

Wie offen können und dürfen Forschungsdaten sein?

Marius Gerecht & Helge Kminek

1 Strategien und Beispiele aus der empirischen Bildungsforschung

Im Zuge der Debatten und Initiativen rund um das Thema „Open Access“ wird auch die Forderung nach einem offenen Zugang zu Forschungsdaten immer lauter. Ein Blick in die Naturwissenschaften zeigt, dass die Zugänglichmachung von Forschungsdaten ein Standard wissenschaftlichen Arbeitens sein kann und auch sein sollte. Dies gewährleistet ein transparentes Forschungshandeln und ermöglicht, Forschungsbefunde nachzuvollziehen, zu replizieren, zu falsifizieren oder gegebenenfalls auch zu revidieren und im Zuge von Sekundäranalysen mehrfach zu nutzen. Letzteres schützt das beforschte Feld präventiv vor Überforschung und erhält so die für die Wissenschaft nötigen Feldzugänge. Der Austausch von und über Forschungsdaten erhöht nicht nur das analytische Auswertungspotenzial der Scientific Community, sondern ermöglicht auch einen verbindlicheren Forschungsdiskurs. Was in anderen Disziplinen schon längst Usus ist, nämlich die Einreichung der dazugehörigen Forschungsdaten zu einer Publikation, wird beispielsweise in der Psychologie, die sich empirisch stark an naturwissenschaftlichen Methoden orientiert, für zukünftige Publikationsstrategien ein Standard wissenschaftlichen Arbeitens sein. In Sozialwissenschaften wie der Soziologie und Politologie ist die Nachnutzung amtlicher Statistiken meist das erste Mittel der Wahl und die Sekundäranalyse nahezu Forschungsalltag. In der empirischen Bildungsforschung tut man sich bisher schwer damit, das analytische Potenzial der vorhandenen Daten vollständig auszuschöpfen. Meist ist die Analyse auf das einzelne Forschungsprojekt selbst beschränkt, und nur in wenigen Ausnahmen, beispielsweise bei der PISA-Studie, bleiben die Daten für nachfolgende Forschergruppen interessant. Die Einrichtung des Leibniz-Instituts für Bildungsverläufe (LIfBi) ist ein Beispiel dafür, dass man auch in der Bildungsforschung verstärkt auf ein sekundäranalytisches Vorgehen setzt.¹

Über die Miteinreichung und Nachnutzung der verwendeten Forschungsdaten hinaus ist durch die Publikation von Forschungsdaten in Forschungsdatenzentren selbst ein wichtiger wissenschaftlicher Mehrwert entstanden. In Zusammenarbeit von Datenproduzenten, Forschungsdatenzentren und Förderorganisationen konstituiert sich derzeit eine Infrastruktur, die Nachhaltig-

1 Im LIfBi werden Forschenden Datensätze verschiedener Kohorten zum Bildungsverlauf über die gesamte Lebensspanne zur Nachnutzung bereitgestellt.

keit und neue innovative Vernetzungsmöglichkeiten in der wissenschaftlichen Gemeinschaft verspricht und darüber hinaus die interdisziplinäre Bearbeitung von Forschungsfragen auf der Grundlage unterschiedlicher Datenquellen ermöglicht. Durch die Zitation digital verfügbarer Forschungsdaten unter Nennung eines Digital Object Identifiers (DOI) wird zudem die Arbeit der Datenproduzenten gewürdigt, und die Daten selbst werden endlich als ein bedeutender wissenschaftlicher Output anerkannt (vgl. Kminek 2018).

Die Dokumentation, Sicherung, Weitergabe und Nachnutzung von Forschungsdaten – unter Wahrung datenschutzrechtlicher und urheberrechtlicher Vorgaben – wird zu einem Hilfsmittel der wissenschaftlichen Arbeit, das eine effiziente Ausnutzung wissenschaftlicher Ressourcen und eine Vernetzung der Scientific Community ermöglicht.

Gegenwärtig wird eine nationale Infrastruktur für Forschungsdaten der empirischen Bildungsforschung aufgebaut. Doch die Bereitschaft, Daten zu Zwecken der Nachnutzung zur Verfügung zu stellen, ist im Forschungsfeld bislang eher schwach ausgeprägt. In der Vergangenheit mag es aufseiten der Anbieter von Datendokumentation und Datensicherung ein deutliches Defizit gegeben haben, mittlerweile aber stabilisieren sich erste institutionalisierte Strukturen zur Unterstützung der Sicherung und Bereitstellung von Forschungsdaten. Der 2012 vom BMBF ins Leben gerufene Verbund Forschungsdaten Bildung (VerbundFDB) hat sich dieses Desiderats angenommen, um im nationalen Kontext Forschungsdaten der empirischen Bildungsforschung zentral sichtbar und zugänglich zu machen. Forschende finden hier Informationen zu datenschutz- und urheberrechtlichen Fragen und zum Forschungsdatenmanagement sowie Best-Practice-Beispiele zum Umgang mit und zur Dokumentation von Forschungsdaten.² Darüber hinaus stellt der VerbundFDB einen zentralen Einstieg für die Suche nach Forschungsdaten und Instrumenten aus Forschungsprojekten bereit (siehe dazu Abschnitt 4). Das ist ein wichtiger Schritt in Richtung der Öffnung der Daten für die Wissenschaft. Darüber hinaus können der interessierten Öffentlichkeit im Rahmen von Public-Use-Files Daten zur Verfügung gestellt werden.

Wer und zu welchem Zweck auf welche Daten zugreifen kann und darf, ist nicht zuletzt eine Frage der rechtlichen Voraussetzungen und der forschungsethischen Sensibilität der vorliegenden Daten. Der diesbezügliche Bewertungsprozess kann am Ende auch zu einem Verschluss der Daten oder zu deren Löschung führen. Aufseiten der datenaufnehmenden Institutionen müssen – in Zusammenarbeit mit der Fachcommunity – zudem Lösungen für die Einschätzung der Archivierungswürdigkeit der Daten entwickelt werden, um die erwarteten Datenmengen auch adäquat be- und verarbeiten zu können. In dieser Hinsicht ist eine deutliche Professionalisierung durch Harmo-

2 <https://www.forschungsdaten-bildung.de/publikationsreihen?la=de>

nisierungsprozesse zwischen den archivierenden Einrichtungen festzustellen. Die hierbei zu beantwortenden zentralen Fragen lauten: Welche Daten können und dürfen gesichert und zu welchem Zweck bereitgestellt werden? Wer darf Zugang dazu erhalten und wer warum nicht? Welche Daten müssen nicht gesichert werden? Welche Angaben sind notwendig, um die Daten retrospektiv auswerten zu können? Wie können Daten langfristig gesichert und in digitale Datenformate überführt und auf entsprechende Datenträger übertragen werden? All diese Fragen sind derzeit Gegenstand nationaler und internationaler Initiativen und Debatten.

Neben dem Stichwort „*open*“ muss im Zusammenhang mit Forschungsdaten auch der Begriff „*FAIR*“ diskutiert werden, der neben der Fairness des Zugangs zu den Daten (d. h. auffindbar und nutzerfreundlich zugänglich) auch auf eine internationale und interdisziplinäre Vernetzung von Forschungsdaten rekurriert. Wirklich offen können nur diejenigen Forschungsdaten sein, die forschungsethisch unbedenklich sind und die rechtlichen Vorgaben erfüllen.

2 FAIRe und offene Daten

Die Bereitstellung und Zugänglichmachung von Forschungsdaten beruht auf der Vernetzung bestehender Archive und Forschungsdatenzentren in föderierten Strukturen, die auf einer übergeordneten Ebene einen zentralen Einstieg ermöglichen. Hierfür stellt der VerbundFDB die nötige Infrastruktur bereit. Die FAIR-Prinzipien sehen es vor, Datenbestände und die dazugehörigen beschreibenden Metadaten „*Findable*“, „*Accessible*“, „*Interoperable*“ und „*Re-Usable*“ zu machen. Dazu bedarf es Anstrengungen auf verschiedenen Ebenen: Die am VerbundFDB beteiligten Forschungsdatenzentren³ und universitären Archive⁴ sprechen sich hinsichtlich solcher Themen wie der Vereinheitlichung von Metadaten, der Archivierungswürdigkeit von Forschungsdaten, der Fragen der Langzeitarchivierung und technischen Innovationen sowie der Harmonisierung ab und setzen die getroffenen Vereinbarungen gemeinsam um. Ziel ist es, ein möglichst umfassendes Angebot auf der Basis einer föderierten Infrastruktur der für die Bildungsforschung relevanten Daten bereitzustellen. Auch wenn die FAIR-Prinzipien bisher noch nicht im-

3 Kernpartner des VerbundFDB unter Koordination des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) sind dessen Forschungsdatenzentrum (FDZ) Bildung, das FDZ des Instituts für Qualitätsentwicklung im Bildungswesen (IQB) und das GESIS – Leibniz-Institut für Sozialwissenschaften. Weitere Partner des Verbunds sind das Deutsche Institut für Erwachsenenbildung (DIE), das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), das Leibniz-Institut für Bildungsverläufe (LifBI) sowie das FDZ des Bundesinstituts für Berufsbildung (BIBB).

4 „Qualiservice“ der Universität Bremen, „Archiv für pädagogische Kasuistik“ (ApaeK) der Goethe-Universität Frankfurt am Main und „Deutsches Gedächtnis“ der Fernuniversität Hagen.

