Dr. Karina Grisse*

Nutzbarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte**

– Rechtliche Bedingungen und Möglichkeiten –

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/EU[3] (UrhR-RL) enthält in den Art. 3

* Dr. Karina Grisse, LL.M. (Edinburgh) ist Wissenschaftliche Mitarbeiterin und Habilitandin am Institut für Medienrecht und Kommunikationsrecht der Universität zu Köln, bei Prof. Dr. *Karl-Nikolaus Peifer*.

** Dieser Beitrag ist im Zusammenhang mit dem von der DFG geförderten Workshop „Strategien für die Nutzbarmachung urheberrechtlich geschützter Textbestände für die Forschung durch Dritte“ am 28.11.2019 und 17.01.2020 in Trier entstanden. Er wird ergänzt durch die Beiträge von *Jotzo* und *Schöch et al*, in diesem Heft (RuZ 2020, 128 und 160). Die Autorin dankt den Teilnehmenden des Workshops für die spannende Diskussion des Themas. Der Dank gilt insbesondere Dr. *Florian Jotzo*, Prof. Dr. *Benjamin Raue* und Prof. Dr. *Christof Schöch* für den wertvollen Gedankenaustausch.

1 Für Anwendungsbeispiele siehe: *Schöch et al*, RuZ 2020, 160, 162 ff.; *Spindler*, GRUR 2016, 1112, 1113; *Raue*, GRUR 2017, 11, 12 f.

2 Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft – UrhWissG, BGBl 2017 Teil I Nr. 61, S. 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.

 DOI: 10.5771/2699-1284-2020-2-143

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

und Art. 4 ebenfalls zwei Schrankenregelungen zu Gunsten 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 Verfügbarmachung blieben selbst dann unklar. Die neuen Richtlinienbestimmungen dürften an diesem Ergebnis und den bestehenden Unsicherheiten[10] nichts ändern. Jedenfalls außerhalb der speziellen Regelung des Art. 3 UrhR-RL für nicht kommerzielle Forschungsorganisationen ist die Aufbewahrung und weitere Nutzung der zum Zwe-

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; zum 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 BMJW, 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.

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.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

cke 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.

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 Tex-

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*, ZfdG 2020.

13 *Schöch et al*, RuZ 2020, 160, 162 ff.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

te.[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. Wiederum 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 Schutzumfangs der Primärtexte liegen. Die abgeleiteten Formate dürfen die schutzbegründenden Merkmale nicht enthalten. Dann werden die durch Urheberrecht und Leistungsschutzrechte geschützten Verwertungsinteressen der Rechteinhaber bei der Nutzung solcher Formate zu Forschungszwecken und zur Forschungstransparenz nicht beeinträchtigt. Die Verwendung und Zugänglichmachung der abgeleiteten Formate für automatisierte Analyseverfahren ist deshalb auch außerhalb der urheberrechtlichen Schrankenbestimmungen möglich.

Die Herausforderung besteht darin, Formate zu finden, die sich technisch zuverlässig ohne zu hohen Rechenaufwand erstellen lassen.[17] Die Formate dürfen den Schutz der Primärtexte nicht tangieren, müssen aber dennoch die forschungsrelevanten 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 Vervielfäl-

14 *Schöch et al*, RuZ 2020, 160, 162 ff., erläutern exemplarisch, welche Informationen für gängige TDM-Analyseverfahren der Digital Humanities benötigt werden; siehe auch *Schöch et al*., ZfdG 2020.

15 *Schöch et al*, RuZ 2020, 160, 165.

16 Siehe für die Beschreibung solcher Formate und deren Verwendungsmöglichkeiten: *Schöch et al*, RuZ 2020, 160, 163 ff.

17 Vgl. *Schöch et al*, RuZ 2020, 160; *Schöch et al*, ZfdG 2020.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

tigungsrechts etabliert.[18] Wird ein Schutzgegenstand oder ein Teil davon[19] technisch derart übernommen, dass er nicht mehr erkennbar ist, ist die Festlegung in dieser Form keine Vervielfältigung des Schutzgegenstandes.[20] Diese aus dem allgemeinen Sprachgebrauch entnommene Erkenntnis[21] des EuGH dürfte nicht auf den Vervielfältigungsbegriff im Bereich des Tonträgerherstellerschutzes 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 heran: Die Rechte des geistigen Eigentums werden nicht schrankenlos gewährt. Ändernde, nicht wiedererkennbare Nutzungen in Ausübung eines anderen Grundrechts[23] 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 Verbotsrecht 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

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.

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 – *Goldrapper*.

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.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

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 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 Schutzumfang.[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]

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*, RuZ 2020, 128, 138. und den Referentenentwurf des BMJV zur Umsetzung von Art. 15 UrhR-RL.

31 Nach der Begründung des EuGH liegt nämlich dann keine Vervielfältigung vor.

32 Wie *Jotzo*, RuZ 2020, 128, 134, 137 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; sowie *Schöch* et al, ZfdG 2020.

