



D-SPIN

**D-SPIN Report R7.1:
Legal Aspects in the Provision of
Language Resources:
The German Context**

June 2009

D-SPIN, BMBF-FKZ: 01UG0801C

Deliverable: R7.1:
Legal Aspects in the Provision of Language Resources:
The German Context

© all rights reserved by Institute for the German Language on behalf of D-SPIN

Editors: Marc Kupietz and Christina Bankhardt

<i>Overview</i>	4
<i>Bestandsaufnahme juristisch relevanter Aspekte bei der Verfügbarmachung von Sprachressourcen</i>	6
1. Textressourcen	6
2. Gesprächsressourcen	9
3. Haftungsproblematik.....	12
<i>Fazit</i>	12
<i>Literaturverzeichnis</i>	13

Overview

Legal and ethical issues are of key importance to the viability of the CLARIN and D-SPIN infrastructure. D-SPIN and CLARIN are committed to the idea of open access. However, we cannot ignore the fact that the use of the majority of digital research resources in the humanities is subject to a variety of legal and ethical restrictions and that this is unlikely to change in the foreseeable future.

For instance, linguistics is faced with the specific problem that not only its secondary research data, i.e. published scientific research, is in general protected by intellectual property rights but also its primary research data, namely recorded observations of language use. This has several aspects and implications. First of all, unlike the rights on typical secondary research data and the primary research data in natural sciences the rights on research data in linguistics typically belong to persons and corporations outside the public sector and, in particular, outside the scientific community. As a consequence, open-access models from other scientific communities are not easily transferable to the CLARIN/D-SPIN. That also means that typically e.g. fair use doctrines or the limitations in the German copyright law – regardless of whether they would or will suffice for the typical use of the data – do not come into play because the researcher or resource provider does not get hold of the data unless they have signed some kind of licence agreement.

In short, licence agreements are typically needed to make language resources available and (re-)usable and it is one of the major tasks of this work package and its other national and CLARIN counterparts to provide model licence agreements in order to make CLARIN and D-SPIN readily receptive to new resources and in order to avoid reinventing the wheel, which would be particularly expensive and full of pitfalls in this domain. Like other infrastructure tasks and re-usability tasks in general, the development of model agreements is challenging because it has to deal with conflicting interests, trade-offs and compatibility issues. The typical trade-off in this context is, that the more extensive the rights granted in a formal agreement are, the more expensive the licence, and the less likely is its signing. And, while some research questions, e.g., require personal information on a speaker or whole texts others do not. While such different needs have to be taken seriously also a certain harmonisation and canonicalization will be needed for the technical implementation of the licence management and in order to be able to control the compatibility in a network of contracts between the different actors: IPR holders, copyright holders, resource providers, data centres, end users, etc. Thus, an analogy with the standardisation of, e.g., metadata and maybe even an analogy with solutions from that domain, like concept and relation registries, is not far-fetched.

Although the handling of legal and ethical restrictions with respect to the use of language data is a complicated matter, the situation sketched in this survey must not be misunderstood as desperate. Good news, e.g., is that the way linguists and researchers from adjacent disciplines use language data usually does not interfere substantially with the economical or other interests of IPR or copyright holders: linguists use language resources in a very specific way, they are not interested in the information they contain, they do not read, see or hear them for their entertainment, and typically do not make profit of them. For that reason the use of licence agreements is not only the single commonly possible way for the LRT community to make research data available, but also a feasible and affordable one. Since the LRT and linguistics communities, as mentioned above, do vitally depend on the donation of resources by copyright holders, the D-SPIN approach is both ethically and strategically the preferable way to pushing or stretching barriers in legally unclear areas. The same holds of course for contractually unclear areas. Not only in the long run, confidence and trust are very valuable goods here. To squander them for some small local gains, could, on the level of a large infrastructure initiative, turn fatal for the whole community that depends on empirical data.

A second piece of good news is that the biggest problem with licence restrictions, viz. that their fulfilment by the end users could not be guaranteed or enforced, can indeed be tackled within the protected and secure environment of the CLARIN/D-SPIN technical infrastructure. Such a controlled

technical infrastructure restricted to the scientific community would also be a prerequisite to benefit from extended fair use doctrines as otherwise e.g. novels could end up being downloadable for everyone or patient interviews could end up in TV shows, regardless of whether they were intended for academic use only.

To summarise, in the case of linguistics and adjacent disciplines, CLARIN and D-SPIN have an even bigger potential to *enable humanities* than research infrastructures in other disciplines, because they can additionally provide the contractual models and technical means to deal with typical legal restrictions in such a way that makes language resources as (re-)usable as legally and ethically possible and thereby enable empirical research in an additional, fundamental sense.

At the national D-SPIN level, which is handled in this first report, the first step to this end is to provide a survey of legally relevant aspects of making available language resources in the German context. In subsequent steps, the outcomes of comparable national surveys will be gathered and integrated at the EU-CLARIN level to serve as input for the development of canonised licence and best practice models which in turn will be localised on national levels.

