbezogener Daten, aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person ist nach Art. 9 I DS-GVO untersagt. Eine Ausnahme reguliert Art. 9 II DS-GVO. Neben den Erlaubnistatbeständen aus Art. 6 DS-GVO ff. ermöglicht Art. 9 III lit. j DS-GVO Öffnungsklausel für die Mitgliedsstaaten. Der deutsche Gesetzgeber hat mit § 27 BDSG und der Landesgesetzgeber in § 22 I davon Gebrauch gemacht. Nach § 22 I LDSG ist die Verarbeitung von Daten auch ohne Einwilligung für wissenschaftliche oder historische Forschungszwecke oder für statistische Zwecke zulässig, wenn die Verarbeitung zu diesen Zwecken erforderlich ist und die Interessen des Verantwortlichen an der Verarbeitung die Interessen der betroffenen Person an einem Ausschluss der Verarbeitung erheblich überwiegen. § 27 I 1 BDSG ist gleich im Wortlaut, sodass angenommen werden kann, dass

⁵⁶ Kramer, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 6 Rn. 44.

⁵⁷ Golla/Hofmann/Bäcker, DuD 2018, 89, 93.

⁵⁸ Golla/Hofmann/Bäcker, DuD 2018, 89, 93.

der Landesgesetzgeber keine Änderungen zu den Ausführungen im Bundesgesetz vornehmen wollte.

Hierfür muss die Verarbeitung drei Voraussetzungen erfüllen. Zuerst muss der Zweck der Datenverarbeitung die wissenschaftliche Forschung sein. Damit ist ein eigenes „konkretes Forschungsvorhaben [gemeint], das seinem ganzen Aufbau und Inhalt nach wissenschaftlichen Ansprüchen genügt“. ⁵⁹ Des Weiteren muss die Datenverarbeitung erforderlich sein. Diese Voraussetzung ist gegeben, soweit das Vorhaben ohne die Verarbeitung der konkreten personenbezogenen Daten undurchführbar ist. ⁶⁰ Zuletzt müssen die beiden konkurrierenden Interessen gegeneinander abgewogen werden. Hierbei ist das wissenschaftliche Interesse und der Schutz der eigenen Daten relevant. Bei Vorliegen der Voraussetzungen aus § 27 I 1 BDSG ist stets der Erlaubnistatbestand der Interessenabwägung aus Art. 6 Abs. 1 lit. f DS-GVO gegeben, da das BDSG strengere Anforderungen aufweist. ⁶¹

Aus § 22 II LDSG ergibt sich, dass die Daten für weitere, mit dem ursprünglichen Zweck vereinbare Zwecke der Forschung verarbeitet werden dürfen.

Zuletzt haben die Forschenden nach § 27 I 1 BDSG, angemessene und spezifische Maßnahmen zur Wahrung der Interessen der betroffenen Person gemäß § 22 II 2 BDSG vorzunehmen.

VII. Informationspflichten, Art. 13 f. DS-GVO

Art. 13 und Art. 14 DS-GVO legen dem Verantwortlichen Informationspflichten gegenüber den Betroffenen der Datenverarbeitung auf. ⁶² Damit bilden die Vorschriften die Grundlage für die Ausübung der Betroffenenrechte. ⁶³ Denn erst, wenn eine betroffene Person hinreichende Informationen über die Datenverarbeitung hat, kann sie ihre Rechte wie den Auskunftsanspruch nach Art. 15 DS-GVO ausüben. ⁶⁴ Sie setzen mithin den allgemeinen Grundsatz der Transparenz aus Art. 5 I a) DS-GVO um. ⁶⁵

⁵⁹ *Simitis*, in: *Simitis*, BDSG, 8. Auflage 2014, § 28 Rn. 310.

⁶⁰ *Simitis*, in: *Simitis*, BDSG, 8. Auflage 2014, § 28 Rn. 310; *Golla/Hofmann/Bäcker*, DuD 2018, 89, 93.

⁶¹ *Golla/Hofmann/Bäcker*, DuD 2018, 89, 93.

⁶² *Eßer*, in: *Auernhammer*, DSGVO BDSG, 6. Auflage 2018, Art. 13 Rn. 1.

