Rechtsfragen bei der Nutzung und Weitergabe linguistischer Daten*

Timm Lehmberg, Christian Chiarcos, Georg Rehm, Andreas Witt

Zusammenfassung

Grundlage dieses Artikels[1] ist das Verbundprojekt „Nachhaltigkeit linguistischer Daten" der drei Sonderforschungsbereiche 441, 538 und 632, dessen Ziel es ist, Lösungen für die nachhaltige Verfügbarkeit der an den SFBs vorhandenen Korpora zu entwickeln. Ein zentraler Aspekt betrifft die Klärung der Rechtslage für die Nutzung und Weitergabe linguistischer Ressourcen, die durch das Urheber- sowie das Datenschutzrecht geschützt sind.

Eine als indifferent wahrgenommene rechtliche Situation wird in der Praxis oft als das entscheidende Hindernis für die Weitergabe linguistischer Daten angeführt. Tatsächlich jedoch sind Nutzung und Weitergabe von Daten zu wissenschaftlichen Zwecken normativ geregelt. Problematisch ist oftmals die Einordnung der speziellen linguistischen Daten als Schutzgegenstand sowie die Tatsache, dass an linguistische Daten und Datensammlungen aufgrund ihrer komplexen und vielschichtigen Beschaffenheit durchaus mehrere Urheber Rechte besitzen können, die sich auf verschiedene Inhalte beziehen.

Der Beitrag gibt einen Überblick über das geltende Recht sowie die juristischen und natürlichen Personen, die potentiell Rechte an linguistisch aufbereiteten Datenkollektionen besitzen. Es ist nicht Gegenstand dieses Artikels, rechtsverbindliche Aussagen zu treffen, die auf eine Nutzung und Weitergabe jedweder Daten angewandt werden. Der Artikel orientiert sich in seiner Struktur und thematischen Tiefe bewusst nicht an einem juristischen Publikum, sondern beschreibt die Problematik aus geisteswissenschaftlicher Perspektive.

Zusammen mit einem Überblick über das vom Umgang mit linguistischen Datensammlungen betroffene Recht, das Urheberrechtsgesetz (Abschnitt 1) und das Bundesdatenschutzgesetz (Abschnitt 2), wird in den jeweiligen Abschnitten auch eine Klassifikation der Daten aus juristischer Sicht vorgenommen. Anschließend werden Lösungsansätze vorgestellt, die im Rahmen des o. g. Verbundprojektes erarbeitet werden (Abschnitt 3).

1 Urheberrecht

Als zentrale Einwände gegen die Nutzung und Weitergabe linguistischer Daten werden zumeist urheberrechtliche Interessen an den Primärtexten genannt (vgl. Jüttner, 2000, Patzelt, 2003).

* Erschienen in: *Datenstrukturen für linguistische Ressourcen und ihre Anwendungen*, Georg Rehm, Andreas Witt, Lothar Lemnitzer (Hrsg.), Tübingen: Gunter Narr Verlag. 2007. Seite 93–102.

[1] Die Autoren bedanken sich bei Dipl.-Jur. Felix Zimmermann (Institut für Rechtsinformatik, Universität Hannover) für zahlreiche sachkundige Hinweise.

Ausschlaggebend hierfür ist, dass aus finanziellen sowie arbeitspraktischen Erwägungen bei der Korpuserstellung oft auf bereits digitalisierte Ressourcen (Zeitungstexte, literarische Werke etc.) zugegriffen wird, an denen naturgemäß die Rechte von Autoren, Editoren etc. bestehen. Obwohl das deutsche Urheberrechtsgesetz (UrhG) eine Nutzung und Weitergabe zum Zweck der wissenschaftlichen Forschung unter bestimmten Bedingungen zulässt (vgl. Abschnitt 1.4), stellt eine uneingeschränkte Weitergabe dieser Daten eine Verletzung des Urheberrechtes dar. Nur selten wird hinsichtlich der Urheberschaft des wissenschaftlichen Personals argumentiert, dass die Aufbereitung der Korpora selbst, also die Entwicklung von Datenstrukturen, Annotation etc. vorgenommen hat.

Der nachfolgende Abschnitt, der sich mit urheberrechtlichen Fragen der Nutzung und Weitergabe linguistischer Daten auseinandersetzt, führt zunächst in das deutsche Urheberrechtsgesetz ein. Darauf aufbauend erfolgt eine Klassifizierung linguistischer Korpora (bzw. Teilen davon, wie Annotation, Zugangssoftware etc.) aus juristischer Sicht sowie die damit verbundene Benennung der jeweiligen Rechteinhaber. Anschließend werden besondere Regelungen des Urheberrechts für Forschung und Lehre sowie deren Geltungsbereich dargestellt.