Because the following survey makes strong use of references to German legal texts and the definition and explanation of even the basic legal technical terms would exceed the scope of this report, it will be held completely in German.

Bestandsaufnahme juristisch relevanter Aspekte bei der Verfügbarmachung von Sprachressourcen

Eine Herausforderung für empirisch arbeitende Linguisten ist die Tatsache, dass ihre Forschungsprimärdaten in der Regel nicht frei von Rechten Dritter sind. Sie arbeiten entweder primär mit urheberrechtlich geschützten Texten oder sie müssen bei Gesprächsdaten datenschutzrechtliche Aspekte beachten. Darüber hinaus können bei Gesprächskorpora wie auch bei Textressourcen persönlichkeitsrechtliche Aspekte eine Rolle spielen, vor allen Dingen dann, wenn deren Inhalte die Privat- oder gar Intimsphäre anderer berühren. Ebenso ist es möglich, dass eine Vielzahl von Individuen „geistiges Eigentum“ im urheberrechtlichen Sinn an Gesprächsressourcen erwerben kann.

Um der Forschergemeinschaft im Rahmen von D-SPIN / CLARIN einen möglichst freien Zugang zu den Sprachressourcen gewähren zu können, muss, unabhängig von anderen europäischen Regelungen, im deutschen Kontext eine Reihe von Aspekten beachtet werden, denen verschiedene Grundrechtskonflikte zugrunde liegen.

1. Textressourcen

Aus urheberrechtlicher Sicht kollidiert auf Grundrechtsebene die Wissenschaftsfreiheit aus Art. 5 III GG mit der Eigentumsgarantie aus Art. 14 GG. Primär betrifft dies Korpora, die Texte enthalten.

Art. 5 GG: Wissenschaftsfreiheit

III. Kunst und Wissenschaft, Forschung und Lehre sind frei. Die Freiheit der Lehre entbindet nicht von der Treue zur Verfassung.

Art 14 GG: Eigentumsgarantie

I. Das Eigentum und das Erbrecht werden gewährleistet. Inhalt und Schranken werden durch die Gesetze bestimmt.

II. Eigentum verpflichtet. Sein Gebrauch soll zugleich dem Wohle der Allgemeinheit dienen.

...

Die Wissenschaftsfreiheit gewährt jedem, der in Wissenschaft und Forschung tätig ist, ein Recht auf Abwehr jeder staatlichen Einwirkung auf den „Prozess der Gewinnung und Vermittlung wissenschaftlicher Erkenntnisse“¹ und das vorbehaltlos. Das bedeutet, die Freiheit der Wissenschaft kann nur durch kollidierendes Verfassungsrecht eingeschränkt werden, wie z.B. die Eigentumsgarantie. Das Eigentum eines Urhebers an seinem Text wird durch das Urheberrecht geschützt. Dessen Schutz umfasst sowohl die ideellen als auch die materiellen Interessen eines Urhebers an seinem Werk². Gemäß § 2 UrhG sind folgende Werke geschützt:

§ 2 UrhG: Geschützte Werke

I. Zu den geschützten Werken der Literatur, Wissenschaft und Kunst gehören insbesondere:

1. Sprachwerke, wie Schriftwerke, Reden und Computerprogramme;

¹ BVerfGE 35, 79/113.

² vgl. Reh binder 2008, Rn. 5f.

2. *Werke der Musik;*
3. *pantomimische Werke einschließlich der Werke der Tanzkunst;*
4. ...

II. Werke im Sinne dieses Gesetzes sind nur persönliche geistige Schöpfungen.

An das zu schützende Werk werden keine qualitativen Anforderungen gestellt, es muss allerdings eine gewisse Schöpfungshöhe erreichen, um als Werk im Sinne des Urheberrechts qualifiziert werden zu können und urheberrechtlichen Schutz genießen zu können. Damit die Schöpfungshöhe erreicht wird, muss ein Text einen gewissen Grad an Individualität aufweisen³. Bei den Texten in Sprachressourcen, die der Forschung zur Verfügung stehen, dürfte dies in aller Regel der Fall sein. Urheberrechte an einem Werk entstehen in Deutschland ohne formelles Verfahren rein mit der Schaffung des Werks⁴ und sind nicht übertragbar. Der Urheber hat das Recht, selbst zu entscheiden, ob und wie das Werk zu veröffentlichen ist (§ 12 UrhG, Urheberpersönlichkeitsrecht). Geschützt sind auch die wirtschaftlichen Verwertungsinteressen des Urhebers (§ 15 UrhG). Verwertet werden kann ein Werk durch Vervielfältigung (§ 16 UrhG) oder Verbreitung in der Öffentlichkeit (§17 UrhG). Wer ein urheberrechtlich geschütztes Werk nutzen will, kann dies auf zwei Arten tun. Entweder kann er sich auf eine der Schranken im Urheberrecht beziehen oder er erwirbt die Nutzungsrechte durch einen Lizenzvertrag.

