Originalveröffentlichung in: Beißwenger, Michael/Lemnitzer, Lothar/Müller-Spitzer, Carolin (Hrsg.): Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium. – Paderborn: Wilhelm Fink, 2022. S. 277-291. (UTB 5711)

POSTPRINT

Pawel Kamocki
Andreas Witt

15. Was darf die sprachwissenschaftliche Forschung? Juristische Fragen bei der Arbeit mit Sprachdaten

Sich in der Linguistik mit rechtlichen Themen beschäftigen zu müssen, ist auf den ersten Blick überraschend. Da jedoch in den Sprachwissenschaften empirisch gearbeitet wird und Sprachdaten, insbesondere Texte und Ton- und Videoaufnahmen sowie Transkripte gesprochener Sprache, in den letzten Jahren auch verstärkt Sprachdaten internetbasierter Kommunikation, als Basis für die linguistische Forschung dienen, müssen rechtliche Rahmenbedingungen für jede Art von Datennutzung beachtet werden. Natürlich arbeiten auch andere Wissenschaften, wie z. B. die Astronomie oder die Meteorologie, empirisch. Jedoch gibt es einen grundsätzlichen Unterschied der empirischen Basis: Im Gegensatz zu Temperaturen, die gemessen, oder Konstellationen von Himmelskörpern, die beobachtet werden, basieren Sprachdaten auf schriftlichen, mündlichen oder gebärdeten Äußerungen von Menschen, wodurch sich juristisch begründete Beschränkungen ihrer Nutzung ergeben.

1. Einleitung

In diesem Kapitel werden einige grundlegende Konzepte des geistigen Eigentums (Abschnitt 2) und des Datenschutzrechts (Abschnitt 3) vorgestellt, die für die sprachwissenschaftliche Forschung schon während Ihres Studiums relevant werden können. In den folgenden Abschnitten erfahren Sie, wie Sie Daten für Ihre ersten Forschungsprojekte zusammentragen und diese verarbeiten, ohne gegen die gesetzlichen Bestimmungen zu verstoßen, und in welcher Form Sie Ihre Ergebnisse anderen Personen zur Verfügung stellen dürfen.

Den Abschnitt zum Schutz des geistigen Eigentums (insbesondere 2.3.–2.5) sollten Sie lesen, wenn Sie für Ihre Forschung eine Sammlung z. B. von Text- oder Audiodaten anlegen oder eine derartige Sammlung von Dritten verwenden möchten. Den Abschnitt zu Open Access und Lizenzmodellen (Abschnitt 2.6) sollten Sie lesen, wenn Sie planen, eine wissenschaftliche Arbeit oder eine selbst erstellte Datensammlung zu veröffentlichen. Der Abschnitt über die Persönlichkeitsrechte (Abschnitt 3) ist relevant für Sie, wenn die von Ihnen erhobenen Daten, z. B. aus einer Befragung oder aus Experimenten, personenbezogene (Meta-)Daten enthalten müssen, damit Sie Ihr Forschungsziel erreichen können.

2. Geistiges Eigentum

2.1 Einleitung oder: Warum benötigte Leibniz keinen rechtlichen Schutz seines geistigen Eigentums?

Wenn Sie in einer Buchhandlung das Regal der belletristischen Neuerscheinungen durchstöbern, werden Sie viele meist hochwertig produzierte neue Ausgaben von Romanen finden, deren Verkaufspreise in Deutschland und Österreich derzeit meist zwischen 20 € und 25 € liegen. Zweifelsohne hat das haptisch erfahrbare Objekt Buch einen Wert, jedoch sind die Kosten dafür, das Papier zu kaufen, zu bedrucken, zu schneiden, zu binden, auszuliefern etc. meist sehr viel geringer als der Preis des Buches, der in den Buchhandlungen zu zahlen ist. Ein Grund für den höheren Preis ist, dass ein Team von Verlagsangestellten eine beträchtliche Anzahl von Arbeitsstunden aufgewandt hat, die Romane zu lektorieren, zu bewerben und zu vertreiben. Die wichtigste Tätigkeit, die vor der Publikation eines Romans geleistet werden

Publikationsserver des Leibniz-Instituts für Deutsche Sprache
URN: http://nbn-resolving.de/urn:nbn:de:bsz:mh39-110671

musste, ist jedoch das Verfassen des Texts. Die Schriftsteller*innen oder die Verfasser*innen der Werke haben meist viel Zeit damit verbracht, die Texte zu schreiben. Ein Buch – oder analog dazu ein Gemälde oder ein Musikstück – hat also einen höheren Wert als der Wert des analogen oder digitalen Mediums, auf dem es gespeichert ist. Das ist grundsätzlich auch beim Handel mit anderen Gütern der Fall. So kostet ein hochwertiges faltbares Fahrrad im Handel erheblich mehr, als seine Herstellung kostet. Ein zentraler Unterschied zwischen schöpferischen Werken und Gebrauchsgütern besteht jedoch darin, dass der Aufwand, eine nutzbare Reproduktion oder Kopie von schöpferischen Werken zu erstellen, wesentlich geringer ist als der Nachbau von Gebrauchsgegenständen. Der Handel mit Wirtschaftsgütern kann entsprechend nur dann sicher und effizient erfolgen, wenn er auf Rechtsvorschriften beruht, wobei Güter, die das Produkt schöpferischer Tätigkeit sind, hierbei besonders geschützt werden müssen.

Geschriebene Texte werden als geistiges Eigentum und damit als immaterieller Vermögenswert angesehen. Sie sind das Produkt menschlicher Kreativität und werden durch Gesetze geschützt. Dies war nicht immer der Fall. In den Zeiten von Leibniz (und davor) gab es keinen gesetzlichen Schutz von geistigem Eigentum. Ein Grund hierfür ist darin zu finden, dass es keinen Bedarf für einen derartigen Schutz gab. Da es schlicht sehr teuer war, Bücher herzustellen, war auch das Kopieren von Texten nicht lukrativ. Aus diesem Grund wurde bis zum Ende des 17. Jahrhunderts der eigenständige wirtschaftliche Wert von Inhalten noch nicht gesehen. Erst mit der Entwicklung einer immer effizienteren Druckindustrie veränderte sich diese Lage. In diesem Zusammenhang entstand das Urheberrecht.

Einfach ausgedrückt ist das Urheberrecht ein Recht an geistigem Eigentum, das den Urheber*innen eine gewisse Kontrolle über ihre Werke einräumt. Um seine Rolle und seine wachsende Bedeutung für die heutige Wirtschaft vollständig zu verstehen, wird im folgenden Abschnitt kurz die Geschichte des Urheberrechts vorgestellt (Abschnitt 2.2). Anschließend wird der Gegenstand des Urheberrechts erörtert, d. h. die Frage, welche Güter eigentlich vom Urheberrecht geschützt werden und welche nicht (Abschnitt 2.3). Zudem wird dargestellt, was es für die Nutzung eines Werkes bedeutet, dass es urheberrechtlich geschützt ist (Abschnitt 2.4). Wie Sie sehen werden, sind die allgemeinen Regeln, die das Urheberrecht formuliert, sehr streng und könnten daher negative Auswirkungen auf die (sprach-)wissenschaftliche Nutzung von Texten und Sprachaufnahmen haben. Um diese negativen Konsequenzen abzuschwächen, hat die Gesetzgebung spezielle Regelungen für die Wissenschaft eingeführt, die in Abschnitt 2.5 vorgestellt werden. Hier erfahren Sie insbesondere, unter welchen Bedingungen die Wissenschaft geschützte Daten auch ohne Erlaubnis der Urheber*innen nutzen darf.

Neben dem Urheberrecht ist auch das Verwertungsrecht und damit verbunden das Thema Lizenzen für die Forschung relevant. Wenn Sie einen wissenschaftlichen Artikel schreiben, ein Textkorpus für eine wissenschaftliche Untersuchung zusammenstellen oder eine Software entwickeln, haben Sie, wie erwähnt, unter bestimmten Bedingungen das Urheberrecht daran. Um Ihr Werk für die Community zugänglicher und nützlicher zu machen (und die Wahrscheinlichkeit zu erhöhen, dass Ihre Arbeit bemerkt und zitiert wird), sollten Sie darüber nachdenken, dieses Werk unter einer entsprechenden Lizenz möglichst offen zur Verfügung zu stellen. In Abschnitt 2.6 erhalten Sie hierzu einige grundlegende, praktisch anwendbare Informationen.