1.1 Das Werk

Das deutsche Urheberrecht ist (wie z. B. das Patentrecht) Teil des *Immaterialgüterrechtes*. Hierbei handelt es sich um den Oberbegriff für alle Regelungen, die dem Schutz nicht-materiellen geistigen Eigentums, also Ideen, Patenten, Gebrauchs- und Geschmacksmustern sowie künstlerischen Werken dienen. Anders als ein Patent muss eine Urheberschaft nicht offiziell angemeldet werden, da sie automatisch bei der Schaffung eines Werkes entsteht.

Das Urheberrecht ist nicht zu verwechseln mit dem angloamerikanischen *Copyright*, das eine stärkere ökonomische Ausrichtung besitzt, da es vornehmlich die hinter den Werken stehenden Investitionen schützen soll. Beim Urheberrecht hingegen steht der Schutz des Schaffenden selbst (Autor, Künstler etc.) im Vordergrund. Es schützt den Urheber in seinem Persönlichkeitsrecht und seinen wirtschaftlichen Interessen in Bezug auf ein von ihm geschaffenes *Werk*.

Der Werksbegriff ist von entscheidender Bedeutung, weil er eine starke Einschränkung hinsichtlich der Schutzfähigkeit mit sich bringt. Als Werk im Sinne des UrhG gelten ausschließlich persönliche geistige Schöpfungen, denen eine gestalterische Tätigkeit zugrunde liegt und die durch ihre Form bzw. ihren Inhalt eine Neuschöpfung darstellen. Das UrhG nennt in diesem Kontext „Schöpfungen der Literatur, Wissenschaft und Kunst“. Dazu zählen Sprach- und Schriftwerke, ebenso wie Sammel- und Datenbankwerke (vgl. § 2 ff UrhG). Entscheidendes Maß für die Einstufung als Werk im Sinne des UrhG ist die *Schöpfungshöhe* (auch: *Gestaltungshöhe*). Gemeint ist der „Grad der Individualität, den ein geistiges Erzeugnis besitzen muss, um eine persönliche geistige Schöpfung nach § 2 Abs. 2 UrhG zu sein.“ (Wandtke / Bullinger / *Bullinger* § 2, UrhG, Rn. 23–25). Dabei handelt es sich „um den quantitativen Gesichtspunkt der Individualität des Werkes“ (ebd.). Ist ein gewisses Maß der Gestaltungshöhe nicht gegeben, so besteht kein urheberrechtlicher Schutz. Die hier formulierten Anforderungen dienen dazu, Alltagserzeugnisse (standardisierte Briefe, einfache Überarbeitungen etc.) von künstlerischen Werken abzugrenzen (ebd.). In der Praxis ist eine objektive Einschätzung im Hinblick auf das Kriterium der Schöpfungshöhe jedoch keineswegs immer möglich, da ihr stets eine subjektive Betrachtung des Werkes zugrunde liegt (vgl. Rehbinder, 2005, Rn. 49). Für eine Einstufung

korpuslinguistischer Bearbeitungen (z. B. Codierung, Annotation, Transkriptionsarbeit etc.) im Sinne des UrhG stellt sie jedoch eine wichtige Herausforderung dar, da sie maßgeblich für die Einstufung von wissenschaftlichem Personal als Rechteinhaber an den Forschungsdaten ist.

Ebenfalls schutzrechtfähig nach dem UrhG sind Bearbeitungen eines geschützten Werkes, sofern es sich dabei um eigene geistige Schöpfungen des Bearbeiters handelt. Der Schutz des Ausgangswerkes bleibt von der Bearbeitung unberührt (§ 3 UrhG). In diesem Zusammenhang ist der in § 70 UrhG geregelte *Schutz wissenschaftlicher Ausgaben* zu nennen, der aus sprachwissenschaftlicher Perspektive für den Umgang mit wissenschaftlichen Editionen historischer (also nicht mehr urheberrechtlich geschützter) Texte relevant ist. Demnach genießen wissenschaftliche Editionen von nicht mehr geschützten Texten einen urheberrechtlichen Schutz. Dieses Recht steht dem Verfasser der Ausgabe zu und erlischt 25 Jahre nach Erscheinen der Ausgabe.

Von großer Bedeutung für den Schutz linguistischer Datensammlungen ist der durch das UrhG formulierte *Schutz von Sammelwerken und Datenbankwerken*. Als Sammlungen gelten hierbei „Sammlungen von Werken, Daten oder anderen unabhängigen Elementen, die aufgrund der Auswahl oder Anordnung der Elemente eine persönliche geistige Schöpfung sind“ (§ 4 Abs. 1 UrhG). Datenbankwerk hingegen ist „ein Sammelwerk, dessen Elemente systematisch oder methodisch angeordnet und einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sind“ (§ 4 Abs. 2 UrhG). Die oben erwähnte Schöpfungshöhe wird in diesem Fall also nicht durch eine kreative Leistung des Urhebers, sondern durch die systematische und methodische Zusammenstellung bzw. Aufbereitung der Daten erreicht; zum *sui generis* Schutz für Datenbankwerke siehe auch Abschnitt 1.6. Der urheberrechtliche Schutz an Sammlungen bzw. Datenbanken besteht hier ungeachtet der Urheberrechte an deren Inhalten.

1.2 Der Urheber

Als Urheber gilt der Schöpfer eines Werkes. Für den Fall, dass mehrere Personen an der Schaffung eines Werkes beteiligt sind, ohne dass sich ihre Anteile gesondert verwerten lassen, werden sie als gleichberechtigte *Miturheber* behandelt (§ 8 Abs. 1 UrhG). Dem Schutz der Person des Urhebers kommt im deutschen Urheberrecht eine besondere Bedeutung zu. Im Gegensatz zum angloamerikanischen Rechtsraum kennt das deutsche Urheberrecht keine Übertragung der originären Urheberrechte an Dritte. Eine solche Übertragung ist ebenfalls beim Verkauf des Werkes nicht möglich. Der Schöpfer kann nach deutschem Recht jedoch Lizenzen einräumen, das von ihm geschaffene Werk in einer bestimmten Art und Weise zu nutzen. Dieses Recht gilt sowohl für ihn als auch für seine Erben bis zu 70 Jahre nach dem Tod des Urhebers.

1.3 Verwertungsrechte

Das Verwertungsrecht (§§ 15–24 UrhG) umfasst in erster Linie Regelungen hinsichtlich der Vervielfältigung, Verbreitung und Ausstellung sowie der öffentlichen Wiedergabe (d. h. Aufführung, Zugänglichmachung im Internet, Wiedergabe durch Bild- und Tonträger etc.). Aus Gründen der Übersichtlichkeit wird an dieser Stelle auf eine vollständige Darstellung der Verwertungsrechte verzichtet. Stattdessen werden in den folgenden Abschnitten die für die öffentliche Zugänglichmachung und Publikation linguistischer Daten relevanten Aspekte des Verwertungsrechts in Verbindung mit konkreten Beispielen wiedergegeben.

1.4 Schranken des Urheberrechts

Ein von Laien häufig missverständlich interpretierter juristischer Fachbegriff ist die Bezeichnung „Schranken des Urheberrechtes" (§§ 44 a–63 a UrhG). Hierbei handelt es sich nicht um Schranken, die das UrhG für die Verwertung und Nutzung geschützter Werke auferlegt, sondern um Einschränkungen der oben dargestellten ausschließlichen Verwertungsrechte des Urhebers. Schrankenbestimmungen räumen für bestimmte Bereiche spezielle Formen der Nutzung ein, die auch ohne Einwilligung des Urhebers erfolgen können. Sie beziehen sich dabei in erster Linie auf die Nutzungsformen, die der Allgemeinheit bzw. der Kulturwirtschaft dienen. Mit Blick auf die Ausgangsfrage des vorliegenden Artikels, die Nutzung und Weitergabe linguistischer Ressourcen, werden im Folgenden lediglich die für die private und wissenschaftliche Nutzung relevanten Schrankenregelungen wiedergegeben.

Besonders interessant für die Publikation von Sprachdaten, Belegen, Korpora etc. ist die in § 52 a Abs. 1 UrhG geregelte *Öffentliche Zugänglichmachung für Unterricht und Forschung*. Sie erlaubt, „veröffentlichte kleine Teile eines Werkes, Werke geringen Umfangs sowie einzelne Beiträge aus Zeitungen oder Zeitschriften zur Veranschaulichung im Unterricht an Schulen, Hochschulen, [...]" öffentlich zugänglich zu machen, wobei dies „ausschließlich für den bestimmt abgegrenzten Kreis von Unterrichtsteilnehmern" erfolgen darf. Dies ist ebenfalls zulässig „für einen bestimmt abgegrenzten Kreis von Personen für deren eigene wissenschaftliche Forschung". Bedingung ist hierbei jedoch, dass die Zugänglichmachung tatsächlich zu diesem Zweck erforderlich ist und nicht zum Zweck einer kommerziellen Nutzung erfolgt. Das UrhG beinhaltet weitere Schrankenregelungen, die Relevanz für Forschung und Lehre besitzen, beispielsweise zum Zitatrecht (§ 51 UrhG), zu den zur öffentlichen Zugänglichmachung erforderlichen Vervielfältigungen (§ 52 a Abs. 3 UrhG) oder Vervielfältigungen zum privaten, wissenschaftlichen und sonstigen eigenen Gebrauch (§ 53 Abs. 2, 3 UrhG), die jedoch an dieser Stelle außer Acht gelassen werden müssen.

Die hier angeführten Schrankenbestimmungen räumen zwar der wissenschaftlichen Nutzung gewisse Privilegien ein, diese unterliegen jedoch Einschränkungen hinsichtlich des Personenkreises und der Größe des Werkes (bzw. der Ausschnitte davon), die aufgrund der Verwendung auslegungsbedürftiger Rechtsbegriffe nicht explizit quantitativ festgelegt sind. So stellen „kleine Teile eines Werkes" in der Praxis 15% des Gesamtwerkes dar (vgl. Beger, 2003, S. 152), ein eingeschränkter Personenkreis ist beispielsweise im Rahmen eines Forschungsseminars gegeben, nicht jedoch einer offenen Vorlesungsveranstaltung. Ein im Prinzip offener Kreis von Nutzern, die sich online für die Nutzung einer Ressource registriert haben, gilt ebenfalls nicht als eingeschränkter Personenkreis.

Die Interpretation von § 52 a UrhG ist seit seiner (ursprünglich auf drei Jahre befristeten) Einführung im Rahmen der Änderungen des „1. Korbs" im Jahr 2003 Gegenstand zahlreicher Diskussionen. Die Streitigkeiten reichen von Fragen der Auslegung bis hin zu Forderungen nach Abschaffung der Norm. Das Aktionsbündnis „Urheberrecht für Bildung und Wissenschaft"[2] fordert z. B. in seiner *Göttinger Erklärung zum Urheberrecht für Bildung und Wissenschaft vom 5. Juli 2004*: „[...] die notwendige Rechtssicherheit herzustellen, indem die für

[2] Bei diesem Bündnis handelt es sich um einen Zusammenschluss aus verschiedenen deutschen Wissenschaftsorganisationen, Verbänden und Institutionen, u. a. der Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V., der Max-Planck-Gesellschaft sowie verschiedenen Hochschulen und Bibliotheken.

Laien nur noch schwer verständlichen und selbst für Juristen kaum verlässlich zu interpretierenden Schrankenbestimmungen im UrhG (insb. §§ 52 a, 53) klar und nachvollziehbar formuliert werden." Zum Zeitpunkt der Entstehung dieses Artikels besitzt § 52 a UrhG eine Befristung bis zum 31.12.2008, die im Juli 2006 vom Deutschen Bundestag beschlossen wurde.

1.5 Linguistische Daten und Urheberrecht

Der folgende Abschnitt unternimmt auf der Grundlage der vorangegangenen Einführung in die Grundzüge des deutschen Urheberrechts eine Einordnung linguistischer Daten aus juristischer Sicht. Zu diesem Zweck werden Korpora, Primärdaten, Sekundärdaten sowie Computerprogramme in separaten Abschnitten behandelt.

Linguistische Korpora als Datenbanken

Als linguistische Korpora werden im Rahmen dieses Artikels Sammlungen digitalisierter (gesprochener wie geschriebener) Sprachdaten verstanden, die zum Zweck der sprachwissenschaftlichen Forschung zusammengestellt und aufbereitet (kodiert, annotiert etc.) wurden. Sie gelten damit aus juristischer Sicht als *Datenbanken*, die hinsichtlich ihres urheberrechtlichen Schutzes eine besondere Stellung einnehmen.[3] Den Rahmen für diese Regelung liefert die *Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11. März 1996 über den rechtlichen Schutz von Datenbanken*.

Anders als bei dem in Abschnitt 1.1 beschriebenen Werksbegriff legitimiert sich der Schutz von Datenbanken nicht aus der Schöpfungshöhe und der innovativen Leistung eines oder mehrerer Urheber, sondern aus dem logistischen Aufwand heraus, der zur Erstellung einer Datenbank notwendig ist. Die oben genannte EU-Richtlinie schaffte 1996 hierzu ein *sui generis Recht*, das diese Leistungen urheberrechtlich schützt. Der Schutz einer Datenbank bezieht sich grundsätzlich auf das Gesamtwerk und niemals auf Teile davon.

Um dem hohen Maß an personellem und finanziellem Aufwand Rechnung zu tragen, der zur Erstellung einer Datenbank erforderlich ist, zerfällt der Schutz von Datenbanken in Leistungs- und Investitionsschutz. *Leistungsschutz* besteht, wenn die Erstellung der Datenbank nach nichttrivialen Prinzipien erfolgt ist. Inhaber sind die Autoren von Datenbankstruktur und Aufbereitungsprinzipien. *Investitionsschutz* hingegen genießt der *Datenbankhersteller*, d. h. diejenige Person, die die Investition getätigt hat, die zur Schaffung der Datenbank notwendig war (vgl. § 87 a Abs. 2 UrhG). Der Datenbankhersteller hat das ausschließliche Recht der Vervielfältigung, Verbreitung und der öffentlichen Wiedergabe der Datenbank.