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 Parodieschranke) 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 Schutzbereichsbeschrän-

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

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 Außenverhältnis[40] geschützte Beziehung zum Werk nicht beeinträchtigt. Der Werkgenuss wird nicht substituiert. Zudem ermöglicht nur das Abstellen auf die Rezipienten eine externe Überprüfbarkeit. Daher ist darauf abzustellen, ob ein durchschnittli-

kung von § 24 UrhG fortbestehen (S. 82). Bei hinreichendem Abstand zum vorbestehenden Werk soll keine 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*, RuZ 2020, 118, 123 f., in diesem Heft. § 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: Schricker/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.: Schricker/Loewenheim-*Peukert*, Urheberrecht, 6. Aufl. 2020, § 13 UrhG Rn. 1; § 12 UrhG sichert dem Urheber das Veröffentlichungsrecht.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

cher, nicht vorinformierter[41] 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 fragmentierten[44] 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 Teilvervielfältigungen 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 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]

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, 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 Schricker/Loewenheim-*Loewenheim*, Urheberrecht, 6. Aufl. 2020, § 16 UrhG Rn. 15 mwN; Dreier/Schulze-*Schulze*, Urheberrechtsgesetz, 6. Aufl. 2018, § 16 UrhG Rn. 11, 12.

46 Hoeren/Sieber/Holznagel-*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.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

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 Modelle[48] o.Ä.) nicht zur Verfügung gestellt, insbesondere nicht mit 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 Einzelfallbetrachtung[49] 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]

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.

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 berechtigte 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 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.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

2. Beispiele[52] und urheberrechtliche Bewertung abgeleiteter Formate

a) Term-Dokument-Matrizen[53]

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 ist,[54] 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)selektion[55]

Ein Format, in dem lediglich alle Worte eines Textes unter Beibehaltung der Wortreihenfolge durch Lemmata[56] 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.

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

52 Für eine Darstellung der hier diskutierten abgeleiteten Formate und deren praktische Verwendungsmöglichkeiten siehe: *Schöch et al*, RuZ 2020, 160, 163 ff.

53 *Schöch et al*, RuZ 2020, 160, 163 ff.

54 Bei sehr kurzen Texten könnte sich aufgrund zwingender grammatikalischer Regelungen aus den wenigen aufgeführten Worten der Texte rekonstruieren lassen.

55 Vgl. *Schöch et al*, RuZ 2020, 160, 164, 167 f.

56 Lemma ist die Grund- oder Nennform eines Wortes, „Lemma (Lexikographie)“, in: Wikipedia, Die freie Enzyklopädie, Stand 28. April 2019, 111:23 UTC, abrufbar unter: https://de.wikipedia.org/w/index.php?title=Lemma_(Lexikographie)&oldid=187995927, zuletzt abgerufen am 19.10.2020.

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.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

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 Sequenzinformation[58]

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 Presseerzeugnissen mit Blick auf das Presseverlegerleistungsschutzrecht[62] 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 Figurenbeschreibungen[64] mit Sicherheit unkenntlich und desto schwieriger wird die Rekonstruktion bzw. sie wird gar unmöglich.

d) N-Gramm-Tabellen[65]

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

58 *Schöch et al*, RuZ 2020, 160, 165 f.
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*, RuZ 2020, 128, 133, 137.
60 Siehe erläuternd: *Jotzo*, RuZ 2020, 128, 133.
61 EuGH, Urt. v. 16.07.2009 – C-5/08, ECLI:EU:C:2009:465 = GRUR 2009, 1041 Tz. 48 – *Infopaq*.
62 Die §§ 87f ff. UrhG sind zwar wegen unterbliebener Notifizierung nicht anwendbar, jedoch sind Nachfolgeregelungen bereits im Entstehen, siehe dazu: *Jotzo*, RuZ 2020, 128, 138 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*, RuZ 2020, 128, 135, mwN.
65 Zur Verwendung von N-Grammen: *Schöch et al*., RuZ 2020, 160, 169 ff.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

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.

e) Formate aus Wort-Vektoren auf Grundlage eines Word Embedding Models[66]

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 Rechtsunsicherheit[67] 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 Zugang[68] zu den Primärtexten besteht.

Sofern geschützte Texte erst auf einem Server gesammelt und damit nicht nur vorübergehend[69] digital vervielfältigt werden, setzt dies die Zustimmung der Rechteinha-

66 *Schöch et al*, RuZ 2020, 160, 171 ff.
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.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ber oder das Vorliegen einer gesetzlichen Schrankenbestimmung voraus. Als Schrankenbestimmungen kommen insbesondere § 60d[70] oder § 60c[71] 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 Schutzmaßnamen 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]

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 vor-

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, S. 41.

73 Die Regierungsbegründung nennt beispielhaft Vervielfältigungen durch Mitarbeiter einer Bibliothek, BT-Drs. 18/12329, S. 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 S. 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.

77 Dreier/Schulze-*Dreier*, Urheberrechtsgesetz, 6. Aufl. 2018, § 87b UrhG Rn. 4; *Jotzo*, RuZ 2020, 128, 140 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*, RuZ 2020, 128, 140; *Spindler* GRUR 2016, 1112, 1114.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

behalten, 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 UrhG[80] 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 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

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.

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

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

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 erziel[t]" werden können.[87] Auf den wirtschaftlichen Wert eines möglichen, aus dem technischen 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

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*.

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.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

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 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.

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.

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 Schricker/Loewenheim-*Peukert*, Urheberrecht, 6. Aufl. 2020, Vor §§ 12 ff. UrhG Rn. 4.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

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 Werk[94] 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, 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 allgemeingesellschaftlich 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.

94 § 11 S. 1 UrhG.

https://doi.org/10.5771/2699-1284-2020-2-143 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA