

21 Forschungsdatenmanagement und Nachnutzung der Daten

Kristina Petzold, Ulrich Heid

21.1 Einleitung – Wozu Forschungsdatenmanagement?

»Qualitätsgesicherte Forschungsdaten bilden einen Grundpfeiler wissenschaftlicher Erkenntnis«, heißt es in der Präambel der »Grundsätze zum Umgang mit Forschungsdaten« der Allianz der Wissenschaftsorganisationen aus dem Jahr 2010. Allerdings geht ein großer Teil dieser so wichtigen Daten mitunter nach dem Ende der Projekte verloren (vgl. Büttner et al. 2011: 5). Dabei sind die z. T. mühsam gesammelten und produzierten Daten wie – im Falle von Rez@Kultur – Textkorpora, Interviewtranskripte und methodische Zwischenprodukte wie Annotationen, Fragebögen und Auswertungsalgorithmen nicht nur für die aktuelle Forschungsfrage relevant, sondern auch für eine mögliche Nachnutzung – sei es zum Zweck der Kontrolle früherer Ergebnisse oder mit ganz neuen Fragestellungen.³⁵ So heißt es denn auch weiter im Text der Allianz: »Die nachhaltige Sicherung und Bereitstellung von Forschungsdaten dient daher nicht nur der Prüfung früherer Ergebnisse, sondern in hohem Maße auch der Erzielung künftiger Ergebnisse« (Allianz der Wissenschaftsorganisationen 2010). Die darauf aufbauende allgemeine Forderung nach einem verantwortungsvollen und nachhaltigen Umgang mit Forschungsdaten hat vor dem Hintergrund der Digitalisierung in den letzten Jahren nochmals an Wichtigkeit gewonnen. Denn digitale Daten erhöhen die Zugänglichkeit von großen Datenmengen und ermöglichen zunehmend auch deren Auswertbarkeit, sodass das Interesse an einem strukturierten und reflektierten Umgang mit Forschungsdaten stetig wächst (vgl.

35 Vgl. hierzu auch die Präambel der »Grundsätze zum Umgang mit Forschungsdaten« der Allianz der deutschen Wissenschaftsorganisationen aus dem Jahr 2010.

Büttner et al. 2011: 5). Mit dem Interesse sind allerdings auch Anforderungen verbunden, sodass sich in vielen Punkten sowohl Forschende als auch Institutionen in einem Lern- und Entwicklungsprozess befinden.

Große Datenmengen in hoher Qualität langfristig und gut zugänglich zu archivieren setzt nicht nur Überlegungen z. B. zur Auffindbarkeit durch Suchanfragen, zum Standort von Hardware und zur langfristigen Kompatibilität von Software voraus, sondern auch die Klärung rechtlicher Fragen. Forschungsdatenmanagement (FDM) umfasst also »verschiedene Vorkehrungen und Maßnahmen [...], die bereits bei der Planung und Durchführung von Forschungsprojekten berücksichtigt werden müssen« (RatSWD 2016: 3). Auch im Fall des Rez@Kultur-Projektes war das FDM von Anfang an integraler Bestandteil des Forschungsprozesses, weshalb ihm im Rahmen dieser Publikation ein eigenes Kapitel gewidmet wird. Die nachfolgenden Erläuterungen zum FDM in Rez@Kultur orientieren sich am Forschungsdatenzyklus und gliedern sich entsprechend in Datenerstellung, Datenverarbeitung und -analyse sowie Metadaten, Datenarchivierung und Nachnutzung (vgl. ICPSR 2012: 8). Innerhalb dieser drei Abschnitte des Datenlebenszyklus werden jeweils spezifische Problemlagen geschildert, das Vorgehen sowie die Lösungen im Datenmanagement von Rez@Kultur beschrieben und dabei auf konkrete technische und rechtliche Aspekte eingegangen. Diesen Abschnitten vorangestellt ist eine kurze Beschreibung der in Rez@Kultur erhobenen und produzierten Datentypen.