Das Urheberrecht ist nicht das einzige Recht an geistigem Eigentum, das sich auf die sprachwissenschaftliche Forschung auswirken kann. Abhängig von der Art der Daten, die Sie für Ihre Forschung verwenden wollen, werden Sie möglicherweise mit einer Reihe von verwandten Schutzrechten konfrontiert, die z. B. die Nutzung von Datenbanken oder wissenschaftlichen Ausgaben gemeinfreier Werke regeln. Diese Rechte werden in Abschnitt 2.7 kurz erläutert.

2.2 Ein bisschen Rechtsgeschichte…

Um die Druckindustrie vor sogenannten Piraten zu schützen, die lediglich Bücher kopieren und sie kostengünstiger herstellen und verkaufen wollten, wurde 1710 in England das erste Urheberrechtsgesetz verabschiedet. Das Gesetz wurde nach der damaligen Königin benannt und ist als *Statute of Anne* in die Geschichte eingegangen. Bis Ende des 18. Jahrhunderts wurden in den USA, die weitgehend die *Statute of Anne* aus England übernommen haben, sowie in Frankreich und in Preußen ebenfalls Urheberrechtsgesetze erlassen. Im Jahr 1886 wurde zudem mit der Berner Übereinkunft erstmals ein Abkommen zur internationalen Anerkennung von Schutzrechten an der Urheberschaft geschlossen. Dieses leitete einen Prozess der internationalen Harmonisierung in diesem Bereich ein, was zum Teil auf die Bemühungen des berühmten französischen Schriftstellers Victor Hugo zurückzuführen war. Die Berner Übereinkunft ist – mit mehreren Überarbeitungen – nach wie vor in Kraft. In der Bundesrepublik Deutschland wurde 1965 das derzeitige Urheberrechtsgesetz in seiner ersten Fassung verabschiedet.

Wie alle anderen Rechtsnormen wird auch das Urheberrecht kontinuierlich im Kontext der gesellschaftlichen und technischen Entwicklungen neu bewertet und bei Bedarf weiterentwickelt. In den vergangenen Jahrzehnten gab es bemerkenswert häufig Anstöße zur Reformierung des Regelwerks. Schon in den 1970er Jahren wurde mit Blick auf die Popularisierung von analogen Geräten zur Reproduktion – Bücher wurden mit Kopiergeräten, Musik wurde mit Kassettenrekordern kopiert – ein besserer Schutz der Ansprüche der künstlerisch-schöpferisch tätigen Menschen und ihrer Vertriebspartner (Verlage, Plattenfirmen) gefordert.

Da die Herstellung von Analogkopien jedoch mit einem relativ hohen Arbeitsaufwand, hohen Kosten und häufig mit Qualitätseinbußen einherging, wurde das Thema Urheberrecht in Spezialdiskussionen behandelt. Eine breite gesellschaftliche Diskussion über das Urheberrecht begann, als durch digitale Medien eine einfache, verlustfreie Reproduzierbarkeit und, durch das Internet, ein umfassender Vertrieb der kopierten Text-, Audio- und Videodateien zu einem Massenphänomen wurde. Auf internationaler Ebene wurde mit einem Übereinkommen der Welthandelsorganisation (WTO) über handelsbezogene Aspekte der Rechte des geistigen Eigentums (1995), dem Urheberrechtsvertrag der Weltorganisation für geistiges Eigentum (1996), dem Digital Millenium Copyright Act in den USA (1998) und nicht zuletzt mit der EU-Richtlinie über das Urheberrecht in der Informationsgesellschaft (2001) auf diese Entwicklungen reagiert. Das letztgenannte Dokument löste eine Welle von Reformen des deutschen Urheberrechts (unterteilt in „drei Körbe") aus, die bis 2014 erfolgten. Anschließend, im Jahr 2017, wurde das deutsche Urheberrecht durch das *Gesetz zur Anpassung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft* (UrhWissG) nochmals umfassend neu geregelt, wobei die Regelung der akademischen Nutzungen und die Verwendung neuer Technologien im Zentrum standen. Im Jahr 2019 wurde zudem eine neue EU-Richtlinie über das Urheberrecht im digitalen Binnenmarkt angenommen, die für alle Staaten der EU bindend ist, was bedeutet, dass die Gesetze aller EU-Mitgliedsstaaten diese Richtlinie umsetzen müssen.

2.3 Überblick über das Urheberrecht

Es gibt viele falsche Vorstellungen darüber, was urheberrechtlich geschützt ist oder geschützt werden kann. In der folgenden Darstellung des Urheberrechts wird speziell auch auf einige dieser falschen Vorstellungen eingegangen.

Das Urheberrecht schützt weder Erfindungen noch Markennamen oder Logos. Erstere können durch Patente geschützt werden, letztere durch das Markenrecht. Das Urheberrecht schützt Werke der Literatur, Wissenschaft und Kunst. Zu dieser Kategorie gehören Sprachwerke, kreative Computerprogramme, Werke der bildenden Künste und Musik sowie Darstellungen wissenschaftlicher oder technischer Art.

Der Urheberrechtsschutz entsteht, im Gegensatz zum Schutz durch Patente, automa-

tisch. Es gibt daher keine Formulare, keine Eintragungspflicht oder andere Formen der Anmeldung dieses Rechts. Mit der Schaffung des Werks durch den kreativ tätigen Menschen greift das Urheberrecht. Allerdings muss im Zweifelsfall nachgewiesen werden können, wer das Werk erstellt hat, was nicht immer einfach ist. In der Praxis besteht eine starke Vermutung, dass diejenigen, deren Namen auf dem Werk erscheinen, die eigentlichen Urheber*innen sind.

Eine weitverbreitete – jedoch falsche – Annahme über das Urheberrecht betrifft die künstlerische Qualität des geschützten Werks. Der künstlerische Wert der Werke ist unerheblich für einen Urheberrechtsschutz. Die einzige Voraussetzung für den Schutz ist die Originalität. Ein Werk gilt im Sinne des Urheberrechts als originell (oder: individuell), wenn es eine eigene geistige Schöpfung darstellt.[1] Das bedeutet, dass alle Werke originell und somit urheberrechtlich geschützt sind, die eine „persönliche Note" der Urheber*innen tragen, die in der Auswahl und Anordnung der enthaltenen Elemente (Wörter, Töne, Formen, Farben usw.) zum Ausdruck kommt. Jeder Prozess, bei dem kreative Entscheidungen möglich sind (Kurzgeschichte, Essay, aber auch Slogan oder Vertrag), kann daher die Entstehung eines schützenswerten Werks zur Folge haben.

Eine praktische Prüfung der Originalität könnte mit Hilfe folgender Fragestellung überprüft werden: Können zwei gleich qualifizierte Personen unabhängig voneinander genau dasselbe Ergebnis erzielen? Es ist äußerst unwahrscheinlich, dass zwei Autor*innen genau denselben Roman schreiben oder zwei Maler*innen dasselbe Gemälde erstellen. Wenn es aber darum geht, einen Slogan zu (er-)finden, liegen die Chancen von zwei Urheber*innen, die gleichen Ergebnisse zu liefern, sehr viel höher. Aus diesem Grund sind kurze Geschichten und Gemälde praktisch immer originell, während beispielsweise nur sehr wenige Fotos des Eiffelturms in diesem Sinne originell sind.

Der Schutz gilt nicht nur für das Werk als Ganzes, sondern auch für seine konstitutiven Teile, solange es sich um Originale handelt, d. h., wenn es unwahrscheinlich ist, dass diese von einer anderen Person unabhängig ebenfalls geschaffen werden. „Ich bin Pastafari, ich glaube an das Spaghettimonster" ist vermutlich ein unoriginelller Satz, Noam Chomskys berühmter Satz „Colorless green ideas sleep furiously" gilt hingegen wahrscheinlich als originell.