Schranken sind Regelungen im Urhebergesetz, die eine Nutzung auch ohne Lizenzvertrag und damit ohne Erlaubnis des Urhebers gestatten. Sie schaffen einen Ausgleich zwischen den Interessen des Urhebers, dem prinzipiell das ausschließliche Nutzungsrecht seines Werkes eingeräumt ist, und gegenläufigen Interessen z.B. Privater oder Wissenschaftler.

Eine der bekanntesten Schranken ist das Zitatrecht. Zitate mit Quellenangaben dürfen gemäß § 51 UrhG vervielfältigt, verbreitet und öffentlich wiedergegeben werden, ohne Zustimmung des Autors. Ebenso bekannt ist die Schranke, die Vervielfältigungen zum privaten Gebrauch erlaubt (§ 53 UrhG), solange kein Kopierschutz umgangen oder eine rechtswidrige Vorlage benutzt wird. Gemäß § 52a UrhG werden die Rechte des Urhebers auch zu Gunsten der Wissenschaft (Forschung und Lehre) eingeschränkt.

§ 52a UrhG: Öffentliche Zugänglichmachung für Unterricht und Forschung

I. Zulässig ist,

1. veröffentlichte kleine Teile eines Werkes, Werke geringen Umfangs sowie einzelne Beiträge aus Zeitungen oder Zeitschriften zur Veranschaulichung im Unterricht an Schulen, Hochschulen, nichtgewerblichen Einrichtungen der Aus- und Weiterbildung sowie an Einrichtungen der Berufsbildung ausschließlich für den bestimmt abgegrenzten Kreis von Unterrichtsteilnehmern oder

2. veröffentlichte Teile eines Werkes, Werke geringen Umfangs sowie einzelne Beiträge aus Zeitungen oder Zeitschriften ausschließlich für einen bestimmt abgegrenzten Kreis von Personen für deren eigene wissenschaftliche Forschung öffentlich zugänglich zu machen, soweit dies zu dem jeweiligen Zweck geboten und zur Verfolgung nicht kommerzieller Zwecke gerechtfertigt ist.

Während Ziffer eins die restriktivere Schrankenregelung für die Lehre enthält, beinhaltet Ziffer zwei die etwas weitere Schrankenregelung für die Forschung. Die Formulierung ist jedoch sehr vage gehalten und scheint auf den ersten Blick zwar einiges zu erlauben. Doch muss man zunächst wissen, was „Teile eines Werkes“ sind. Dazu gibt es noch keine gerichtliche Entscheidung, und in der rechtswissenschaftlichen Literatur gehen die Meinungen dazu auseinander. Die Schätzung, ein Teil

³ vgl. Reh binder 2008, Rn. 55ff.

⁴ vgl. Reh binder 2008, Rn. 97.

eines Werkes liege „deutlich unter 50%“⁵ des gesamten Werkes, wird kontrovers diskutiert⁶; überwiegend wird davon ausgegangen, dass die Teile das gesamte Werk nicht ersetzen dürfen. Genauso kritisch wird erörtert, was unter Werken „geringen Umfangs“ zu verstehen ist⁷. 2. Frage: Was ist ein „bestimmt abgegrenzter Personenkreis“? Wille des Gesetzgebers war es, mit dieser Regelung kleine Forschergruppen zu privilegieren, die gemeinsam in einem Projekt arbeiten⁸. Die gesamte LRT-Community ist daher wohl zu groß, um von dieser Schranke profitieren zu können. Darüber hinaus ist diese Schrankenregelung zeitlich beschränkt und niemand kann sagen, ob diese Schranke auch nach 2012 noch besteht.

§ 137k UrhG: Übergangsregelung zur öffentlichen Zugänglichmachung für Unterricht und Forschung

§ 52a ist mit Ablauf des 31. Dezember 2012 nicht mehr anzuwenden.

Eine derartige zeitliche Beschränkung ist für die Interessen von Infrastrukturprojekten sehr kritisch. Im Hinblick auf eine zeitliche Beschränkung ist für die Forschung außerdem die lange Dauer des Urheberrechtsschutzes relevant: Texte sind bis 70 Jahre nach dem Tod des Urhebers geschützt (§ 64 UrhG), nach Ablauf dieser Schutzfrist werden sie gemeinfrei und können von jedem frei verwertet werden (die Rechte an Tonträgern erlöschen 50 Jahre nach Erscheinen des Tonträgers (§ 85 II UrhG), Datenbanken sind 15 Jahre nach der Veröffentlichung der Datenbank geschützt (§ 87d UrhG)).