⁶³ *Schmidt-Wudy*, in: *BeckOK Datenschutzrecht*, Wolff/Brink, 42. Edition, Stand: 01.11.2022, Art. 13 Rn. 2.

⁶⁴ *Schmidt-Wudy*, in: *BeckOK Datenschutzrecht*, Wolff/Brink, 42. Edition, Stand: 01.11.2022, Art. 13 Rn. 2.

⁶⁵ *Eßer*, in: *Auernhammer*, DSGVO BDSG, 6. Auflage 2018, Art. 14 Rn. 2.

Art. 13 DS-GVO regelt alle Fälle, in denen Verantwortliche die personenbezogenen Daten direkt bei den Betroffenen erhoben haben. Art. 14 DS-GVO ist anzuwenden, soweit die Erhebung bei einem Dritten und nicht dem Betroffenen selbst erfolgt. Letzteres hat in diesem Kontext Vorrang. In den Absätzen 1 – 3 zählt die Vorschrift enumerativ alle Angaben auf, die sich nach dem Absatz 4 auch auf die Weiterverarbeitung beziehen müssen.

Art. 14 V lit. b DS-GVO enthält allerdings eine Ausnahmvorschrift für wissenschaftliche oder historische Forschungszwecke. Hiernach entfällt die Informationspflicht, soweit die Erteilung dieser Informationen sich als unmöglich erweist oder einen unverhältnismäßigen Aufwand erfordern würde. Davon ist bei kleineren Projekten in der Regel auszugehen.

VIII. Zusammenfassung

Die DS-GVO bildet eine sogenannte „Magna Charta der zulässigen Datenverarbeitung“.⁶⁶ Sie hat den Schutz personenbezogener Daten zum Ziel. Der Anwendungsbereich ist weit, sodass praktisch jede Verarbeitung den Anforderungen der DS-GVO unterliegt.

Die Digital Humanities müssen die Regelungen beachten, soweit sie mit einer großen Datenmenge arbeiten, da diese oftmals personenbezogene Daten beinhaltet. Hierbei sind insbesondere die Grundsätze für die Verarbeitung personenbezogener Daten nach Art. 5 DS-GVO einzuhalten. Beispielsweise müssen die Daten auf rechtmäßige Weise verarbeitet werden, sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein. Zudem müssen die Daten sachlich richtig sein. Auch müssen sie in einer Weise verarbeitet werden, die eine angemessene Sicherheit der personenbezogenen Daten gewährleistet, einschließlich Schutz vor unbefugter oder unrechtmäßiger Verarbeitung und vor unbeabsichtigtem Verlust, unbeabsichtigter Zerstörung oder unbeabsichtigter Schädigung durch geeignete technische und organisatorische Maßnahmen. Dabei sind die Grundsätze nach Art. 5 I DS-GVO, die regelmäßig dadurch erfüllt sind, soweit sich die Digital Humanities an die Grundsätze guter wissenschaftlicher Praxis halten.

Die Verarbeitung ist als Verbotsnorm mit Erlaubnisvorbehalt ausgestaltet, vgl. Art. 6 I 1 DS-GVO. Dabei ist der sicherste Erlaubnistatbestand derjenige der Interessenabwägung. So streng die Verordnung auch geregelt ist, sie soll einen

⁶⁶ *Kramer*, in: Auernhammer, DSGVO BDSG, 6. Auflage 2018, Art. 5 Rn. 1.

Ausgleich zwischen dem Schutz personenbezogener Daten und dem freien Verkehr solcher Daten gewährleisten, vgl. Art. 1 I DS-GVO. Die Interessenabwägung wird mithin regelmäßig zugunsten der Datenverarbeiter zu wissenschaftlichen Zwecken ausfallen, soweit sie sich an die Grundsätze aus Art. 5 I DS-GVO halten.

Hinsichtlich der Datenverarbeitung zu wissenschaftlichen Zwecken ist die Öffnungsklausel nach Art. 85 und Art. 89 DS-GVO zu beachten. Wissenschaftliche Forschungszwecke ist „jede wissenschaftliche Tätigkeit, das heißt [...] alles, was nach Inhalt und Form als ernsthafter planmäßiger Versuch zur Ermittlung von Wahrheit anzusehen ist.“⁶⁷ Durch die Öffnungsklausel ergeben sich Ausnahmenvorschriften, die im ganzen Gesetz gestreut sind.