Aufgrund der besonderen Form des urheberrechtlichen Schutzes existieren für die Rechte des Datenbankherstellers gesonderte Schrankenregelungen. So ist eine „Vervielfältigung eines nach Art oder Umfang wesentlichen Teils einer Datenbank" sowohl zum privaten als auch dem wissenschaftlichen Gebrauch zulässig, sofern sie hierfür erforderlich ist und nicht zu gewerblichen Zwecken erfolgt (§ 87 c UrhG). Ebenfalls zulässig ist „die Benutzung zur Veranschaulichung des Unterrichts" (ebd.), wobei das UrhG im Falle der wissenschaftlichen Nutzung bzw. der Nutzung zu Unterrichtszwecken eine eindeutige Quellenangabe fordert.

[3] Diese Definition gilt auch, wenn sie aus technischer Sicht keine (z. B. relationale) Datenbanken sind.

Anders als das Urheberrecht an Werken im „klassischen“ Sinne (s. o.) erlischt das Recht des Datenbankherstellers bereits 15 Jahre nach Veröffentlichung einer Datenbank bzw. 15 Jahre nach ihrer Herstellung, sofern keine Veröffentlichung stattgefunden hat (vgl § 87 d UrhG).

Angewandt auf linguistische Korpora, die nicht selten unter Aufwand von öffentlichen bzw. Drittmitteln erstellt und aufbereitet werden, bedeutet die hier dargestellte Rechtslage, dass je nach Situation die Forschungseinrichtung bzw. fördernde Institution als Datenbankhersteller betrachtet werden kann, sofern keine anders lautenden vertraglichen Vereinbarungen bestehen.

Ebenfalls durch das Urheberrecht geregelt ist der Schutz von Computerprogrammen, die zur Aufbereitung einer Datenbank oder für den Zugriff auf die Daten erforderlich sind. Ein solches Programm wird grundsätzlich als separater Schutzgegenstand behandelt (§ 4 Abs. 2 UrhG).

Schutz von Primärdaten

Wie bereits erwähnt, sind bestehende Rechte von Urhebern an den Primärdaten einer linguistischen Datensammlung eines der zentralen Hindernisse bei einer Weitergabe bzw. öffentlichen Zugänglichmachung der Daten (z. B. online über das WWW, auf Datenträgern etc.).

Zwar erlauben die in Abschnitt 1.4 angeführten Schrankenreglungen eine Weitergabe an einen eingeschränkten Personenkreis zum Zweck der nicht-kommerziellen wissenschaftlichen Forschung, diese Restriktion lässt sich in der Praxis, z. B. bei der Zugänglichmachung über das Internet, nur schwer realisieren, da zum einen eine Einschränkung auf einen festen Kreis von Personen nicht durchführbar ist, zum anderen die Gewährleistung einer ausschließlich nicht-kommerziellen Nutzung zu Forschungszwecken vom Archivhalter kaum möglich ist.

Die Rechte an den Primärtexten bestehen ungeachtet etwaiger Rechte von Datenbankautoren und -herstellern (s. o.). Bei einer Weitergabe und/oder öffentlichen Zugänglichmachung sind also die Rechte einer Vielzahl von Autoren bzw. Editoren zu berücksichtigen.

Schutz von Sekundärdaten

Linguistische Korpora beinhalten in der Regel neben den Primärdaten auch Annotationen, die häufig unter hohem materiellem und personellem Aufwand zu den Daten hinzugefügt wurden und entscheidend den Wert eines Korpus für wissenschaftliche Zwecke mitbestimmen.

Ungeachtet der Struktur und Form der Speicherung („embedded“ – in derselben Datei wie die Primärdaten oder „standoff“ – in einer separaten Datei) kann die Annotation eines Textes als *Bearbeitung* (Abschnitt 1.1) betrachtet werden. Dies gilt auch für Verfahren wie die Transkription gesprochener Sprache oder die Transliteration von Handschriften. Alle genannten Standardverfahren der Bearbeitung linguistischer Daten können als eigenständiger Schutzgegenstand angesehen werden, wobei als Urheber die Personen gelten, die die Bearbeitung vorgenommen haben. Notwendige Bedingung hierfür ist jedoch, dass die Bearbeitung die in Abschnitt 1.1 beschriebene Schöpfungshöhe erreicht, um urheberrechtlichen Schutz zu genießen.

Zum Zeitpunkt der Entstehung dieses Artikels existiert keine höchstrichterlich gefestigte Rechtsprechung, die eine Einstufung der Schöpfungshöhe gängiger korpuslinguistischer Verfahren vornimmt. Die Frage, ab welchem Grad ein urheberrechtlicher Schutz der Bearbeitung vorliegt, ist jedoch hochrelevant für die weitere Nutzung der Daten, da die Bearbeiter als Urheber im Prinzip Einfluss hierauf nehmen können. Grundsätzlich ist zur Klärung dieser Frage eine

getrennte Einstufung von Annotationsschemata, Transkriptionsstandards u. ä. auf der einen Seite und der auf ihrer Grundlage durchgeführten Bearbeitung auf der anderen Seite erforderlich. Es ist beispielsweise denkbar, dass in konkreten Fällen die Richtlinien und Schemata der Datenaufbereitung die Schöpfungshöhe eines Werks erreichen, die Durchführung hingegen als unwesentliche Bearbeitung keinen Schutz genießen.