Da die Schranken nicht ausreichend sind, müssen, um ausreichende Nutzungsrechte sicher zu stellen, diese durch Lizenzverträge eingeräumt werden. Nutzungsrechte können vom Urheber bzw. Verwertungsrechteinhaber ausschließlich oder einfach eingeräumt werden. Bei ausschließlichen Nutzungsrechten darf niemand außer dem Rechteinhaber selbst einen Text verwerten, bei einfachen Nutzungsrechten darf der Text auch von anderen verwertet werden⁹. Es ist allerdings aus Ressourcenhaltersicht nicht immer klar ersichtlich, wer die jeweiligen Rechte wirklich inne hat. Denn Texte werden teilweise von mehreren Autoren verfasst, Verleger haben meist die ausschließlichen Nutzungsrechte dieser Texte inne. Problematisch ist, dass es keinen gutgläubigen Erwerb von Nutzungsrechten gibt¹⁰. Im Zweifel haftet demnach der Ressourcenhalter, der die Texte weitergibt.

Weitere rechtliche Ebenen entstehen zum einen mit den durch das Urheberrecht gewährleisteten Leistungsschutzrechten sowie mit der Annotation von Texten, die im Rahmen der linguistischen Aufbereitung hinzugefügt wird. Es ist bisher ungeklärt, ob auch an Annotationen neue Urheberrechte entstehen können. Problematisch ist dies einerseits bei Annotationen, die von kommerziellen Tools (halb)automatisch hinzugefügt werden und andererseits bei manuell hinzugefügten Annotationen. Insbesondere bei letzteren sollte man davon ausgehen, dass diese Annotationen die erforderliche Schöpfungshöhe erreichen. Die Rechteinhaber an den Annotationen sind damit aber wiederum andere wie die der ursprünglichen Texte. Bei der Feststellung von möglichen Urheberrechten kommerzieller Werkzeuge sind v.a. die Lizenzverträge für die Nutzung der Werkzeuge zu berücksichtigen.

Außerdem können bei der Zurverfügungstellung von Zeitungsartikeln zu wissenschaftlichen Zwecken auch Persönlichkeitsrechte Dritter, die in den Artikeln erwähnt werden, betroffen sein (vgl. 2.), insbesondere dann, wenn Eigennamen erwähnt werden.

⁵ Lüft 2009, § 52a Rn. 12.

⁶ vgl. u.a. Dreier/Schulze 2008, § 52a Rn. 9; Loewenheim 2006, § 52a Rn. 13.

⁷ vgl. Lüft 2009, § 52a Rn. 7; Dustmann 2008, § 52a Rn. 8; Loewenheim 2006, § 52a Rn. 13.

⁸ vgl. BT-Drucksache 15/837, S. 34.

⁹ vgl. Schack 2007, Rn. 539.

¹⁰ vgl. Czychowski 2003, § 13 Rn. 241; Hertin 2008, Rn. 348.

2. Gesprächsressourcen

Auf dieser Ebene kollidieren auf der einen Seite wiederum die schrankenlos gewährleistete Wissenschaftsfreiheit und das allgemeine Persönlichkeitsrecht auf der anderen Seite.

Das allgemeine Persönlichkeitsrecht ist ein aus der Zusammenschau der freien Entfaltung der Persönlichkeit aus Art. 2 I i.V.m. Art. 1 I GG dem Schutz der Menschenwürde entwickeltes Grundrecht¹¹, das auch post mortem gilt.

Art. 2 GG: Freie Entfaltung der Persönlichkeit

I. Jeder hat das Recht auf die freie Entfaltung seiner Persönlichkeit, soweit er nicht die Rechte anderer verletzt und nicht gegen die verfassungsmäßige Ordnung oder das Sittengesetz verstößt.

Art. 1 GG: Menschenwürde

I. Die Würde des Menschen ist unantastbar. Sie zu achten und zu schützen ist Verpflichtung aller staatlichen Gewalt.

Die für die Verfügbarmachung von Sprachressourcen relevante Ausprägung des allgemeinen Persönlichkeitsrechts ist insbesondere das Recht auf informationelle Selbstbestimmung. Es soll jedem Einzelnen ermöglichen, selbst zu bestimmen, welche Daten er von sich selbst preisgeben möchte¹². (Daneben sind davon auch das Recht am eigenen Bild, gesprochenen Wort, geschriebenen Wort und der Schutz der persönlichen Ehre erfasst.)

Die Datenschutzgesetze versuchen einen Ausgleich zwischen diesen beiden kollidierenden Grundrechten zu schaffen.

Das Bundesdatenschutzgesetz schützt gemäß § 1 I BDSG den Einzelnen davor, dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird.

§ 1 BDSG: Zweck und Anwendungsbereich des Gesetzes

I. Zweck dieses Gesetzes ist es, den Einzelnen davor zu schützen, dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird.

Personenbezogene Daten in diesem Sinne sind gemäß § 3 I BDSG Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren Person, wie der Name z.B.

§ 3 Weitere Begriffsbestimmungen

I. Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener).

Desweiteren sind Wissenschaftler gemäß § 40 II BDSG gehalten, ihre Forschung möglichst auf der Grundlage anonymisierter Daten durchzuführen.