21.2 Datentypen in Rez@Kultur

Wie bereits weiter oben beschrieben, handelt es sich bei Rez@Kultur um ein gemischmethodisches Forschungsdesign, dessen zentrales Charakteristikum in der Heterogenität seiner Daten besteht. Bei den zu analysierenden Primär- bzw. Rohdaten handelte es sich um die drei folgenden Datentypen:

- *Typ A:* 6 qualitative Interviews, die von Rez@Kultur geführt wurden, im Audio-Format vorlagen und nach fachspezifischen Regeln transkribiert und in MAXQDA codiert wurden.
- *Typ B:* 36 Rezensionendateien zur qualitativen Analyse, die uns als PDF-Screenshots vorlagen und ebenfalls in MAXQDA codiert wurden.

- *Typ C*: Insgesamt 231.854 Rezensionen von 3 großen Plattformen und 20 Weblogs für die quantitative Analyse, die im XML-Format und mit UTF-8-Zeichenkodierung vorlagen und mit linguistischen Tools nach fachspezifischen Richtlinien annotiert wurden (z. B. Tokenisierung, Lemmatisierung, POS-Tagging, Sentimentanalyse, vgl. Kapitel 8).

Die bei der Datenverarbeitung und -analyse entstandenen Sekundärdaten, wie Annotationen und Codierungen zu Rezensionstexten sowie Ergebnisse verschiedener automatischer Analyseverfahren, werden ebenfalls langfristig archiviert. Da sie aber sowohl rechtlich als auch technisch unproblematisch verarbeitet und archiviert werden konnten, werden sie nachfolgend nicht gesondert betrachtet. Die Beschreibung der Daten wurde zu Beginn des Projekts in einem Datenmanagementplan vorgenommen, der mithilfe des DMPTY-Programms erstellt wurde (vgl. »Datenmanagementplan« von CLARIN-D).

Schon die Datenerhebung erforderte die Klärung der rechtlichen Rahmenbedingungen. Neben forschungsethischen Fragestellungen zählten dazu das Urheberrecht (UrhG), das Urheberrechts-Wissensgesellschafts-Gesetz (UrhWissG) sowie die EU-Datenschutzgrundverordnung (DSGVO). Erschwerend kam hinzu, dass die beiden letztgenannten erst im Projektverlauf in Kraft traten. Die Unsicherheit in Bezug auf die Auslegung einzelner Paragraphen des UrhWissG und der EU-DSGVO aufgrund noch ausstehender Gerichtsurteile verlangsamte den institutionellen Entscheidungsprozess zusätzlich.

21.3 Erhebung der Interviewdaten (A)

Für die Nutzung der qualitativen Interviewdaten wurde entsprechend bildungswissenschaftlicher Fachkonventionen eine Einverständniserklärung erstellt, welche von allen Teilnehmer_innen jeweils vor dem Interview unterschrieben wurde. Damit wurde das Kriterium der »eindeutige[n] bestätigende[n] Handlung [...], mit der freiwillig, für den konkreten Fall, in informierter Weise« Zustimmung bekundet wird, erfüllt (vgl. Abs. 32, Amtsblatt EU 4/5/2016, 119/6). Entsprechend der EU-DSGVO enthielt die Einverständniserklärung Informationen sowohl zum Zweck und zur Art der Verarbeitung als auch zur Nachnutzung. Es wurde die Möglichkeit angeboten, das

Einverständnis jederzeit zurückzuziehen und eine Löschung zu veranlassen. Daher mussten bereits zu diesem frühen Zeitpunkt die Optionen der Langzeitarchivierung berücksichtigt werden. Zum Zweck der Anonymisierung wurden biografische und demographische Informationen in einem separaten Fragebogen erhoben und getrennt von den inhaltlich erhobenen Daten aufbewahrt. Die Verarbeitung der Daten erfolgte ausschließlich in anonymisierter Form, d. h. ohne Angabe des Namens. Informationen, die Rückschlüsse auf die Identität der Interviewpartner_in zuließen, wurden anhand von Anonymisierungsregeln bei der Transkription der Audio-Interviews eliminiert. Das nachfolgende Beispiel verdeutlicht, welche Relevanz die Anonymisierung der Transkripte haben kann: »[D]ann [Pause] denke ich mir schon öfter mal, das muss jetzt schon auch gut sein [Eigener Name], also du kannst da jetzt nicht so irgend so einen Mist veröffentlichen« (IP 02: 641-643). Die Anonymisierungs- bzw. Pseudonymisierungsregeln wurden in einem separaten Dokument gespeichert, welches ebenfalls mit den gesammelten Daten archiviert wird.