Leseempfehlung zur vertiefenden Lektüre: *Kamocki*, in: Hawkins, *Access and Control in Digital Humanities*, 249.

⁶⁷ BVerfG, NJW 1973, 1176.

Verzeichnis der Autorinnen und Autoren

Karolina Benedyk

war wissenschaftliche Mitarbeiterin am Institut für Recht und Digitalisierung Trier (IRDT) bei Prof. Dr. Benjamin Raue (Universität Trier) und arbeitete im interdisziplinären Forschungsprojekt Mining and Modeling Text (MiMoText, Universität Trier). Sie ist Volontärin beim Weser Kurier

Frédéric Döhl

ist Strategiereferent der Generaldirektion der Deutschen Nationalbibliothek in Leipzig und Frankfurt am Main, sowie Privatdozent für Musikwissenschaft am Institut für Theaterwissenschaft der Freien Universität Berlin

Katharina de la Durantaye

ist Inhaberin der Professur für Bürgerliches Recht und Recht der Digitalisierung an der Humboldt-Universität Berlin

Katharina Erler-Fridgen

war wissenschaftliche Mitarbeiterin am Institut für Recht und Digitalisierung Trier (IRDT) bei Prof. Dr. Benjamin Raue (Universität Trier) und arbeitete im interdisziplinären Forschungsprojekt Mining and Modeling Text (MiMoText, Universität Trier). Sie arbeitet jetzt als Communications Officer bei der Universität Luxemburg

Evelyn Gius

ist Professorin für Digital Philology und Neuere deutsche Literaturwissenschaft am Institut für Sprach- und Literaturwissenschaft der Technischen Universität Darmstadt

Karina Grisse

ist Wissenschaftliche Mitarbeiterin und Habilitandin am Institut für Medienrecht und Kommunikationsrecht der Universität zu Köln

Maria Hinzmann

ist wissenschaftliche Mitarbeiterin am Trier Center for Digital Humanities (TCDH) und war Koordinatorin des interdisziplinären Forschungsprojekts Mining and Modeling Text (MiMoText, Universität Trier)

Fotis Jannidis

ist Inhaber des Lehrstuhls für Computerphilologie und Neuere Deutsche Literaturgeschichte an der Universität Würzburg

Florian Jotzo

ist Leiter des Referats für Digitalisierung und Datenschutz des Ministeriums für Inneres, Kommunales, Wohnen und Sport Schleswig-Holstein

Peter Leinen

ist Leiter des Fachbereichs Informationsinfrastruktur bei der Deutschen Nationalbibliothek

Benjamin Raue

ist Inhaber der Professur für Zivilrecht, Recht der Informationsgesellschaft und des Geistigen Eigentums sowie Direktor des Instituts für Recht und Digitalisierung (IRDT) an der Universität Trier

Achim Rettinger

ist Inhaber der Professur für Computerlinguistik an der Universität Trier

Jörg Röpke

ist Leiter der Abteilung Informationstechnologie, Forschungs- und Publikationsdienste der Universitätsbibliothek Trier

Christof Schöch

ist Professor für Digital Humanities an der Universität Trier und wissenschaftlicher Ko-Direktor des Trier Center for Digital Humanities (TCDH)

Peer Trilcke

ist Inhaber der Professur für deutsche Literatur des 19. Jahrhunderts mit dem Schwerpunkt Theodor Fontane an der Universität Potsdam, Leiter des Theodor-Fontane-Archivs an der Universität Potsdam sowie Sprecher des Potsdamer „Netzwerks für Digitale Geisteswissenschaften“

Benjamin Raue & Christof Schöch (Hrsg.)

Recht und Digital Humanities

Dieser Sammelband führt insbesondere die Beiträge zusammen, die aus dem Projekt „Mining and Modeling Text (MiMoText)“ (2019-2023) hervorgegangen sind, das den Bereich der quantitativen Methoden zur Extraktion, Modellierung und Analyse geisteswissenschaftlich relevanter Informationen aus umfangreichen Textsammlungen weiterentwickelt und aus interdisziplinärer (geistes-, informatik- und rechtswissenschaftlicher) Perspektive erforscht hat.