§ 40 BDSG: Verarbeitung und Nutzung personenbezogener Daten durch Forschungseinrichtungen

II. Die personenbezogenen Daten sind zu anonymisieren, sobald dies nach dem Forschungszweck möglich ist. Bis dahin sind die Merkmale gesondert zu speichern, mit denen Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren Person zugeordnet werden können. Sie dürfen mit den Einzelangaben nur zusammengeführt werden, soweit der Forschungszweck dies erfordert.

¹¹ vgl. Pieroth/Schlink 2008, Rn. 373; Ladeur 2008, § 7 Rn. 2.

¹² vgl. Jarass/Pieroth 2007, Art. 2 Rn. 44; BVerfGE 65, 1/43.

Problematisch aus linguistischer Sicht ist, dass entgegen der oft vertretenen Meinung, anonymisierte Daten seien für die Forschung ausreichend¹³, Forschungsprojekte der Linguistik, durchaus auf personenbezogene oder gar sensible Daten angewiesen sein können. Eine Untersuchung von Gesprächen zwischen einem Arzt und seinem Patienten bspw. ist ohne Kenntnis sensibler Details zum Gesundheitszustand nicht ohne Weiteres möglich.

Ist eine Anonymisierung nicht möglich, dürfen Daten auch in pseudonymisierter Form verarbeitet werden. Nur wenn auch dies nicht möglich ist, dürfen personenbezogene Daten genutzt werden. Es bedarf aber einer sorgfältigen datenschutzrechtlichen Prüfung:

Es stellt sich daher die Frage: Ist die Anonymisierung bzw. Pseudonymisierung von Gesprächsaufzeichnungen möglich und sind die Daten für eine Nachnutzung dann überhaupt noch brauchbar?

Die Anonymisierung ist die Veränderung der personenbezogenen Daten in derart, dass diese Daten nicht mehr einer bestimmten Person zugeordnet werden können. Da eine absolute Anonymisierung nur selten möglich ist, ist eine sog. faktische Anonymisierung ausreichend¹⁴. Bei einer faktischen Anonymisierung kann ein Betroffener nur mit einem unverhältnismäßig großen Aufwand reidentifiziert werden. Bei der Pseudonymisierung wird der Name oder ein anderes Identifikationsmerkmal durch ein Pseudonym ersetzt, um die Identifizierung des Interviewten durch Unberechtigte auszuschließen oder wesentlich zu erschweren¹⁵.

In Bezug auf Primärdaten bedeutet dies nun:

Bei den Audioaufnahmen müssten sämtliche personenbezogene Daten aus den Gesprächen gelöscht oder verfremdet werden. Die betreffenden Stellen in der Aufnahme, wenn z.B. der Name einer natürlichen Person innerhalb eines Gespräches fällt, könnten komplett gelöscht oder durch Messtöne wie Pfeiftöne oder Rauschen ersetzt werden. Fraglich ist, ob die Daten bereits als faktisch anonymisiert gelten. Problematisch ist nämlich, dass die Stimme, durch die eine Person durchaus identifiziert werden kann, erhalten bleibt. Es könnte als weitere Maßnahme die Stimme z.B. durch eine Tonhöhenveränderung verfremdet werden, doch kann diese mit entsprechender Software nachträglich wieder rückgängig gemacht werden. Damit wäre der Schutz der Interviewten nicht mehr gewährleistet, die Aufnahmen könnten de-anonymisiert werden. Allerdings dürfte dies i.d.R. einem unverhältnismäßig großen Aufwand entsprechen.

Weitere Nachteile sind: Qualitätseinbußen müssten in Kauf genommen werden und erschweren zusätzlich zu Pfeiftönen oder Rauschen eine linguistische Auswertung. Klanginformationen und inhaltliche Informationen, die für die Forschung von Bedeutung sein könnten, können verloren gehen. Zudem erfordert diese Art der Anonymisierung/Pseudonymisierung einen hohen technischen Aufwand. Eine Nachnutzung würde größtenteils wohl verhindert.

Bei Videodaten stellt sich darüber hinaus das Problem, dass neben dem Ton auch das Bildmaterial verfremdet werden müsste. Das Gesicht könnte mit einem schwarzen Balken maskiert werden oder mit anderen Effekten unkenntlich gemacht werden. Die Videodaten ermöglichen jedoch nur in ihrem unverfremdeten Zustand z.B. multimodale Untersuchungen, da es dort auf die Auswertung des Gesichtsausdrucks, der Mimik, Körperhaltung, Bewegung und Interaktion im Raum ankommt. Daher wären die Videoaufnahmen nach solch einer Anonymisierung/Pseudonymisierung für die Forschung schlicht unbrauchbar.

In Bezug auf Gesprächstranskriptionen ist eine Anonymisierung bzw. Pseudonymisierung grundsätzlich möglich, personenbezogene Daten wie der Namen usw. können durch Decknamen

¹³ vgl. Gola/Schomerus 2007, § 40 Rn. 14; Simitis § 40 Rn. 67.