Bei der Erstellung der Einverständniserklärung nach EU-DSGVO war die Konsultation des Datenschutzbeauftragten der Universität Hildesheim unumgänglich. Dabei gingen die juristischen Meinungen aufgrund der Novität des Gesetzes teilweise auseinander. So sahen beispielsweise einige Jurist_innen in der o. g. Form der Anonymisierung den Wunsch auf Löschung personenbezogener Daten bereits erfüllt, sodass diese nicht zusätzlich hätte in der Einverständniserklärung angeboten werden müssen.

21.4 Erhebung qualitativer und quantitativer Rezensionsdaten (B und C)

Bei der Erhebung von digitalen rezensiven Äußerungen stellten sich Herausforderungen auf mindestens drei Ebenen: (1) forschungspraktische Fragen, (2) technische Fragen und (3) rechtliche Fragen.

Auf der forschungspraktischen Ebene musste geklärt werden, welche Arten von Online-Rezensionen und -Plattformen im Rahmen von Rez@Kultur analysiert werden sollten. Diese Sampling-Entscheidungen waren inhaltlicher Natur und werden in Kapitel 2 sowie in Kapitel 3 erläutert.

Die technischen Fragen ergaben sich in erster Linie aus der Suche nach der besten Offline-Speicherlösung für digitale rezensive Texte. Auch wenn

es trivial erscheint, ergaben sich bei der technischen Umsetzung der Datenerhebung von Online-Rezensionen selbst bei kleinen Mengen bereits Hürden. So ist es entsprechend der Empfehlungen zur Online-Forschung (vgl. Marx 2017: 105) vorgesehen, wenigstens die qualitativ zu analysierenden Rezensionen nicht nur als Textdateien (XML) zu speichern, sondern inklusive der grafischen Seiteninformationen als PDF. Die Suche nach einem geeigneten Programm zum Export all dieser Daten erwies sich als sehr komplex und die Lösung bis zum Schluss nicht als vollständig befriedigend. Am Ende wurden je nach Seiten- und Quelltextstruktur verschiedene Programme zum Erstellen von PDF-Dateien verwendet u. a. der MAXQDA Web Collector und verschiedene Add-ons zu Mozilla Firefox (u. a. PDF Mage).

Für die Erhebung von Online-Rezensionen als Massendaten ergaben sich andere technische Herausforderungen. Hier spielte die Verfügbarkeit von API-Schnittstellen eine wichtige Rolle sowie die Möglichkeit, Webinhalte mithilfe einer Software zu crawlen. In den meisten Fällen konnte die inhaltlich begründete Auswahl von Plattformen und Weblogs jedoch technisch umgesetzt werden. Erstaunlich war dabei eher, wie offen einige vor allem private Plattformen mit ihren Mitgliederdaten umgingen. Auf *BücherTreff.de* ließen sich beispielsweise nicht nur Rezensionen, sondern auch Profildaten der Nutzer_innen abrufen. Nicht möglich war die Erhebung von Daten auf der Plattform *LovelyBooks*.

Die größten Herausforderungen bei der Sammlung von Online-Rezensionen waren aber nicht inhaltliche oder technische Aspekte, sondern rechtliche. Im Zentrum stand dabei die Frage, ob Online-Rezensionen im Rahmen der Forschung überhaupt erhoben und verarbeitet werden dürfen und wenn ja unter welchen Voraussetzungen. Muss dafür das Einverständnis der Urheber_innen eingeholt werden (was auf Plattformen wie Twitter umständlich und in vielen Fällen auch schlicht nicht möglich ist)? Müssen die Daten im Sinne der EU-DSGVO anonymisiert werden oder fallen sie – im Gegenteil – unter das Urheberrecht und eine Namensnennung ist zwingend erforderlich? Die juristischen Stellungnahmen, die Rez@Kultur zur Klärung dieses Problems einholte, fielen in ihrer Einschätzung z. T. sehr unterschiedlich aus.