Warum „vermutlich", warum „wahrscheinlich"? In viele Fällen ist die Entscheidung, ob eine bestimmte Wortfolge aus juristischer Sicht als urheberrechtlich schützenswert betrachtet wird oder nicht, schwer vorhersagbar. So entschied das Oberlandesgericht München im Dezember 2019, dass der Satz „Früher war mehr Lametta.", den der Satiriker Loriot in einem Sketch äußerte, keine hinreichende Schöpfungshöhe für einen Schutz besitze.[2] Ganz ähnlich stellt sich die Frage nach dem Werkcharakter von Postings in öffentlichen Chat-Räumen oder anderen Formen internetbasierter Kommunikation: Welchen Grad von Originalität müssen diese aufweisen, damit sie als Werk im Sinne des Urheberrechts betrachtet werden können und somit als schützenswert gelten? Diese Frage ist für empirische Untersuchungen im Bereich der Medienlinguistik und für den Aufbau von Korpora internetbasierter Kommunikation (→ Kapitel 26 [Korpora internetbasierter Kommunikation] in diesem Band) nicht unerheblich (vgl. Beißwenger et al. 2017).

Nur der Ausdruck (*die Form*) eines Werks ist geschützt, nicht aber der Inhalt oder die „Idee". Das bedeutet, dass Sie zwar keine Kopie eines Harry Potter Bandes von J. K. Rowling verkaufen dürfen. Sehr wohl dürfen Sie jedoch einen Roman über einen Waisenjungen verkaufen, der feststellt, dass er magische Kräfte hat. Tatsächlich gab es solche Geschich-

1 Diese Definition wurde 2009 vom Gerichtshof der Europäischen Union angenommen (Rechtssache C-5/08 Infopaq), mit der das Konzept der Originalität in der gesamten EU harmonisiert wurde. Erwähnenswert ist, dass sie sich nur geringfügig von der traditionellen Definition der Originalität im deutschen Urheberrechtsrecht unterscheidet.

2 OLG München, Urteil vom 20.12.2019, Az. 6 W 927/19.

ten schon lange vor Harry Potter und neue kommen hinzu. Das Urheberrecht schränkt nicht die Kreativität der Kunstschaffenden ein, sondern fördert vielmehr das kreative Schaffen, indem es Urheber*innen ermöglicht, von ihrem Werk zu leben. Zumindest ist das die Grundidee, wobei jedoch immer auch Fragen der Freiheit der Kunst bei rechtlichen Auseinandersetzungen thematisiert werden.[3]

Darüber hinaus ist es sinnvoll zu verstehen, dass urheberrechtlich geschützte Werke abstrakte, nicht reale Gegenstände sind. Die meisten Werke sind in einem physischen Träger fixiert (als *Corpus mechanicum* bezeichnet, z. B. auf Papier gedruckt oder in Stein gemeißelt); das Urheberrecht berührt jedoch nicht das Eigentum an diesen Gegenständen. Das Urheberrecht schützt Werke in ihrem abstrakten Sinne: Wenn der materielle Träger einer Skulptur zerstört wird, ist seine „metaphysikalische" Form nach wie vor urheberrechtlich geschützt (sofern die Schutzdauer nicht abgelaufen ist, wie nachstehend erläutert wird), und nur der Urheber/die Urheberin (und nicht der Eigentümer/die Eigentümerin der physischen Skulptur) kann rechtmäßig eine weitere Kopie davon anfertigen oder die Anfertigung einer solchen Kopie genehmigen.

Schließlich ist das Urheberrecht zuerkannte Eigentum – im Gegensatz zu einem Eigentum an materiellen Gegenständen – zeitlich begrenzt. Es gilt in Deutschland für die Lebenszeit der Urheberin oder des Urhebers und weitere siebzig Jahre. Nach Ablauf dieses Zeitraums gelangt das Werk in die Öffentlichkeit und kann frei genutzt werden. Retrospektiv ist eine starke Tendenz der Gesetzgebung beobachtbar, die Dauer des Urheberrechts zu verlängern — interessanterweise sah die *Statute of Anne* erst nach der Eintragung des Werks einen Schutz von 14 Jahren vor.

2.4 Praktische Auswirkungen des Urheberrechtsschutzes

Wie oben erläutert gewährt die Urheberschaft den Autor*innen Eigentumsrechte an ihren Werken, die als abstrakte, immaterielle Objekte betrachtet werden. Aber was bedeutet es, Exklusivrechte an einem immateriellen Gut wie einem urheberrechtlich geschützten Werk zu haben?

Der Zweck der Exklusivrechte an Werken bestand zu Beginn des Urheberrechts darin, die Druckindustrie zu schützen. Daher bestand das Urheberrecht vom ersten Tag an aus dem Exklusivrecht, Kopien eines Werks anzufertigen und diese Kopien an die Öffentlichkeit zu verteilen. Dazu ist die Erlaubnis des Urhebers/der Urheberin erforderlich. Die zwei Grundrechte sind dann das Vervielfältigungsrecht und das Recht der öffentlichen Zugänglichmachung. Ein drittes, traditionell anerkanntes ausschließliches Recht ist das Anpassungsrecht, d. h. das Recht, Anpassungen, Übersetzungen und andere abgeleitete Werke auf der Grundlage des Originalwerks (z. B. ein auf einem Buch basierenden Film) vorzunehmen bzw. zu erstellen.

In der digitalisierten Welt wurde es viel komplizierter. Wie Sie sich erinnern, ist ein Werk unabhängig vom Medium, auf dem es dargeboten und gespeichert ist, geschützt: Digitale Werke genießen den gleichen Schutz wie Bücher. Ein Verbot, Kopien von gedruckten Büchern anzufertigen, mag gelegentlich als Unannehmlichkeit angesehen werden. Der Umstand, dass das gleiche Verbot auch für digitale Texte, z. B. E-Books, gilt, hatte gravierende Auswirkungen auf die Gestaltung der digitalen Welt. Alles, was Sie mit digitalen Werken tun – auch wenn Sie sie nur auf Ihrem Bildschirm anzeigen – bedeutet zwangsläufig, dass Sie Kopien davon erstellen, und sei es auch nur temporär im Arbeitsspeicher Ihres Computers. Wenn wir strenge urheberrechtliche Regeln für digitalisierte Werke anwenden, würde dies in der Tat bedeuten, dass buchstäblich nichts ohne Erlaubnis der

3 In einem mehr als 20 Jahren währenden Rechtsstreit zwischen den Musikern Moses Pelham und der Band Kraftwerk ging es im Kern um die Frage, ob (HipHop-)Musiker*innen ungefragt Ausschnitte urheberrechtlich geschützter Werke anderen Musiker*innen für das Sampling verwenden dürfen.

Rechteinhaber*innen mit fremden Texten gemacht werden kann.

Glücklicherweise gibt es eine Reihe von Ausnahmen (Urheberrechtsschranken) von diesen strengen, allgemeinen Regeln – darunter eine, die es ermöglicht, im Internet zu surfen, ohne ständig Berechtigungen erfragen oder erwerben zu müssen.[4] Es gibt auch Ausnahmen für den Einsatz von Texten im Unterricht und in der Forschung, die im nächsten Unterabschnitt vorgestellt werden. Es ist jedoch wichtig zu wissen, dass diese Urheberrechtsausnahmen nur zu eng abgegrenzten Nutzungsformen berechtigen. Wann immer ein digitales Werk auf eine Weise verwendet werden soll, die nicht von einer Ausnahme erfasst wird, muss hierfür eine Erlaubnis der Rechteinhaber*innen eingeholt werden.

Neben den Verwertungsrechten für Vervielfältigung, öffentliche Zugänglichmachung und Anpassung haben Autor*innen auch eine Reihe von Persönlichkeitsrechten. Das wichtigste dieser Rechte ist die Anerkennung der Urheberschaft, d. h. das Recht eine Namensnennung einfordern zu dürfen, wenn seine oder ihre Werke zitiert oder anderweitig verwendet werden. Ein weiteres wichtiges Persönlichkeitsrecht ist das Recht, die Entstellung des Werkes zu verhindern (z. B. die Neuordnung von Sätzen in einem Zitat, um dem Autor eine falsche Meinung unterzuschieben).