¹⁴ vgl. Tinnefeld 2005, S. 287.

¹⁵ vgl. Metschke/Wellbrock, S. 23.

ersetzt werden. Bei narrativen Interviews mit erheblich biographischem Anteil ist es wohl aber nicht ausreichend, wenn lediglich der Name und sonstige persönliche Bezüge aus einem Interview gestrichen werden, um von einer Anonymisierung sprechen zu können. Man müsste so sehr in den Text eingreifen und ihn bspw. durch Legung falscher Fährten verfremden, dass es ungeeignet zu sein scheint, dass diese Transkription für eine Nachnutzung überhaupt noch brauchbar sein könnte, da Kontexte entstehen können, die nicht mehr nachvollziehbar oder gar sinnverzerrend sind. Problematisch bei alignierten Transkripten ist die Tatsache, dass sie mit der Audioaufnahme und damit mit der Stimme der Sprecher verbunden sind.

Zur Rechteinräumung empfiehlt es sich bei datenschutzrechtlich geschützten Daten, immer mit Einwilligungserklärungen zu arbeiten, v.a. bei der Veröffentlichung von Gesprächsaufzeichnungen. Denn gemäß § 40 BDSG:

§ 40 BDSG: Verarbeitung und Nutzung personenbezogener Daten durch Forschungseinrichtungen

III. Die wissenschaftliche Forschung betreibenden Stellen dürfen personenbezogene Daten nur veröffentlichen, wenn

- 1. der Betroffene eingewilligt hat oder*
- 2. ...*

Es gibt verschiedene formale und inhaltliche Anforderungen an eine solche Einwilligungserklärung. Zu den formalen Anforderungen gehört u.a. die Schriftform (§ 4a BDSG). Die schriftliche Fixierung des Einverständnisses soll zum einen den Betroffenen vor einer übereilten oder unbedachten Einwilligung schützen. Sie dient zum anderen aber auch dem Wissenschaftler zu Beweis Zwecken¹⁶. Aufgrund der natürlichen Skepsis gegenüber Kleingedrucktem ist an der Schriftform problematisch, dass sie manche potentielle Interviewte aber auch abschrecken kann, am Projekt überhaupt teilzunehmen. Je umfangreichere Rechte in einer solchen Erklärung eingeräumt werden sollen, umso länger wird diese und umso unwahrscheinlicher wird es, dass sie unterschrieben wird. Die Einwilligung ist darüber hinaus grundsätzlich vor der Aufnahme einzuholen¹⁷. Auch das kann zu Problemen führen, v.a. da die Interviewten im Vorhinein über den Forschungszweck (s. inhaltliche Anforderungen) aufgeklärt werden müssen.

Inhaltliche Anforderungen sind u.a. (*informed consent*)¹⁸: Träger und Leiter des Forschungsvorhabens müssen genannt werden. Informationen über die Erhebung, den Forschungszweck und die weitere Verarbeitung der Daten müssen an den Interviewten weitergegeben werden. Das Wissen über den Forschungszweck kann gerade Gespräche, die möglichst authentisch sein sollen, aber beeinflussen. Werden die Interviewten vorab aber nicht über den Forschungszweck aufgeklärt oder sogar mit Hilfe einer *cover story* über den wahren Zweck der Untersuchung getäuscht, kann die einzelne Person nicht entscheiden, welche Informationen sie über ihr Verhalten von sich selbst preisgeben möchte und genau dies stünde im Widerspruch zum Recht auf informationelle Selbstbestimmung. Problem bei Sprachressourcen in Infrastrukturprojekten ist außerdem die Vielzahl von Forschungszwecken, für die Gesprächsdaten genutzt werden sollen. Der Zeitpunkt der Löschung personenbezogener Daten sollte möglichst konkret benannt werden¹⁹. Diese Metadaten wären nicht mehr nachnutzbar. In der Einverständniserklärung muss weiterhin auf die Freiwilligkeit der Teilnahme hingewiesen werden und die Möglichkeit des Widerrufs genannt werden²⁰. Selbst

¹⁶ vgl. Holznagel/Sonntag 2003, S. 689.

¹⁷ vgl. Gola/Schomerus 2007, § 4a Rn. 2, 15; Simitis 2006, § 4a Rn. 27ff.

¹⁸ vgl. Holznagel/Sonntag 2003, S. 695f; Metschke/Wellbrock 2002, S. 25ff; Gola/Schomerus 2007, § 4a Rn. 10ff.

¹⁹ vgl. Metschke/Wellbrock 2002, S. 27.