Zentral für die Lösung dieses Problems war die Einschätzung dazu, ob Online-Rezensionen als personenbezogene Daten anzusehen sind oder ob sie als publizierte Werke mit einer Schöpfungshöhe angesehen werden. Letzteres würde einerseits urheberrechtlichen Schutz rechtfertigen und

gleichzeitig die Nutzung im Rahmen der Forschungsschranke des UrhWissG erlauben. Gegen eine vollständige Anonymisierung der Daten sprachen letztendlich zwei Argumente: Erstens wäre eine Anonymisierung von Online-Veröffentlichungen in Zeiten von Google faktisch nicht möglich, da Texte häufig über die Suchmaschine gefunden und somit der Verfasser_in wieder zugeordnet werden könnten. Zweitens wäre ein solches Vorgehen juristisch gar nicht nötig, da die beabsichtigte Datenverarbeitung durch Art. 6 Abs. 1 lit. f DSGVO gedeckt ist, denn die berechtigten Interessen der Forschenden an der Durchführung der Forschung dürften die Interessen der Betroffenen überwiegen, die ihre Beiträge selbst veröffentlicht haben und damit gegebenenfalls ihre persönlichen Daten selbst in öffentlich zugängliche Quellen eingegeben haben. Von einer Veröffentlichung der Daten wird ausgegangen, wenn es sich um offene Webseiten, Plattformen und soziale Netzwerke handelt, bei denen mehr als nur ein Freundeskreis erreicht wird (vgl. § 15, Abs. 3, Satz 1 UrhG) und zu dem jede_r grundsätzlich (möglicherweise nach einer Anmeldung) Zugriff hat. Eine differenzierende Lösung, bei der in jedem Einzelfall über die Schöpfungshöhe und die Frage nach Klarnamen entschieden werden muss, wurde sowohl aus pragmatischen als auch aus juristischen Gründen (u. a. der Verhältnismäßigkeit) verworfen. Am Ende der Konsultationen und vieler projektinterner Diskussionen wurden die folgenden Beschlüsse gefasst:

- a. Die Nutzung von online frei verfügbaren Rezensionsdaten durch Rez@Kultur wird als datenschutz- und urheberrechtskonform angesehen.
- b. Einer Anonymisierung bedarf es nicht – auch nicht bei der Zitation.
- c. Größere Rezensions-Plattformen, auf denen sich jeder anmelden kann, werden als Öffentlichkeit im Sinne des UrhG (§ 15) betrachtet. Nur wenn Texte von geschlossenen Foren genutzt werden sollen, wird die Einverständniserklärung der Urheber_in eingeholt.
- d. Das Verfassen von Rezensionen unter Pseudonym wird als Wille der Rezensent_in gedeutet, anonym zu bleiben. Dieser Wille wird von Rez@Kultur respektiert, indem in diesem Fall bei der Datenverarbeitung, -weitergabe und -zitation kein Klarnamen genannt wird, auch wenn dieser möglicherweise ermittelt werden kann.

Im Ergebnis konnte die Mehrzahl der von Rez@Kultur inhaltlich vorgesehenen Online-Rezensionen zum Zwecke der Forschung gecrawlt und analysiert

werden. Ausnahmen bildeten Plattformen, zu denen entweder kein technischer Zugang bestand oder die als geschlossene Foren betrachtet wurden wie beispielsweise *Lectory*.