2.5 Urheberrecht und Wissenschaft: Ausnahmen für Unterricht und Forschung

Wie Sie sich vorstellen können, ist der Prozess der Einholung bestimmter Nutzungsrechte bei den Autor*innen nicht nur zeitaufwändig, sondern oft auch faktisch unmöglich. Wenn Linguist*innen mit großen Textsammlungen arbeiten, um z. B. den Sprachwandel zu erforschen, müssten tausende Rechteinhaber*innen, d. h. Autor*innen oder deren Erb*innen, um Erlaubnis gefragt werden.

Die Verwendung urheberrechtlich geschützter Werke für Forschungs- und Lehrzwecke wurde lange Zeit aus urheberrechtlicher Sicht als nahezu irrelevant angesehen. In der Tat konnte ein Gelehrter vor der Mitte des 20. Jahrhunderts nur manuell Auszüge von Büchern aus Bibliotheken kopieren und sie dann seinen Schüler*innen vorlesen, eine Aktivität, die die Verlagsbranche nicht störte. Man kann sagen, dass das Urheberrecht und die Wissenschaft friedlich koexistierten. Das änderte sich, als das Kopieren in den Fluren der Schulen, Universitäten und in den Bibliotheken allgegenwärtig wurde.

Die Berner Übereinkunft verlieh in ihrer ursprünglichen Fassung von 1886 den nationalen Gesetzgebern, „Befugniß, Auszüge oder Stücke aus Werken der Literatur und Kunst in Veröffentlichungen, welche für den Unterricht bestimmt oder wissenschaftlicher Natur sind, oder in Chrestomathie[5] aufzunehmen, vorzusehen."[6] Während der Überarbeitung des Übereinkommens von 1967 geschah etwas ziemlich Überraschendes. Der Verweis auf „wissenschaftliche Natur" wurde gestrichen, und die Formulierung wurde geändert, um nur Verwendungen „zur Veranschaulichung des Unterrichts" zuzulassen.[7]

Als die europäische Gesetzgebung um die Jahrtausendwende an der EU-Richtlinie zum Urheberrecht arbeitete, musste sie den internationalen Verpflichtungen der Mitgliedstaaten nachkommen, die alle an die Berner Übereinkunft gebunden waren. Dies erklärt teilweise die Formulierung der harmonisierten Ausnahme für Forschung und Lehre in dieser 2001 verabschiedeten Richtlinie.

Die Richtlinie (genauer gesagt Artikel 5.3 (a)) erlaubt es den EU-Mitgliedstaaten, in ih-

4 Diese Ausnahme wurde im Artikel 5.1 der EU-Urheberrechtsrichtlinie von 2001 formuliert. Diese Ausnahme ist die einzige obligatorische Ausnahme in der Richtlinie, d. h. sie musste von allen EU-Mitgliedstaaten umgesetzt werden.

5 Das heute sehr wenig gebräuchliche Wort Chrestomathie bezeichnet eine Zusammenstellung von Texten zu Lehrzwecken.

6 Artikel 8 der ersten Fassung der Berner Übereinkunft.

7 Artikel 10.2 der aktuellen Fassung der Berner Übereinkunft.

ren nationalen Rechtsvorschriften Ausnahmen „für die Nutzung ausschließlich zur Veranschaulichung im Unterricht oder für Zwecke der wissenschaftlichen Forschung, sofern – außer in Fällen, in denen sich dies als unmöglich erweist – die Quelle, einschließlich des Namens des Urhebers (...) angegeben wird und soweit dies zur Verfolgung nicht kommerzieller Zwecke gerechtfertigt ist", zu erlassen. Mit anderen Worten: In den Ländern der EU konnten durch diese Richtlinie Urheberrechtsschranken für nichtkommerzielle Lehre und Forschung beschlossen werden – eine Ausnahme für die kommerzielle Forschung wäre mit dem EU-Recht jedoch unvereinbar.

Bemerkenswert ist, dass nichts in der Richtlinie die Größe der Auszüge einschränkt, die verwendet werden können (es könnte theoretisch möglich sein, eine ganze Enzyklopädie zu kopieren und zu teilen, solange dies für nichtkommerzielle Forschungszwecke ist), oder auch die Anzahl der Begünstigten: Akademiker*innen, Student*innen oder auch Bürgerwissenschaftler*innen können hier gemeint sein.

In der Praxis entscheiden sich die nationalen Gesetzgeber jedoch meist für relativ restriktive Schrankenregelungen und fügen viele Bedingungen hinzu, die in der Richtlinie nicht vorgeschrieben sind. Dies war bis 2018 auch in Deutschland der Fall, wo Forscher*innen nur kleine Teile eines Werkes und Werke geringen Umfangs wie einzelne Gedichte oder Artikel verwenden durften und einer Verwertungsgesellschaft (wie der *VG Wort*) eine angemessene Vergütung für die Verwendung zu zahlen hatten (die in der Praxis die Verwendung der Ausnahme zur institutionalisierten Forschung reduzierte) und selbst dann die Werke nur mit einem bestimmten abgegrenzten Kreis von Personen teilen konnten[8]. Es ist unnötig zu erwähnen, dass solche Ausnahmen für Forscher*innen, die digitale Daten verwenden, keinen praktischen Nutzen hatten.

In den letzten zehn Jahren wurde die Legislative in vielen Ländern auf die Bedeutung von Technologien wie Data Mining für die Forschung aufmerksam gemacht. Infolgedessen wurden in (ehemaligen) EU-Mitgliedstaaten wie Großbritannien, Frankreich oder Deutschland neue Ausnahmen für das Data Mining zu Forschungszwecken eingeführt. Aus den oben erläuterten Gründen konnten diese Ausnahmen jedoch nur nichtkommerzielle Unterrichts- und Forschungstätigkeiten abdecken.

Tab. 1: Urheberrechtsschranken für die Wissenschaften

Welche Werke können benutzt werden?	Alle Werke; einzige Ausnahme: Es ist nicht erlaubt, während öffentlicher Vorträge, Aufführungen oder Vorführungen eines Werkes diese auf Bild- oder Tonträger aufzunehmen und später öffentlich zugänglich zu machen.

Wie groß sind die Teile von Werken, die benutzt werden können?	• Für eigene Nutzungsformen, zu denen nicht das Teilen zählt: bis zu 75 Prozent eines Werkes
	• Für eigene Nutzungsformen, zu denen auch das Teilen zählt: bis zu 15%
	• Vollständig genutzt werden können zudem in Ausnahmefällen:
	  • einzelne Beiträge aus derselben Fachzeitschrift oder wissenschaftlichen Zeitschrift
	  • Werke geringen Umfangs, d. h. einzelne Gedichte, Kurzgeschichten oder andere gedruckte Werke, die kürzer als 25 Seiten sind, Bilder, Fotografien, Ton- oder Videoaufnahmen, die kürzer als 5 Minuten sind
	  • vergriffene Werke

Wer kann die Werke oder Teile des Werkes benutzen?	Jede*r
Zu welchem Zweck?	Zum Zweck der nicht-kommerziellen wissenschaftlichen Forschung
Mit wem können die Werke oder Teile des Werkes geteilt werden?	Nur mit einem bestimmt abgegrenzten Kreis von Personen (z. B. innerhalb eines Forschungsteams) und mit einzelnen Dritten, soweit dies der Überprüfung der Qualität wissenschaftlicher Forschung dient.

8 § 52a UrhG (bis 1. März 2018).

In Deutschland fand die Reform im Jahr 2018 statt, als ein neues Gesetz unter dem vielversprechenden Titel *Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft (UrhWissG)* in Kraft trat. Mit dem UrhWissG wurden eine Reihe neuer Bestimmungen und Änderungen im Urhebergesetz (UrhG) eingeführt, die insbesondere Bereiche wie Unterricht und Lehre (§60a), Lehrmedien (§60b), nicht-kommerzielle Forschung (§60c), Text und Data Mining (§60d) sowie Bibliotheken (§60e) betreffen. Nachfolgend konzentrieren wir uns auf die Paragraphen §60c (Wissenschaftliche Forschung) und §60d (Text und Data Mining).