²⁰ vgl. Gola/Schomerus 2007, § 4a Rn. 17ff.

wenn man allerdings aufgrund der Einverständniserklärung dazu berechtigt wäre, im Falle eines Widerrufs, die Daten in anonymisierter Form weiterhin bereit zu stellen, ist die Frage, ob dies auch ethisch noch vertretbar ist. Zudem müssen die Daten und Metadaten so aufbereitet sein, dass es möglich ist, die Daten derjenigen aus dem Datenkorpus zu entfernen, die von ihrem Widerrufsrecht Gebrauch machen. Ähnliches gilt bei der Verletzung von Persönlichkeitsrechten in Zeitungsartikel, die einzeln aus einem Korpus evtl. entfernbar sein müssen.

Weitere Probleme ergeben sich bei nicht einwilligungsfähigen Personen wie Kindern oder Aphasikern²¹. Werden Gesprächsaufzeichnungen bspw. in Schulen gemacht, muss oft die Zustimmung von Schulleiter, Schulverwaltung, Lehrer, Ministerien, Schüler und Eltern eingeholt werden. Darüber hinaus wird von Schulbehörden auch die Zustimmung der Datenschutzbehörden gefordert.

Aufgrund der föderalen Organisation des deutschen Staates gibt es neben dem Bundesdatenschutzgesetz auch die Landesdatenschutzgesetze, die abweichende Regelungen enthalten können und je nach Kooperation evtl. auch in einer Zusammenschau mit dem Bundesdatenschutzgesetz gesehen werden müssen.

3. Haftungsproblematik

Auch wenn die Wissenschaft primär keine kommerziellen Absichten verfolgt, die den Interessen von Urhebern oder Rechteinhabern zuwiderlaufen könnten, sind bestimmte Haftungsrisiken zu bedenken. Denn die Folgen können schwerwiegend sein nicht nur in finanzieller Hinsicht, auch der Ruf einer Institution kann schweren Schaden nehmen.

Bezüglich der Textressourcen stellt sich z.B. die Frage, wer haftet, wenn ein Ressourcenhalter nicht wirklich die Rechte inne hatte, die er glaubte inne zu haben und z.B. Texte zur Verfügung stellt, die er nicht hätte bereitstellen dürfen? Heikel hinsichtlich der Gesprächsressourcen ist die Tatsache, dass bei der Verfügbarmachung solcher Ressourcen, Gesprächsdaten auch für brisante Forschungsfragen verwendet werden können, für die evtl. keine Einwilligung gegeben wurde.

Es ist außerdem problematisch, wenn alle erforderlichen Nutzungsrechte und Einwilligungen eingeholt wurden, die Daten aber in die falschen Hände geraten und nicht zu wissenschaftlichen Zwecken gebraucht werden.

Schwierig bei bereits existierenden Ressourcen ist darüber hinaus, dass oft die Nutzungsrechte unklar sind. Es gibt bspw. ganze Korpora von Gesprächsaufzeichnungen, bei denen keine oder nur unzureichende Einverständniserklärungen vorliegen. Es ist sehr aufwendig mitunter auch unmöglich, die erforderlichen Nutzungsrechte im Nachhinein noch einzuholen. Auch bereits bestehende Lizenzverträge mit Textgebern decken nicht ohne weiteres die Aufnahme der Korpora in eine europäische Infrastruktur ab.

Bei Verletzungen des allgemeinen Persönlichkeitsrechts oder einer Urheberrechtsverletzung kommen insbesondere Unterlassungs- und Beseitigungs- sowie Schadensersatzansprüche in Betracht.

Fazit

Bezüglich der Textressourcen müssen für eine ausreichende Rechtesicherung Lizenzverträge abgeschlossen werden, da die Schrankenregelung (§ 52a UrhG) für die öffentliche Zugänglichmachung für Unterricht und Forschung in unserem Fall nicht greift. Hinsichtlich datenschutzrechtlich geschützter Ressourcen müssen diese nach Möglichkeit in anonymisierter Form vorliegen und die Einverständniserklärungen müssen auch die Verfügbarmachung für die

²¹ vgl. Metschke/Wellbrock 2002, S. 30f.

Community abdecken. Fehlende, aber erforderliche Rechte im Nachhinein einzuholen ist oft mit einem erheblichen Aufwand verbunden bzw. unmöglich. Aufgrund der teilweise relativ offenen Formulierungen der einschlägigen gesetzlichen Regelungen ist es nach unserer Interpretation der zur Anwendung kommenden Normen empfehlenswert, immer vertraglich die Erhebung, die Bereitstellung, die Archivierung und die Distribution von Sprachressourcen zu regeln. Verträge müssen zwischen unterschiedlichen Akteuren geschlossen werden. Die Gruppe der Akteure ist abhängig von den Ressourcen und den Ressourcentypen und umfasst u.a. Urheber (z.B. Autoren), Verwerter (z.B. Verlage), Ressourcen-Provider (z.B. LRT-Zentren) und Endnutzer.