21.5 Praktische Fragen der Datenverarbeitung und -analyse

Für die projektinterne gemeinsame Arbeit an den o.g. Forschungsdaten wurden verschiedene Verarbeitungs- und Speicherlösungen gefunden und entwickelt, die praktischen Anforderungen (wie der einfachen Zugänglichkeit durch alle Projektbeteiligten) gerecht werden mussten, aber auch die Einhaltung vertraglicher und rechtlicher Rahmenbedingungen gewährleisten sollten. Gemeint sind dabei insbesondere die Wahrung der Anonymität der Interviewpartner_innen durch separate Speicherung von Interviewdaten und Klarnamen sowie das ausschließliche Zugriffsrecht auf die Daten zum Zwecke der Forschung (vgl. Forschungsschranke § 60c, Abs. 1.1, UrhWissG).

Als praktikabel erwies sich die gemeinsame Nutzung von Dokumenten in der niedersächsischen Hochschulcloud *Academic Cloud*, wo Zugänge und Speicherplatz für alle Mitarbeiter_innen des Projekts zur Verfügung standen. Da diese allerdings keine Backup-Funktion anbietet, mussten Backups regelmäßig manuell auf lokalen Rechnern durchgeführt werden. Für die Massendatenkorpora musste zusätzlich ein eigener Serverspeicherplatz am Rechenzentrum der Universität Hildesheim eingerichtet werden.

Als wichtiges gemeinsames Werkzeug im Bereich der qualitativen Analyse fungierte die Codiersoftware MAXQDA. Alle Speicherorte und Programme waren durch Passwörter oder individuelle Lizenzen vor dem Zugriff Unbefugter geschützt. Insgesamt könnte man den organisatorischen Aufwand in Einzelprojekten aber sicherlich verringern, wenn auf Landes- oder Universitätsebene Standardlösungen z. B. zum Projektmanagement, zum Datenmanagement oder auch zur Speicherung und Sicherung (Stichwort Backup) auch großer Datenmengen zur Verfügung gestellt würden.

21.6 Metadaten

Unter Metadaten werden hier alle nicht im Korpus selbst enthaltenen Informationen verstanden, die einzelne Einheiten von Forschungsdaten (z. B. Plattformkorpus, Einzelrezension oder Interview) spezifizieren. Auch die Metadaten zu den in Rez@Kultur erhobenen Forschungsdaten mussten unterschiedlichen Anforderungen gerecht werden. Dazu zählen allein die unterschiedlichen disziplinären Standards und Datenformate, Unterschiede in den Auswertungszielen bei den verschiedenen Datenformaten sowie verschiedene rechtliche Rahmenbedingungen. Während es z. B. bei der massenhaften Analyse von Rezensionsdaten sinnvoll ist, die Rezensionen über ein Metadatum ›Urheber_in‹ identifizierbar zu halten, um beispielsweise Netzwerkanalysen leichter durchführen zu können, ist ein solches Vorgehen bei den qualitativen Interviews gar nicht zulässig, da Anonymisierung garantiert wurde. Ein weiteres Beispiel ist das Erstellungsdatum, das bei qualitativen Interviews zwar als Metadatum hinterlegt wird, aber von deutlich geringerem Interesse bei der Analyse ist, als das Publikationsdatum von Online-Rezensionen. Aus letzteren konnten im Rahmen der Massendatenauswertungen beispielsweise Zeitverlaufsanalysen generiert werden. Die Erstellung einer verbindlichen Metadatenliste erforderte daher nicht nur die Auseinandersetzung mit den jeweiligen Analysezielen und unterschiedlichen disziplinären Standards, sondern auch Differenzierungen zwischen Typen von Quellen. Nicht zuletzt mussten die Möglichkeiten des Datenrepositoriums HilData zur Hinterlegung von Metadaten berücksichtigt werden. Im Ergebnis entstand die folgende Metadatenliste, die sich erstens im Blick auf Interview- und Rezensionsdaten unterscheidet und sich zweitens grob in die verschiedenen Metadatendimensionen (1) Werk, (2) Rezension, (3) Plattform, (4) Analyse und (5) allgemeine Metadaten unterteilen lässt:

Tabelle 21.1: Metadaten Rez@Kultur

Rezensionen (Massendaten)	
(1) Werk	Autor (wenn vorhanden)
	Titel
(2) Rezension	Rezensions-ID
	Autor bzw. Autor-ID
	Titel
	Veröffentlichungsdatum
	Länge der Rezension
	Publikationsort/Plattform
	Anzahl der Kommentare (wenn vorhanden)
	Anzahl der Likes (wenn vorhanden)
Anzahl vergebener Sterne (wenn vorhanden)	
(3) Plattform	Plattformname
	Plattformtyp
	Öffentlichkeitslevel/Zugangstyp
	Informationen zur Erhebung (Suchstrategie)
	Erhebungszeitraum
(4) Methodische Metadaten	Clusterzugehörigkeit der Rezension
	Schlagworte
	Daten zur computerlinguistischen Annotation/Prozessierung
(5) Formale Metadaten	Dokumentname
	Ressource/Erreichbarkeit
	Qualitätssicherung/Aufbereitung
	Datenstruktur/Datenorganisation
	Format
	Technische Daten
	Rechtliche Daten (Aufbewahrungsdauer, Zugangslizenz)

Rezensionen (Einzelfalldaten)	
(1) Methodische Metadaten	Arbeitsschritte
	Sample-Kriterien
(2) Formale Metadaten	Dateityp
	Dateiformat
	Sprache
(3) Inhaltliche Metadaten	Plattformentyp
	Erscheinungsdatum
Qualitative Interviews	
(1) Methodische Metadaten	Arbeitsschritte
	Sample-Kriterien
	Erhebungsmethode
(2) Formale Metadaten	Dateityp
	Dateiformat
	Sprache
	Erhebungsort
	Erhebungszeit
	Dauer
	Dateigröße
(3) Rechtliche Metadaten	Anonymisierung
	Einverständniserklärung
	Speicherung
(4) Inhaltliche Metadaten	Alter (aggregiert)
	Geschlecht
	Sparte (Kunst oder Literatur)
	Startzeitpunkt der Veröffentlichung von Online-Rezensionen
	Häufigkeit der Veröffentlichung von Online-Rezensionen

Quelle: Kristina Petzold/Universität Hildesheim

21.7 Datenarchivierung und Nachnutzung

In Deutschland gibt es derzeit mehrere regionale und nationale Datenrepositorien, die mit unterschiedlichen Speicher- und Katalogisierungssystemen (und Finanzierungsmodellen) die Forschungsdaten verschiedener Wissenschaftsdisziplinen langfristig archivieren. Ein interdisziplinäres Projekt wie Rez@Kultur steht bei der Auswahl eines passenden Datenrepositoriums vor besonderen Herausforderungen: Welches Repositorium ist technisch in der Lage und geeignet, die unterschiedlichen Datentypen zu archivieren? Welches Repositorium ist fachlich geeignet? Und ist es zielführend, die Daten eines Projekts zu separieren und in verschiedenen Repositorien zu archivieren, damit die disziplinäre Zuordnung zum Repositorium jeweils gewahrt bleibt? Oder sollten die gesamten Projektdaten besser vollständig an *einem* Ort archiviert werden, mit dem Nachteil, dass sie für Einzeldisziplinen möglicherweise schlechter auffindbar sind? Welche rechtlichen Rahmenbedingungen müssen vom Projekt und vonseiten des Repositoriums beachtet werden? Wer darf Zugriff auf die Daten erhalten und wie lange? Und wie kann man die Suche nach Antworten auf diese Fragen sinnvoll und ressourcenorientiert gestalten und möglichst früh im Forschungsprozess berücksichtigen?

Auf dem Weg zur Beantwortung dieser Fragen war ein wichtiger Schritt die sondierende Korrespondenz mit mehreren Repositorien. Eine stärkere Vernetzung der Repositorien untereinander sowie der Ausbau und die zunehmende Bekanntheit allgemeiner Standards (wie beispielsweise CMDI oder Dublin Core) und zentraler Informationspunkte könnten diesen Prozess sicher optimieren. Auch die beratende Unterstützung des Referats Forschungsdatenmanagement sowie die juristische und datenschutzrechtliche Beratung an der Universitätsbibliothek Hildesheim waren sehr wertvoll.