Die erste dieser Schranken ist zwar restriktiver, als es sich die akademische Gemeinschaft gewünscht hatte, hat jedoch den Vorteil, dass sie ziemlich klar formuliert ist (was beim vorherigen Rahmen nicht der Fall war). Sie kann in einer Tabelle dargestellt werden (Tab. 1).

Es scheint, dass diese Ausnahme eher traditionelle, ‚konsumierende' Verwendungen abdeckt, etwa das sog. close reading. ‚Nicht konsumierende' Verwendungen, wie *distant reading*, fallen in den Geltungsbereich einer bestimmten Ausnahme (§60d, Text und Data Mining), die in Tabelle 2 zusammengefasst wird.

Möglicherweise stellen Sie sich jetzt die Frage, wie Sie zwischen kommerzieller und nichtkommerzieller Forschung unterscheiden können. Das ist in der Tat nicht immer klar zu beantworten.[9]

Es ist durchaus möglich, dass sich die akademische Community in Kürze auch nicht mehr mit der kommerziellen oder nichtkommerziellen Unterscheidung befassen muss. 2019 wurde eine neue EU-Richtlinie zum Urheberrecht im digitalen Binnenmarkt verabschiedet. Sie verpflichtet die Mitgliedstaaten unter anderem, urheberrechtliche Ausnahmen für das Data Mining für Forschungsarbeiten einzuführen, die an Universitäten und Gedächtnisorganisationen (z. B. Archive, Museen) durchgeführt werden, unabhängig davon, ob es sich um kommerzielle oder nichtkommerzielle Zwecke handelt. Dieser Schritt steht im Einklang mit der Förderung öffentlich-privater Partnerschaften durch die Europäische Kommission.

2.6 Urheberrechtslizenzierung: Creative Commons und die Open Access-Bewegung

Vermutlich haben Sie bereits von der Bewegung gehört, die den offenen Zugang (Open Access) zu wissenschaftlichem Material fördert, und Sie fragen sich möglicherweise, wie

Tab. 2: Urheberrechtsschranke für das Text und Data Mining

Welche Werke können benutzt werden?	Alle Werke ohne Einschränkungen
Bis zu welchem Umfang können Auszüge benutzt werden?	Es gibt keine Einschränkungen, die Werke können vollständig genutzt werden.
Wer kann die Werke oder Teile des Werkes benutzen?	Jede*r
Zu welchem Zweck?	Um eine Vielzahl von Werken (Ursprungsmaterial) für die nicht kommerzielle wissenschaftliche Forschung automatisiert auszuwerten.
Mit wem können die Werke oder Teile des Werkes geteilt werden?	Das Korpus kann mit einem bestimmt abgegrenzten Kreis von Personen (z. B. innerhalb eines Forschungsteams) und mit einzelnen Dritten ('soweit dies der Überprüfung der Qualität wissenschaftlicher Forschung dient') geteilt werden.
WICHTIGE EINSCHRÄNKUNG	Das Korpus und die Vervielfältigungen des Ursprungsmaterials sind nach Abschluss der Forschungsarbeiten zu löschen oder den Bibliotheken oder Archiven zur dauerhaften Aufbewahrung zu übermitteln.

9 Für Ihr Studium, für Ihre Bachelor- oder Masterarbeit ist die Situation jedoch eindeutig, Sie müssen sich keine Sorgen machen.

sie in das Bild passt, das wir gerade skizziert haben.

Die Open-Access-Bewegung ist historisch mit der Free/Open-Source-Software Movement verbunden, die in den 1980er Jahren von Richard Stallman ins Leben gerufen wurde und Software fördert, die nicht nur kostenlos ist, sondern auch von den Benutzer*innen geändert werden kann. Dieser Ansatz beeinflusste die Software-Welt, nicht zuletzt durch Erfolgsgeschichten wie Linux oder Open Office. Die gleiche Logik kann für wissenschaftliches Material gelten, das auch eine breitere Wirkung erzielen kann, wenn es allgemein zugänglich und möglicherweise sogar modifizierbar ist.

Dieser Ansatz, verbunden mit egalitären Ideen, führte zu Beginn des 21. Jahrhunderts zur Annahme von drei Gründungserklärungen der Open-Access-Bewegung, die oft als 3B bezeichnet werden: die Erklärungen von Budapest, Bethesda und Berlin. Die letzte (deren vollständiger Titel lautet: Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen), die am 22. Oktober 2003 während einer von der Max-Planck-Gesellschaft veranstalteten Konferenz unterzeichnet wurde, ist in unserem Zusammenhang von besonderem Interesse. Gemäß diesem Dokument ist eine Veröffentlichung[10] „Open Access", wenn gilt: „die Urheber und die Rechteinhaber solcher Veröffentlichungen gewähren allen Nutzern unwiderruflich das freie, weltweite Zugangsrecht zu diesen Veröffentlichungen und erlauben ihnen, diese Veröffentlichungen – in jedem beliebigen digitalen Medium und für jeden verantwortbaren Zweck – zu kopieren, zu nutzen, zu verbreiten, zu übertragen und öffentlich wiederzugeben sowie Bearbeitungen davon zu erstellen und zu verbreiten, sofern die Urheberschaft korrekt angegeben wird. (...) Weiterhin kann von diesen Beiträgen eine geringe Anzahl von Ausdrucken zum privaten Gebrauch angefertigt werden."

Open-Access-Material kann entweder von den Autor*innen selbst in einem Open-Access-Repository hinterlegt werden (der „grüne" Weg) oder gegen Bezahlung in einem Open-Access-Journal (der „goldene" Weg) veröffentlicht werden. Wir werden die organisatorischen und politischen Aspekte von Open Access hier nicht weiter betrachten.[11] Stattdessen konzentrieren wir uns auf die urheberrechtlichen Aspekte und erklären einerseits, wie Sie Ihre Arbeit als Open-Access-Publikation zugänglich machen können, und andererseits, wie Sie solche Werke rechtmäßig verwenden können.

Zuvor haben wir in diesem Kapitel erklärt, dass die Urheber*innen eines urheberrechtlich geschützten Werks einige ausschließliche Rechte an dem Werk haben (Vervielfältigung, öffentliche Zugänglichmachung, Anpassung...). Sie könne diese Rechte selbst ausüben oder anderen erlauben, dies gemäß einer Vereinbarung zu tun. Eine Lizenz für geistiges Eigentum ähnelte früher einem Mietvertrag: Es wurde in den meisten Fällen einer einzelnen Rechtsperson – häufig einem Verlag – gewährt. Open-Access-Veröffentlichungen hingegen erlauben der Öffentlichkeit, d. h. jeder und jedem,[12] diese Rechte auszuüben. Der rechtliche Rahmen hierfür ergibt sich daraus, dass die Urheber*innen formell versprechen, dass sie niemanden wegen der Verwendung ihrer Werke verklagen werden. Den potenziellen Nutzer*innen der Open-Access-Veröffentlichungen gibt dieses formale Versprechen Rechtssicherheit.

Die mit Abstand am häufigsten genutzte Familie öffentlicher Lizenzen sind die Creative-Commons (CC) -Lizenzen, die auch für Forschungsdaten gut geeignet sind. Die neueste Version 4.0 wurde 2013 veröffentlicht, frühere Versionen können jedoch weiterhin verwendet werden.

Die CC-Lizenzen sind modular aufgebaut und bestehen aus vier Bausteinen mit jeweils

10 Der Begriff Veröffentlichung umfasst ‚originäre wissenschaftliche Forschungsergebnisse ebenso wie Ursprungsdaten, Metadaten, Quellenmaterial, digitale Darstellungen von Bild- und Graphik-Material und wissenschaftliches Material in multimedialer Form'.

11 Wenn Sie an den verschiedenen Aspekten der OA-Bewegung interessiert sind, könnten Sie Suber (2012) lesen. Zudem beleuchten Graf et al. (2020) dieses Thema für den Bereich der Sozialwissenschaften.