Computerprogramme

Einen weiteren relevanten Schutzgegenstand, der bisher im Rahmen dieses Artikels lediglich am Rande erwähnt wurde, stellen Computerprogramme dar, die zur Aufbereitung oder für den Zugriff auf linguistische Daten genutzt werden.

Computerprogramme genießen denselben Schutz wie Sprachwerke (§ 69 a Abs. 3, 4 UrhG). Dieser Schutz bezieht sich auf den Quellcode des Programms sowie das Entwurfsmaterial, nicht jedoch auf die hinter dem Programm stehende Idee oder das Layout. Der Schutz besteht unabhängig von der Qualität des Programms, gilt also für einfache Skripte ebenso wie für umfassende Applikationen.

Da ein Großteil an Software im Rahmen von Arbeitsverhältnissen entwickelt wird, also eine alleinige Urheberschaft der Entwickler einen Vertrieb erschweren würde, schreibt der Gesetzgeber hier eine besondere Handhabung der Verwertungsrechte vor. Für den Fall, dass ein Programm „von einem Arbeitnehmer in Wahrnehmung seiner Aufgaben oder nach de Anweisungen seines Arbeitgebers“ geschaffen wird, ist ausschließlich dieser „zur Ausübung aller vermögensrechtlichen Befugnisse an dem Programm berechtigt, sofern nichts anderes vereinbart ist“ (§ 69 a UrhG).

Falls nicht vertraglich anders geregelt, ist also im Fall von Software, die im Rahmen von Forschung und Lehre durch wissenschaftliches Personal entwickelt wird, die jeweilige Forschungseinrichtung Inhaber der Verwertungsrechte. Diese Tatsache hat unter anderem unmittelbaren Einfluss auf die Möglichkeit, Software, die im Wissenschaftsbetrieb entwickelt wurde, von dem wissenschaftlichen Personal selbst unter öffentliche Lizenzen (GPL etc.) zu stellen. Notwendige Bedingung hierfür ist nämlich, dass der Entwickler alleiniger Inhaber aller Verwertungsrechte ist, was in dieser Situation nicht zutrifft.

2 Datenschutzrecht

Gemessen an der Komplexität der Fragestellungen, die das Urheberrecht mit sich bringt, sind die Regelungen durch das Datenschutzrecht vergleichsweise einfach auf den Umgang mit linguistischen Daten anzuwenden.

Betroffen sind *personenbezogene Daten*. Dabei handelt es sich um jede Information, die eindeutig einer bestimmten natürlichen Person zugeordnet werden kann. Als Korpusdaten fallen hierunter Ton- und Videoaufzeichnungen von Probanden sowie deren Transkriptionen, aber auch Metadaten, die Rückschlüsse auf Personen erlauben. In Deutschland ist der Umgang mit personenbezogenen Daten durch das Bundesdatenschutzgesetz (BDSG) und die entsprechenden Landesdatenschutzgesetze geregelt. Es dient dem Schutz des Einzelnen vor der Beeinträchtigung seines Persönlichkeitsrechts durch den Umgang mit seinen personenbezogenen Daten

(vgl. § 1 Abs. 1 BDSG). Unter dem Begriff *Umgang* werden dabei die Erhebung, Verarbeitung und Nutzung der Daten zusammengefasst.

Ungeachtet der datenschutzrechtlichen Situation haben sich in der sprachwissenschaftlichen Praxis beim Umgang mit personenbezogenen Daten, nicht zuletzt aus ethischen Erwägungen, die Verfahren der *Anonymisierung* und *Pseudonymisierung* durchgesetzt. Als Anonymisierung bezeichnet das BDSG ein derartiges Verändern der Daten, „dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können" (§ 3 Abs. 6 BDSG). Pseudonymisierung hingegen ist das „Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren" (§ 3 Abs. 6 a BDSG). Im Gegensatz zur Anonymisierung werden bei der Pseudonymisierung die Ersetzungen der Identifikationsmerkmale dokumentiert. Diese werden dann in der Regel getrennt aufbewahrt, sind (z. B. unter Verwendung des Pseudonymisierungsschlüssels) jedoch rekonstruierbar. Dieses Verfahren bietet sich unter anderem in Fällen an, in denen personenbezogene Daten (z. B. Informationen über Herkunft und Alter von Probanden) forschungsrelevant sind.