Die Überlegungen führten zu dem Ergebnis, die Forschungsdaten von Rez@Kultur im universitätseigenen Datenarchiv HilData zu hinterlegen und in möglichst vielen Datenkatalogen anderer Repositorien zu verlinken. Diese Entscheidung hatte in erster Linie rechtliche Gründe. Denn während die anonymisierte Archivierung von Interviewdaten (A) durch die Einverständniserklärungen zulässig war, traf dies auf die gecrawlten Rezensionen (B und C) nicht zu. Die Weitergabe dieser Daten an ein Forschungsdatenrepositorium wurde sowohl vom Justizariat als auch vom Datenschutzbeauftragten der Universität Hildesheim als *problematisch* eingestuft, da wir

gerade nicht über die notwendigen Rechte zur Weitergabe verfügen, sondern lediglich auf Basis der Forschungsschranke die Speicherung und Verarbeitung zu eigenen wissenschaftlichen Zwecken gerechtfertigt war. Dass Datenrepositorien auf derartige Fälle (die ja sicher in Zukunft häufiger sein dürften, da sie jede Art von gecrawlten Online-Daten betreffen) bisher nicht eingestellt sind, zeigte sich daran, dass sie keine andere Vertragsoption anboten außer der Rechteübertragung vom Datengeber auf das Repositorium. Hier sind Datenschutz und FDM-Ziele im Widerspruch. Dass es dabei auch Informationslücken und Anpassungsprobleme gibt, wird in einer Stellungnahme der Deutschen Gesellschaft für Soziologie (DGS) deutlich. Diese betont das Recht der »Primärforschenden [...], die Bedingungen für die Nachnutzung einzuschränken«, verweist auf den »Unterstützungsbedarf« durch Forschungseinrichtungen und Datenzentren und auf die Gewährleistungspflicht der letzteren, die Daten in »verantwortlicher und fachlich kompetenter Weise« zu speichern (DGS 2019).

Im Fall von Rez@Kultur konnte keine Einigung mit einem externen Datenrepositorium erzielt werden – wobei dies letztlich nicht an der mangelnden Kooperationsbereitschaft der Repositorien lag, sondern an der Einschätzung der Rechtssituation. Eine immerhin gangbare, rechtskonforme Lösung wurde darin gefunden, dass die Forschungsdaten von Rez@Kultur vom Projektteam im Datenrepositorium der Universität Hildesheim selbst verwaltet werden. In diesem Fall muss keine Rechteweitergabe an einen externen Partner vorgenommen werden, weil die Daten faktisch beim Projektteam verbleiben. Auch der Zugang für andere Forscher_innen kann vorerst nicht gestattet werden. Auch wenn diese Lösung den Maximen von Open Science und Nachnutzung kaum gerecht wird, wird damit immerhin der Verlust der Daten nach dem Projektende vermieden und die Option offengehalten, die Daten in Zukunft nach einer Änderung der Rechtslage doch noch der wissenschaftlichen Community zugänglich zu machen.

Die Interviewdaten (A) werden ebenfalls in HilData hinterlegt, allerdings mit einer Lizenz zur freien Nutzung im Rahmen wissenschaftlicher Forschung. Um die Daten von Rez@Kultur auch jenseits der Universität Hildesheim auffindbar zu machen, werden die Meta-Daten sowohl der Interviewdaten (A) als auch der Rezensionsdaten (B und C) im Katalog des Verbunds Forschungsdaten Bildung und im Netzwerk CLARIN-D indiziert.

21.8 Lessons Learned oder Empfehlungen für ein Forschungsdatenmanagement der Zukunft

Die Gestaltung und Umsetzung eines praktikablen, rechtskonformen und nachhaltigen FDM für das Rez@Kultur-Projekt erforderte Zeit- und damit Arbeitsressourcen. Es war daher absolut lohnenswert, dies auch bereits bei der Projektplanung und bei der Beantragung von Fördergeldern zu berücksichtigen. Eine ausreichende Arbeitszeitkalkulation sowie die frühe Planung und Auseinandersetzung mit technischen, rechtlichen und organisatorischen Aspekten des Forschungsdatenmanagements verhindert nicht, dass Umwege gegangen werden (müssen) und garantiert auch nicht, dass befriedigende Lösungen für alle Probleme gefunden werden können, aber es *erlaubt und ermöglicht* die (derzeit noch) notwendige *Suche* nach den bestmöglichen Lösungen.