12 Daher stammt der häufig gebrauchte Name *Jedermann-Lizenz*.

einem entsprechenden Symbol. Es wird oft gesagt, dass CC-Lizenzen den Autor*innen den Übergang von der Logik „Alle Rechte vorbehalten" zu „Einige Rechte vorbehalten" ermöglichen – jeder Baustein entspricht einem exklusiven Recht.

Die vier in der Tabelle erläuterten Bausteine (BY, SA, NC, ND) können zu sechs Lizenzen kombiniert werden, wobei BY das obligatorische Element jeder dieser Lizenzen ist. Diese Lizenzen lauten wie folgt: CC BY, BY-SA, BY-NC, BY-ND, BY-NC-ND, BY-NC-SA (beachten Sie, dass sich die ND- und SA-Anforderungen gegenseitig ausschließen und deshalb nicht kombiniert werden können).

Nicht alle dieser Lizenzen entsprechen vollumfänglich dem, was allgemein unter „Open Access" verstanden wird. Tatsächlich sind nur zwei Anforderungen – BY und SA – mit der oben genannten Definition von Open Access kompatibel. NC schränkt die Verwendungszwecke ein und ND verbietet die Herstellung von abgeleiteten Werken. Daher können Werke, die unter Lizenzen verfügbar sind, die diese Anforderungen enthalten, nicht als Open-Access-Publikationen betrachtet werden. Wenn Sie das Material, das Sie in Ihren Forschungsprojekten produzieren (z. B. Artikel, Korpora, Aufzeichnungen oder anderes), open-access unter einer CC-Lizenz veröffentlichen möchten (was unserer Meinung nach der richtige Weg ist), müssen Sie zwischen CC BY und CC BY-SA wählen. Die CC-BY-Lizenz ist sehr liberal und ermöglicht es dem Benutzer / der Benutzerin, praktisch alles mit dem so lizenzierten Werk zu machen während sichergestellt wird, dass Ihr Name als Urheber*in erwähnt wird. Sie kann daher für Forschungsarbeiten empfohlen werden. Die von Wikipedia bekannte CC BY-SA ist für Verbundprojekte interessant, da sie sicherstellt, dass niemand ihr Material (z. B. Baumbanken) zu einem kommerziellen Produkt entwickeln kann (da das modifizierte Material immer unter einer CC-BY-SA-Lizenz verfügbar bleiben muss).

Wenn Sie ein unter einer CC-Lizenz lizenziertes Werk verwenden möchten, denken Sie einfach daran, dass Ihre Verwendung (sofern sie nicht durch eine gesetzliche Ausnahme abgedeckt ist) als Urheberrechtsverletzung angesehen werden kann, wenn Sie gegen die Bestimmungen der Lizenz verstoßen.

2.7 Nicht nur das Urheberrecht: Ein paar Worte zu anderen Rechten

Das Urheberrecht ist die wichtigste Rechtsvorschrift zum Schutze des geistigen Eigentums – es ist aber nicht die einzige. Die sogenannten *Verwandten Schutzrechte* können sich auch auf Ihre Forschungsprojekte auswirken und es ist nützlich, auch darüber ein wenig zu wissen.

Wie oben erläutert, schützt das Urheberrecht originale Werke. Nicht originales Mate-

Tab. 3: CC-Lizenzen

Akronym	Bedeutung	Symbol	Beschreibung
BY	Namensnennung (Attribution)		Der Benutzer / die Benutzerin muss den Namen des Autors / der Autorin, einen Link zur verwendeten Lizenz und (falls möglich) einen Link zum Originalwerk angeben und zusätzlich angeben, ob er / sie das Werk geändert hat.
SA	Weitergabe unter gleichen Bedingungen (Share-Alike)		Wenn der Benutzer / die Benutzerin eine abgeleitete Version (Bearbeitung) erstellt (den Inhalt ändert, eine Bearbeitung oder Übersetzung erstellt), muss diese unter derselben Lizenz wie das Original freigegeben werden (d. h. sie muss offen bleiben).
NC	Nicht kommerziell (Non-Commercial)		Der Benutzer / die Benutzerin kann das Werk nicht für kommerzielle Zwecke nutzen.
ND	Keine Bearbeitung (No Derivatives)		Der Benutzer / die Benutzerin kann keine abgeleiteten Werke basierend auf dem Werk erstellen (d. h. er / sie kann das Werk nicht modifizieren, übersetzen oder bearbeiten); Die Arbeit kann nur „wie sie ist" verwendet werden.

rial, z. B. die Zusammenstellung einer Anthologie, kann durch andere Rechte geschützt werden. Datenbanken (Zusammenstellungen von Daten, die allgemein verstanden werden, einschließlich beispielsweise Websites oder Karten) sind durch das Datenbankherstellerrecht geschützt, sofern eine wesentliche Investition für die Beschaffung, die Überprüfung oder die Darstellung ihres Inhalts vom Datenbankbetreiber geleistet wurde. Dieses Recht besteht EU-weit und verhindert, dass wesentliche Teile geschützter Datenbanken ohne Erlaubnis ihres Herstellers (d. h. des Unternehmens, das in ihre Erstellung investiert hat) kopiert und weitergegeben werden. Die Dauer des Rechts beträgt 15 Jahre nach der Erstinvestition, wird jedoch nach jeder weiteren wesentlichen Investition erneuert.

Ein anderes verwandtes Recht schützt die wissenschaftlichen Ausgaben, d. h. neue Ausgaben gemeinfreier Werke (z. B. mittelalterliche Poesie), die sich erheblich von zuvor bekannten Ausgaben unterscheiden und aus einer wissenschaftlichen Tätigkeit resultieren (z. B. Verwendung eines kritischen Apparats). Das Recht liegt bei dem Verfasser / der Verfasserin der Ausgabe und erlischt 25 Jahre nach ihrer Veröffentlichung. Dieses Recht besteht nur in einigen EU-Mitgliedstaaten und wurde auf EU-Ebene nicht harmonisiert. Daher ist es möglich, dass eine in Deutschland geschützte wissenschaftliche Ausgabe in anderen EU-Staaten nicht geschützt ist.

Ein eher überraschendes Recht schützt Lichtbilder, d. h. Bilder, die mit der gleichen Technologie wie Fotografie aufgenommen wurden, aber das Kriterium für Originalität nicht erfüllen. Es gilt für Fotos, die nicht die Kriterien einer persönlich-geistigen Schöpfung erfüllen. Auch diese sind quasi-urheberrechtlich geschützt, allerdings nur die ersten 15 Jahre nach ihrer Herstellung. Beispiele für nicht originelle Fotos sind Satellitenfotos, Fotos, die in einer Fotokabine aufgenommen wurden, viele Selfies und andere Bilder, die mit Smartphones aufgenommen wurden oder auch die bereits erwähnten Fotos des Eiffelturms. Dieses Recht, das nur in wenigen Ländern Europas besteht, könnte zum Zeitpunkt der Verabschiedung des UrhG (1965) wirtschaftlich gerechtfertigt gewesen sein, da für das Erstellen von Fotos einige finanzielle und technische Investitionen erforderlich waren.

Glücklicherweise gelten die Urheberrechtsschranken auch für die verwandten Schutzrechte. Darüber hinaus umfasst die neueste Version der CC-Lizenzen (4.0 – siehe oben) nicht nur das Urheberrecht, sondern auch alle verwandten Schutzrechte. Wenn Sie CC-lizenzierte Inhalte verwenden oder wenn Ihre Verwendung mit einer Forschungsausnahme kompatibel ist, müssen Sie sich daher keine Sorgen um verwandte Schutzrechte machen.

3. Personenbezogene Daten

3.1 Einführung

Vermutlich haben Sie schon von der Datenschutz-Grundverordnung (DSGVO) gehört. Die DSGVO regelt den Schutz personenbezogener Daten und kommt in der EU seit 2018 zur Anwendung.

In Deutschland, einem Land, das Pionier des Datenschutzes in der EU ist, wurde 1977 das erste Bundesdatenschutzgesetz (BDSG) verabschiedet. Das deutsche Gesetz war eine Inspiration für die EU-Richtlinie über personenbezogene Daten von 1995. Zum größten Teil basiert die DSGVO (Allgemeine Datenschutzverordnung) auf denselben Grundsätzen wie die Richtlinie, die sie ersetzt hat. Der wohl revolutionärste Teil der DSGVO ist, dass die Verwaltungsstrafen für Verstöße erheblich erhöht wurden, was die Aufmerksamkeit der Unternehmen – und der Öffentlichkeit – auf sich zog.