Für die „Verarbeitung und Nutzung personenbezogener Daten durch Forschungseinrichtungen" werden in § 40 BDSG gesonderte Regelungen angeführt. So dürfen personenbezogene Daten, die zum Zweck der wissenschaftlichen Forschung erhoben bzw. gespeichert wurden, auch nur zu diesem Zweck verarbeitet oder genutzt werden (vgl. § 40 Abs. 1 BDSG). Dabei sind die personenbezogenen Daten zu anonymisieren, „sobald dies nach dem Forschungszweck möglich ist" wobei die Merkmale bis zu diesem Zeitpunkt gesondert gespeichert werden müssen (vgl. § 40 Abs. 1 BDSG). Eine Veröffentlichung personenbezogener Daten durch die Forschungseinrichtung ist nur zulässig, wenn die betroffenen Personen ihr zugestimmt haben, oder dies „für die Darstellung von Forschungsergebnissen über Ereignisse der Zeitgeschichte unerlässlich ist" (vgl. § 40 Abs. 1 BDSG).

Angesichts der Tatsache, dass eine vollständige Unkenntlichmachung personenbezogener Daten in vielen Fällen nicht möglich ist, schreibt der Gesetzgeber hinsichtlich des Grades von Anonymisierung/Pseudonymisierung lediglich den oben zitierten „Aufwand an Zeit, Kosten und Arbeitskraft" vor, der zur Rekonstruktion der Daten notwendig wäre. In diesem Fall werden die Daten im Sinne des BDSG nicht mehr als personenbezogen angesehen und es bestehen keine Einschränkungen hinsichtlich der Nutzung und Weitergabe der Daten.

3 Lösungsansätze

Das eingangs erwähnte Projekt untersucht vor dem Hintergrund der hier umrissenen Rechtslage Lösungsansätze, die eine nachhaltige Nutzbarkeit linguistischer Ressourcen zum Zweck der wissenschaftlichen Forschung gewährleisten können. Sie lassen sich in Verfahren zur Modifikation der Inhalte sowie zur Regelung der Zugänglichkeit und Weitergabe der Daten gliedern.

3.1 Modifikation der geschützten Inhalte

Anonymisierung und Pseudonymisierung stellen sichere Verfahren zur datenschutzrechtlichen Absicherung des Umgangs sowie der Weitergabe linguistischer Ressourcen dar, wenn diese per-

sonenbezogene Daten beinhalten. Im Rahmen der Nachhaltigkeitsinitiative wird darüber hinaus die Etablierung einer Treuhandstelle diskutiert, die die Speicherung der personenbezogenen Daten übernimmt und eine gesetzes- und zweckkonforme Weitergabe sicherstellt.

Ein weiterer Ansatz ist, analog zur Unkenntlichmachung personenbezogener Daten durch Anonymisierung/Pseudonymisierung (s. o.), urheberrechtlich geschützte Inhalte zu maskieren. Dies besitzt insbesondere bei der Nutzung von Baumbanken große Attraktivität, da hier die strukturelle Annotation auch ohne die urheberrechtlich geschützten Primärdaten als Datengrundlage für linguistische Untersuchungen dienen kann. Ein Beispiel für die separate Behandlung von Annotation und Primärdaten ist das Lizenzverfahren beim Erwerb der Tübinger Baumbank des Deutschen (TüBa-D/Z, Telljohann et al., 2004), einem syntaktisch annotierten Korpus auf der Grundlage der Zeitung *die tageszeitung*. Nutzer, die Zugriff auf die annotierten Ressourcen erhalten möchten, müssen als Voraussetzung hierfür zunächst eine Lizenz für die Nutzung der Primärdaten erwerben. Eine Nutzung ohne diesen Lizenzerwerb wäre möglich, wenn die urheberrechtlich geschützten Texte im Korpus maskiert würden (Rehm et al., 2007).

3.2 Regelung der Zugänglichkeit und Weitergabe

Ausgehend von den Schrankenregelungen für den Umgang mit urheberrechtlich geschützten Daten zum Zweck der wissenschaftlichen Nutzung (Abschnitt 1.4) hat sich in der Praxis durchgesetzt, den Zugriff auf öffentlich (zumeist online) zugänglich gemachte Korpora einzuschränken. In der Regel erhalten Nutzer nach Registrierung und Anerkennung von Nutzungsvereinbarungen (dieser Vorgang kann sowohl online, als auch auf dem Postweg erfolgen), Zugang auf die Daten. Korpusabfragen liefern nur kleine Ausschnitte der Primärdaten (also im juristischen Sinn „kleine Teile eines Werkes"), die urheberrechtlich geschützten vollständigen Texte bleiben dem Nutzer verborgen. Diesem Verfahren gehen meistens umfassende vertragliche Vereinbarungen zwischen den publizierenden Forschungseinrichtungen und Urhebern bzw. Verlagen voran (vgl. Jüttner, 2000, S. 12). Im Rahmen der Nachhaltigkeitsinitiative werden zu diesem Zweck Vorlagen für diese Vereinbarungen konzipiert und zur Verfügung gestellt.