Als besonders hilfreich erwiesen sich dabei für das Rez@Kultur-Projekt die verschiedenen Ansprechpartner_innen und Expert_innen an und jenseits der Universität Hildesheim, wie die Stelle für Forschungsdatenmanagement, der Datenschutzbeauftragte Thomas Mandl und das Justizariat der Universität Hildesheim. Perspektivisch ist der weitere Ausbau der Zusammenarbeit dieser Stellen zum Zwecke der ganzheitlichen Beratung der Forschenden hinsichtlich des Forschungsdatenmanagements sicherlich sinnvoll.

Auch in Bezug auf den Workflow und die unterschiedlichen Phasen und Segmente des Forschungsdatenmanagements wären ganzheitliche Lösungen wünschenswert. Es gibt zwar inzwischen Angebote beispielsweise zum Erstellen und Versionieren von Datenmanagementplänen; nach unserem Wissensstand existieren derzeit aber keine Softwarelösungen, die den gesamten Prozess des Forschungsdatenmanagements, inklusive der Klärung rechtlicher und technischer Fragestellungen, in einer integrativen Anwendungsumgebung abbilden. Auch das z. T. notwendige Aufteilen einzelner Teilfragen und Schritte (wie beispielsweise das Erstellen von Einverständniserklärungen für qualitative Interviews) wäre einfacher, wenn diese in einem solchen System systematisch eingegliedert und nachvollziehbar wären.

Die beschriebenen Schwierigkeiten bei der Suche nach einem geeigneten Forschungsdatenrepositorium ließen sich in Zukunft vermeiden oder wenigstens reduzieren, wenn eine stärkere Standardisierung und Zentralisierung stattfinden würde bzw. die bisherigen Anstrengungen in diese

Richtung (etwa im Rahmen der Nationalen Forschungsdateninfrastruktur – NFDI) weiter ausgebaut werden. In diesem Zusammenhang wäre es auch wünschenswert, klare Verhältnisse hinsichtlich der Finanzierungsmodelle zu schaffen und dabei sowohl die Interessen der Repositorien als auch die der Datengeber_innen zu berücksichtigen. Bisher scheinen die Meinungen darüber noch sehr weit auseinander zu gehen, ob es sich bei der Archivierung von Forschungsdaten um eine Dienstleistung des Repositoriums oder um eine Bereitstellung (im Sinne einer Datenspende) seitens der Datengeber_innen handelt. Hier bedarf es zukünftig noch einiger Aushandlungen. Andere Probleme, die beispielsweise auf dem mangelnden Erfahrungsschatz einiger Repositorien beruhen, lösen sich mit der Zeit vermutlich von selbst.

Auf der Basis unserer Erfahrungen im FDM schließen wir uns der kritischen Stellungnahme der Deutschen Gesellschaft für Soziologie (2019) an, in der darauf hingewiesen wird, dass die Möglichkeiten des Forschungsdatenmanagements – und insbesondere die der Sekundärnutzung von Daten – nicht die Entscheidung über die Bewilligung von Geldern und damit die Ausrichtung der Forschungsfrage beeinflussen sollten, ebenso wenig, wie sie den Feldzugang einschränken dürfen. Auch beim FDM gilt es, zwischen unterschiedlichen, teilweise konträren Zielsetzungen abzuwägen. Vorgehensweisen und Lösungen, die dabei für das Forschungsprojekt Rez@Kultur sinnvoll waren, müssen dies nicht im selben Maße für andere Projekte sein. Aber wir hoffen, dass wir mit dieser Darstellung zur Transparenz und Verbesserung des FDM im Allgemeinen einen Beitrag leisten konnten. Denn auch im Rahmen unserer Projektarbeit hat sich gezeigt, wie wichtig und lohnend es ist, seine Forschungsdaten umsichtig zu verwalten.