In den folgenden Abschnitten werden das Konzept der personenbezogenen Daten (3.2), einige der allgemeinen Grundsätze für die Verarbeitung personenbezogener Daten (3.3) und die Rechte der betroffenen Personen (3.4) vorgestellt.

3.2 Das Konzept der personenbezogenen Daten

Entgegen einer verbreiteten Meinung sind personenbezogene Daten nicht nur Nominativdaten (Vor- und Nachname), Kontaktdaten (Adresse, E-Mail, Telefonnummer) oder andere Informationen, die sich z. B. im Personalausweis finden, wie u. a. Geburtsdatum und -ort. Der Begriff ist viel weiter gefasst und umfasst alle Informationen, die sich auf eine identifizierte oder auch nur auf eine identifizierbare Person beziehen. Dies würde alle oben genannten Informationen über die Person einschließen, aber auch Sprachaufnahmen und Bilder von Personen erlauben heutzutage eine Identifizierung.

Eine Person ist zudem identifizierbar, wenn sie mit „allen Mitteln identifiziert werden kann, die vernünftigerweise verwendet werden können", einschließlich Querverweise auf andere verfügbare Informationen. Das bedeutet praktisch, dass Sie personenbezogene Daten bereits dann nutzen, wenn Ihnen die Daten, die Sie über eine Person haben, ermöglichen, Informationen über diese Person im Internet zu finden. Es wird dabei nicht unterschieden, ob sich die Daten der Person auf „öffentliche" oder „private" Lebensbereiche eines betroffenen Menschen beziehen. Die Informationen über den Beruf, die Ausbildung oder die Karriere eines Menschen stellen ebenso wie sein Familienstand und seine familiäre Situation personenbezogene Daten dar.

Bestimmte Kategorien personenbezogener Daten genießen jedoch besonderen Schutz. Dies ist der Fall bei Informationen, die Auskunft geben über die ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschaftsmitgliedschaften, genetische Daten, biometrische Daten, wenn sie zur eindeutigen Identifizierung der Person verwendet werden können, Daten zum Gesundheitszustand, zur sexuellen Orientierung[13] sowie Daten zu strafrechtlichen Verurteilungen und Straftaten.[14]

Es ist möglich, die Verbindung zwischen den Informationen und der Person zu lösen – der Prozess wird als „Anonymisierung" bezeichnet. Anonymisierte Daten können nicht mehr mit einer Person in Verbindung gebracht werden, zumindest nicht mit Mitteln, die mit vernünftiger Wahrscheinlichkeit verwendet werden. Diese Daten sind daher keine personenbezogenen Daten und können frei verarbeitet werden. Die Anonymisierung ist eher ein technisches als ein rechtliches Problem und wird in diesem Kapitel nicht weiter behandelt.[15] Sie sollten sich jedoch bewusst sein, dass Anonymisierung viel mehr ist als nur das Entfernen von Vor- und Nachnamen. Eine Anonymisierung ist zudem irreversibel und unterscheidet sich hierdurch von der Pseudonymisierung, bei der personenbezogene Daten reversibel durch Pseudonyme ersetzt werden (d. h. die Information, dass „A001" tatsächlich Herr Peter Müller aus Berlin ist, bleibt erhalten). Pseudonymisierte Daten sollten weiterhin als personenbezogene Daten betrachtet werden.

3.3 Ausgewählte Grundsätze für die Verarbeitung personenbezogener Daten

Die Verarbeitung personenbezogener Daten sollte einer Reihe von Grundsätzen entsprechen, die in Artikel 5 der DSGVO aufgeführt sind. In erster Linie muss die Verarbeitung rechtmäßig sein, d. h. auf einem der in Artikel 6 aufgeführten rechtlichen Gründe (d. h. einem „gültigen Grund") beruhen. Einer dieser Gründe ist die Einwilligung der Person, auf die sich die Daten beziehen (der betroffenen Person). Nach der DSGVO muss die Einwilligung nicht schriftlich erfolgen, sollte jedoch freiwillig (d. h. die Person sollte eine echte Wahl haben, ohne befürchten zu müssen, im

13 Artikel 9 DSGVO.

14 Artikel 10 DSGVO.

15 Weitere Informationen über Anonymisierung finden Sie in der Stellungnahme 5/2014 zu Anonymisierungstechniken vor Artikel-29-Datenschutzgruppe (https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_de.pdf).

Falle einer Ablehnung negative Folgen zu haben), für den bestimmten Fall (d. h. in Bezug auf einen bestimmten Verarbeitungszweck, nicht nach dem Motto „Sie können mit meinen Daten machen, was Sie wollen"), in informierter Weise (siehe unten) und unmissverständlich (ausgedrückt in einer klaren positiven Handlung, z. B. Aktivieren eines Kontrollkästchens) abgegeben werden. Die Einwilligung ist jedoch nicht der einzig mögliche rechtliche Grund. Eine weitere Option, die insbesondere für Forschungszwecke relevant ist, ist die Verarbeitung personenbezogener Daten auf der Grundlage von „berechtigten Interessen". Diese Grundlage erfordert jedoch einen dokumentierten „Abwägungstest", d. h. das Abwägen der „berechtigten Interessen" gegen die Interessen der betroffenen Person.[16]

Die Verarbeitung muss auch transparent sein. Dies bedeutet, dass der betroffenen Person bei jeder Verarbeitung personenbezogener Daten Informationen über unter anderem die für die Verarbeitung zuständige Person oder Einrichtung (für die Verarbeitung Verantwortlicher), den Zweck der Verarbeitung, die Kategorien von Daten, die verarbeitet werden sowie die Informationen über seine Rechte in Bezug auf die Verarbeitung zur Verfügung gestellt werden sollten. Diese Verpflichtung gilt unabhängig davon, ob die Daten direkt von der betroffenen Person stammen oder nicht (z. B. wenn sie aus dem Internet heruntergeladen wurden). Jetzt verstehen Sie sicherlich besser, warum Sie beim Erstellen eines Kontos auf einer Website das Kontrollkästchen aktivieren müssen, das besagt, dass Sie die Datenschutzrichtlinie akzeptieren.

Ein weiteres Prinzip, das die Verarbeitung personenbezogener Daten regelt, ist die Zweckbindung. Nach diesem Prinzip können Daten nur für festgelegte, eindeutige und legitime Zwecke (über die die betroffene Person informiert werden soll) verarbeitet werden, beispielsweise für Ihr Forschungsprojekt, und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden.

Datenminimierung ist ein Prinzip, das mit datenintensiver Forschung völlig unvereinbar ist. Nach diesem Prinzip sollte die Verarbeitung auf die Daten beschränkt sein, die zur Erreichung des vordefinierten Zwecks unbedingt erforderlich sind. In der datenintensiven Forschung gibt es natürlich keine „unnötigen Daten". Bei der Verarbeitung personenbezogener Daten in Ihrem Forschungsprojekt sollten Sie sich jedoch bemühen, dieses Prinzip zu respektieren, indem Sie irrelevante Informationen entfernen (müssen Sie z. B. das Alter, den Beruf oder die Herkunft der Person kennen, um Ihre Fragestellung zu beantworten?).

Personenbezogene Daten sollten auch sachlich richtig und erforderlichenfalls auf dem neuesten Stand sein. Dieser Grundsatz wird durch das Recht der betroffenen Person geschützt, die Berichtigung ungenauer Daten zu verlangen (siehe unten, Abschnitt 3.4).

Personenbezogene Daten sollten auch so verarbeitet werden, dass eine angemessene technische und organisatorische Sicherheit gewährleistet und vor unbefugtem Zugriff oder versehentlichem Verlust, Zerstörung oder Beschädigung geschützt ist. Sie sollten keine persönlichen Daten auf einem USB-Stick mit sich herumtragen oder in einen kommerziellen Cloud-Dienst hochladen! Stattdessen ist die Cloud Ihrer Universität ein sicherer Ort.