Eine vor allem unter dem Aspekt der Datennachhaltigkeit attraktive Möglichkeit der Regelung der Weitergabe ist die bereits erwähnte Etablierung einer Treuhandstelle. Eine solche Institution kann die Speicherung ebenso wie die Regelung der Nutzung und Weitergabe personenbezogener Daten übernehmen (vgl. Schach et al., 1995, S. 63). Denkbar ist auch eine Ausweitung dieses Vorgangs auf komplette linguistische Datensammlungen. Anfragen bezüglich der Nutzung von Daten könnten an die Treuhandstelle gerichtet werden, die ihrerseits auf der Grundlage von Vereinbarungen mit Urhebern sowie der jeweils gültigen Rechtslage eine Weitergabe der Daten regeln bzw. einschränken würde.

4 Ausblick

Der vorliegende Artikel gewährt lediglich Einblicke in die zentralen juristischen Aspekte, die sich vor dem Hintergrund der nachhaltigen Nutzung und Weitergabe linguistischer Daten ergeben. Angesichts der Heterogenität und Vielschichtigkeit der betroffenen Ressourcen und den damit verbundenen komplexen Rechtsfragen ist in der Regel jedoch eine fallspezifische Klärung notwendig, die nicht von juristischen Laien geleistet werden kann. Aus diesem Grund kann

langfristig nur die Etablierung von Institutionen, die sowohl als Treuhänder (vgl. Abschnitt 3.2) als auch als Kompetenzzentrum für Rechtsfragen fungieren, eine dauerhafte Verfügbarkeit der Ressourcen zum Zweck der linguistischen Forschung gewährleisten.

Die Angaben in diesem Beitrag beziehen sich auf folgende Gesetzestexte

- Bundesdatenschutzgesetz (BDSG) in der Fassung der Bekanntmachung vom 14.01.2003 (BGBl. I S. 66), zuletzt geändert durch Artikel 1 des Gesetzes vom 22.08.2006 (BGBl. I S. 1970).
- Urheberrechtsgesetz (UrhG) vom 09.09.1965 (BGBl. I S. 1273), zuletzt geändert durch das Gesetz vom 10.11.2006 (BGB I S. 2587).
- Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11.03.1996 über den rechtlichen Schutz von Datenbanken.

Literaturverzeichnis

Beger, G. (2003): "Internet und Recht". In: *Geschichte und Neue Medien in Forschung, Archiven, Bibliotheken und Museen. Historisches Forum*, herausgegeben von Burckhardt, D.; Hohls, R. und Ziegeldorf, V. Band 7/2005, S. 149–154.

Benecke, M. (2004): "Was ist ›wesentlich‹ beim Schutz von Datenbanken?" *Computer und Recht* 8: S. 608–613.

Jüttner, I. (2000): "Mannheimer Korpus und Urheberrecht. Die Einbeziehung zeitgenössischer digitalisierter Texte in die computergespeicherten Korpora des IDS und ihre juristischen Grundlagen". *Sprachreport* 3: S. 11–13.

Aktionsbündnis ›Urheberrecht für Bildung und Wissenschaft‹ (2004): "Göttinger Erklärung zum Urheberrecht für Bildung und Wissenschaft". `http://www.urheberrechtsbuendnis.de`.

Patzelt, J. (2003): "Unter juristischem Blickwinkel: Textkorpora und Urheberrecht". In: *Korpuslinguistik deutsch: Synchron – diachron – kontrastiv*, herausgegeben von Schwitalla, J. und Wegstein, W.

Rehbinder, M. (2005): *Urheberrecht. Ein Studienbuch*. München: Beck, 14. Auflage.

Rehm, G.; Witt, A.; Zinsmeister, H. und Dellert, J. (2007): "Corpus Masking: Legally Bypassing Licensing Restrictions for the Free Distribution of Text Collections". Proc. of Digital Humanities 2007.

Schach, E.; Kilian, W.; Podlech, A. und Schlink, B. (1995): *Daten für die Forschung im Gesundheitswesen, Statistische und rechtliche Klärung ihrer Weitergabe*. Darmstadt: Toeche-Mittler.

Schmidt, T.; Chiarcos, C.; Lehmberg, T.; Rehm, G.; Witt, A. und Hinrichs, E. (2006): "Avoiding Data Graveyards: From Heterogeneous Data Collected in Multiple Research Projects to Sustainable Linguistic Resources". In: *Proc. of the E-MELD 2006 Workshop on Digital Language Documentation*.

Telljohann, H.; Hinrichs, E. und Kübler, S. (2004): "The TüBa-D/Z Treebank – Annotating German with a Context-Free Backbone". In: *Proc. of LREC 2004*. Lissabon, Portugal.

Thum, K. (2005): "Urheberrechtliche Zulässigkeit von digitalen Online-Bildarchiven zu Lehr- und Forschungszwecken". *Kommunikation und Recht* 11: S. 490–498.

Wandtke, A.-A. und Bullinger, W. (Herausgeber) (2006): *Praxiskommentar zum Urheberecht*. München: Beck, 2. Auflage.