In Lizenzverträgen zwischen Ressourcen-Providern auf der einen Seite und Urhebern/ Verwertern auf der anderen Seite muss auf die Umgebung Bezug genommen werden, in der die Ressourcen von Endnutzern verwendet werden. Diese Umgebung muss im Hinblick auf die Sicherheit der Authentifizierung und Autorisierung dem aktuellen Stand der technischen Entwicklung entsprechen. Nur so kann gewährleistet bzw. glaubhaft vermittelt werden, dass Datenzugriffe nicht durch Unberechtigte erfolgen können und dass die Nutzung durch Berechtigte im Hinblick auf die Einhaltung der Lizenzbedingungen ggf. kontrolliert werden kann. Die Zusicherung der Anwendung der besten verfügbaren technischen Möglichkeiten in diesem Bereich wird auch in zunehmendem Maße die Voraussetzung bilden, mit Ressourcengebern Lizenzvereinbarungen abschließen zu können. Die gilt nicht zuletzt, da in der D-SPIN/CLARIN-Infrastruktur auch solche Ressourcen zur Verfügung gestellt werden können müssen, die nur mit kostenpflichtigen Lizenzen genutzt werden dürfen.

Literaturverzeichnis

Bundesverfassungsgericht: Urteil vom 29.05.1973. In: Entscheidungen des Bundesverfassungsgericht, Jg. 35, S. 79–170.

Bundesverfassungsgericht: Urteil vom 15.12.1983. In: Entscheidungen des Bundesverfassungsgericht, Jg. 65, S. 1–71.

Czychowski, Christian (2003): § 13. Elektronischer Handel mit Geistigem Eigentum. In: Bröcker, Klaus Tim / Czychowski, Christian / Schäfer, Detmar (Hg.): Praxishandbuch Geistiges Eigentum im Internet. Gewerblicher Rechtsschutz, Urheberrecht, angrenzende Rechtsgebiete. München, S. 841–948.

Deutscher Bundestag: Beschlussempfehlung und Bericht des Rechtsausschusses (6. Ausschuss). Drucksache 15/837.

Dreier, Thomas / Schulze, Gernot (Hg.) (2008): Urheberrechtsgesetz. Urheberrechtswahrnehmungsgesetz, Kunsturhebergesetz. Kommentar. 3. Aufl. München.

Dustmann, Andreas (2008): § 52a. Öffentliche Zugänglichmachung für Unterricht und Forschung. In: Nordemann, Wilhelm / Fromm, Friedrich Karl (Hg.): Urheberrecht. Kommentar zum Urheberrechtsgesetz, Verlagsgesetz, Urheberrechtswahrnehmungsgesetz. 10. Aufl. Stuttgart, S. 983–989.

Gola, Peter / Schomerus, Rudolf (Hg.) (2007): BDSG Bundesdatenschutzgesetz. Kommentar. 9. Aufl. München.

- Hertin, Paul W. (2008): Urheberrecht. Zum "2. Korb"! 2. Aufl. München.
- Holznagel, Bernd / Sonntag, Matthias (2003): Die Einwilligung des Betroffenen. In: Roßnagel, Alexander (Hg.): Handbuch Datenschutzrecht. Die neuen Grundlagen für Wirtschaft und Verwaltung. München, S. 678–714.
- Jarass, Hans D. / Pieroth, Bodo (Hg.) (2007): Grundgesetz für die Bundesrepublik Deutschland. Kommentar. 9. Aufl. München.
- Ladeur, Karl-Heinz (2008): § 7. Das allgemeine Persönlichkeitsrecht als Grundrecht. In: Götting, Horst-Peter / Schertz, Christian / Seitz, Walter (Hg.): Handbuch des Persönlichkeitsrechts. München, S. 121–129.
- Loewenheim, Ulrich (2006): § 52a. Öffentliche Zugänglichmachung für Unterricht und Forschung. In: Schrickler, Gerhard (Hg.): Urheberrecht. Kommentar. 3. Aufl. München, S. 1062–1069.
- Lüft, Stefan (2009): § 52a. Öffentliche Zugänglichmachung für Unterricht und Forschung. In: Wandtke, Artur-Axel / Bullinger, Winfried (Hg.): Praxiskommentar zum Urheberrecht. 3. Aufl. München, S. 766–772.
- Metschke, Rainer / Wellbrock, Rita (2002): Datenschutz in Wissenschaft und Forschung. 3. Aufl. Berlin.
- Pieroth, Bodo / Schlink, Bernhard (2008): Grundrechte. 24. Aufl. Heidelberg (= Schwerpunkte, 14,1).
- Rehbinder, Manfred (2008): Urheberrecht. Ein Studienbuch. 15. Aufl. München.
- Schack, Haimo (2007): Urheber- und Urhebervertragsrecht. 4. Aufl. Tübingen.
- Simitis, Spiros (Hg.) (2006): Bundesdatenschutzgesetz. 6. Aufl. Baden-Baden.
- Tinnefeld, Marie-Theres / Ehmann, Eugen / Gerling, Rainer W. (2005): Einführung in das Datenschutzrecht. Datenschutz und Informationsfreiheit in europäischer Sicht. 4. Aufl. München / Wien.