Last but not least sollte die für die Verarbeitung verantwortliche Person in der Lage sein, die Einhaltung aller oben genannten Grundsätze nachzuweisen. Im Falle einer Untersuchung durch die Datenschutzbehörde ist es Sache der für die Verarbeitung Verantwortlichen nachzuweisen, gesetzeskonform gehandelt zu haben (!). Aus diesem Grund ist es von größter Bedeutung, alle Ihre Verarbeitungsvorgänge mit personenbezogenen Daten zu dokumentieren.

16 Weitere Informationen finden Sie in der Stellungnahme 06/2014 zum Begriff des berechtigten Interesses des für die Verarbeitung Verantwortlichen gemäß Artikel 7 der Richtlinie 95/46/EG vor Artikel-29-Datenschutzgruppe (https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_de.pdf).

3.4 Rechte der betroffenen Personen

Die betroffenen Personen haben bestimmte Rechte, wenn es um ihre persönlichen Daten geht. Diese Rechte sollten klar von den im ersten Teil dieses Kapitels vorgestellten Rechten an geistigem Eigentum unterschieden werden, da sie nicht wirtschaftlicher Natur sind und nicht Teil von Handelsgeschäften sein können (d. h. man kann nicht für ein Versprechen bezahlt werden, keine Rechte an seinen persönlichen Daten auszuüben). Einige dieser Rechte werden abschließend vorgestellt:

Das wohl wichtigste Recht der betroffenen Personen ist das Auskunftsrecht.[17] Wenn jemand Ihre Daten verarbeitet, können Sie jederzeit Informationen zur Verarbeitung und eine Kopie der Daten anfordern. Eng verbunden mit dem Zugriffsrecht ist das Recht auf Berichtigung, d. h. das Recht, die Berichtigung oder Löschung ungenauer Daten zu verlangen.[18]

Darüber hinaus hat die betroffene Person immer das Recht, eine Einwilligung zu widerrufen,[19] wenn die Daten auf der Grundlage einer Einwilligung verarbeitet werden, ohne die Entscheidung begründen zu müssen. Die Auswirkungen des Widerrufs sind jedoch nicht rückwirkend – die vor dem Widerruf durchgeführten Vorgänge werden nicht rechtswidrig, es ist jedoch keine weitere Verarbeitung möglich (es sei denn, es liegt ein anderer Rechtsgrund vor).

4. Fazit

Grundkenntnisse zur Rechtslage bezüglich des Schutzes geistigen Eigentums und insbesondere des Urheberrechts sind heutzutage für Forschende und Studierende der Sprachwissenschaften erforderlich, wenn sie mit selbst erstellten oder erhobenen Daten arbeiten oder Daten Dritter (z. B. Korpora) nutzen. Bei der Arbeit mit Sprachdaten ist häufig das Urheberrecht relevant. Moderne Texte, egal aus welchen Quellen, dürfen nicht ohne Erlaubnis verwendet (z. B. in ein Korpus integriert) werden, es sei denn, eine solche Verwendung wird durch gesetzliche Ausnahmetatbestände ermöglicht.

Ein grundlegendes Verständnis des Urheberrechts ist aber auch nützlich, wenn Sie Ihre Werke unter Open-Access-Bedingungen veröffentlichen möchten oder sollen. Ebenso wichtig ist es zu verstehen, dass Sprachdaten auch als personenbezogene Daten gelten können, wodurch ihre Verarbeitung in einen engen rechtlichen Rahmen erfolgen muss, der von der DSGVO festgelegt wurde.

Zum Weiterlesen

Wenn Sie diese Themen interessant finden, enthält das Literaturverzeichnis weitere Informationen für Sie. Die jüngste Reform des deutschen Urheberrechts wird von Hoeren (2018), die Bedeutung von Open Access im Kontext der akademischen Forschung von Baldwin (2018) und die gesetzliche Ausnahme für Text- und Data Mining von Kamocki et al. (2018) erörtert. Vielleicht möchten Sie auch einige Texte der Artikel-29-Datenschutzgruppe oder den Leitfaden des European University Institute zur guten Datenschutzpraxis in der Forschung lesen.

17 Artikel 15 DSGVO.
18 Artikel 16 DSGVO.
19 Artikel 7.3 DSGVO.

Literatur

Artikel-29-Datenschutzgruppe, Stellungnahme 4/2007 zum Begriff „personenbezogene Daten", [online] https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_de.pdf.

Artikel-29-Datenschutzgruppe, Stellungnahme 06/2014 zum Begriff des berechtigten Interesses des für die Verarbeitung Verantwortlichen gemäß Artikel 7 der Richtlinie 95/46/EG [online] https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_de.pdf.

Artikel-29-Datenschutzgruppe, Stellungnahme 5/2014 zu Anonymisierungstechniken vor Artikel-29-Datenschutzgruppe, [online] https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_de.pdf.

Baldwin, Peter (2018): Why are Universities Open Access Laggards? Bulletin of the German Historical Institute, 63 (Fall 2018), S. 67, [online] https://history.ucla.edu/sites/default/files/why_are_universities_open_access_laggards.pdf .

Beißwenger, Michael, Harald Lüngen, Jan Schallaböck, John H. Weitzmann, Axel Herold, Pawel Kamocki, Angelika Storrer und Julia Wildgans (2017): Rechtliche Bedingungen für die Bereitstellung eines Chat-Korpus in CLARIN-D. Ergebnisse eines Rechtsgutachtens, in: Beißwenger, Michael (Hrsg.), *Empirische Erforschung internetbasierter Kommunikation* (=Empirische Linguistik / Empirical Linguistics, Band 9), Berlin: de Gruyter, S. 7–46, [online] https://www.degruyter.com/view/title/534368.

Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen vom 22. Oktober 2003 [online] https://openaccess.mpg.de/68053/Berliner_Erklaerung_dt_Version_07-2006.pdf.

Berner Übereinkunft zum Schutz von Werken der Literatur und Kunst (Pariser Fassung: BGBl. 1973 II S. 1069, geändert am 2. Oktober 1979: BGBl. 1985 II S. 81).

European University Institute, Guide on Good Data Protection Practice in Research, 2019 [online] https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf.

Graf, Dorothee, Yuliya Fadeeva und Katrin Falkenstein-Feldhoff (Hrsg.) (2020): *Das Thema „Bücher im Open Access. Ein Zukunftsmodell für die Geistes- und Sozialwissenschaften?* Opladen/Berlin/Toronto: Verlag Barbara Budrich, [online] https://duepublico2.uni-due.de/receive/duepublico_mods_00072237.

Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz, UrhG) vom 9. September 1965 (BGBl. I S. 1273).

Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft (Urheberrechts-Wissensgesellschafts-Gesetz – UrhWissG) vom 1. September 2017 (BGBl. I S. 3346).

Hoeren, Thomas (2018): Das Urheberrechts-Wissensgesellschafts-Gesetz – Ein richtiger Schritt für das Urheberrecht in Bildung und Forschung? IWRZ3/2018, S. 120, [online] https://www.itm.nrw/wp-content/uploads/Das-Urheberrechts-Wissensgesellschafts-Gesetz.pdf.

Kamocki, Pawel, Erik Ketzan, Julia Wildgans und Andreas Witt (2018): New exceptions for Text and Data Mining and their possible impact on the CLARIN infrastructure, in: *Selected papers from the CLARIN Annual Conference 2018*. Linköping Electronic Conference Proceedings 159, S. 66, [online] https://ep.liu.se/ecp/159/007/ecp18159007.pdf.

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, [online] http://data.europa.eu/eli/dir/2019/790/2019-05-17.

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 (Informationsgesellschaftsrichtlinie), [online] http://data.europa.eu/eli/dir/2001/29/oj.

Suber, Peter (2012): *Open Access*. MIT Press [online] https://cyber.harvard.edu/hoap/Open_Access_(the_book).

Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung, DSGVO), [online] https://eur-lex.europa.eu/eli/reg/2016/679/oj.

Die Adressen aller Webseiten und Online-Ressourcen in diesem Beitrag wurden zuletzt auf Aktualität überprüft am 15. Juni 2021.