plementiert sind und sich die Konzepte zu ihrer praktischen Umsetzung und die Messbarkeit von FAIRness noch in der Entwicklung befinden, zeigen diese Initiativen, dass sich in der Bildungsforschung eine deutliche Besserung der Lage abzeichnet. Eine weitere Initiative auf nationaler Ebene stellt der Aufbau einer nationalen Forschungsdateninfrastruktur (NFDI) dar. Dabei sollen Forschungsdaten über Disziplingrenzen hinweg für die Wissenschaft und die interessierte Öffentlichkeit auffindbar und verfügbar gemacht werden. In diesem Kontext ist der vom Bundesministerium für Bildung und Forschung geförderte Rat für Sozial- und Wirtschaftsdaten (RatSWD)⁵ als wichtiger Motor der strategischen Weiterentwicklung der Forschungsdateninfrastruktur zu nennen. Auf internationaler Ebene haben sich Deutschland, Frankreich und die Niederlande im Rahmen der GoFAIR-Initiative⁶ zusammengeschlossen, um Daten und Services nach den FAIR-Prinzipien auszurichten. Auch bei der Entwicklung und Implementierung einer „European Open Science Cloud (EOSC)“⁷ spielen diese Prinzipien eine entscheidende Rolle. Damit soll die Nachnutzbarkeit über Landesgrenzen und Disziplingrenzen hinweg europaweit sichergestellt werden. Eng damit verbunden ist die Harmonisierung der Metadaten, um sowohl die Systeme als auch die Nachnutzung technischer Lösungen interoperabel zu halten und damit Synergien bei der Entwicklung der Forschungsdateninfrastruktur zu erzielen.

3 Forschungsdaten in der empirischen Bildungsforschung

Auf Tagungen der an der Bildungsforschung beteiligten Disziplinen erfreut sich das Thema der Sicherung, Bereitstellung und Nachnutzung von Forschungsdaten immer größerer Bedeutung und Nachfrage, wie die Durchführung von Symposien und Workshops zu diesen Themen in jüngster Zeit belegt. Eng verbunden mit der Sicherung und Bereitstellung von Forschungsdaten ist auch der Umgang mit den Daten selbst, also das projektinterne Forschungsdatenmanagement, eine wichtige Voraussetzung für ein gelingendes Forschungsvorhaben nach den Regeln zur Sicherung der guten wissenschaftlichen Praxis⁸ sowie einer späteren Dokumentation in einem geeigneten Repositorium bzw. Forschungsdatenzentrum. Mittlerweile wird von den Förderorganisationen und Ethikkommissionen auch explizit ein Forschungsdatenmanagementplan verlangt und Interessenten die Möglichkeit gegeben, Ressourcen für die Forschungsdatendokumentation zu beantragen. Dabei soll neben der Replizierbarkeit der Befunde durch die Dokumentation und Bereit-

5 <https://www.ratswd.de>

6 <https://www.go-fair.org/>

7 <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

8 <https://doi.org/10.1002/9783527679188.oth1>

stellung der Daten über ein Forschungsdatenzentrum auch die Nachnutzung von Forschungsdaten stimuliert werden, sofern dies rechtlich und forschungsethisch möglich ist. Die datenschutzrechtlichen Vorgaben spielen vor allem bei personenbezogenen Daten eine wichtige Rolle. Ist eine Anonymisierung der Daten nicht möglich oder wegen des Verlustes analytischer Potenziale unsinnig, müssen die Möglichkeit der Datendokumentation und die möglichen Nachnutzungsszenarien besonders streng geprüft werden. Entscheidend für eine Datenweitergabe ist bei personenbezogenen Daten die Einwilligungserklärung, in welcher der Ort der Sicherung der Daten selbst sowie eine mögliche Datenweitergabe und die damit verbundenen Auswertungszwecke genannt sind. Da diese Vorgaben bei Erhebungen in der Vergangenheit oftmals nicht berücksichtigt wurden, ist die Nachnutzung von Datenbeständen, die personenbezogene Daten enthalten, besonders zu prüfen und einzuschränken beziehungsweise bei besonders sensiblen Daten sogar ganz auszuschließen.

4 Zugang zu Forschungsdaten der empirischen Bildungsforschung in Deutschland

Einen zentralen Einstieg zu Forschungsdaten der empirischen Bildungsforschung bietet der VerbundFDB an. Hier können nicht nur Forschungsdaten „abgegeben“, sondern auch Datenbestände und erprobte Forschungsinstrumente aufgefunden werden. Einen zentralen Bestand stellen die Projekte des vom BMBF geförderten „Rahmenprogramms für Empirische Bildungsforschung“⁹ dar. Die Daten des Rahmenprogramms wurden gemäß den Archivierungsschwerpunkten an die beteiligten Forschungsdatenzentren FDZ am IQB, FDZ Bildung des DIPF und GESIS zur Sicherung und Dokumentation verteilt. Darüber hinaus weisen diese Kernpartner des Verbundes einen Großteil ihrer für die Bildungsforschung relevanten Daten über forschungsdatenbildung.de nach. GESIS und IQB haben sich auf die Dokumentation quantitativer Datensätze spezialisiert, während sich das FDZ Bildung am DIPF zu einer zentralen Anlaufstelle für die Sicherung und Bereitstellung von qualitativen Forschungsdaten entwickelt. Daneben bieten GESIS mit seiner „Zusammenstellung sozialwissenschaftlicher Items und Skalen“ (ZIS)¹⁰ und das FDZ Bildung über sein Angebot „Datenbank zur Qualität von Schule“ (DaQS)¹¹ erprobte Fragebogen- und Testinstrumente für die kostenfreie Nachnutzung an. Damit sollen die Verfügbarkeit von Forschungsinstrumenten und der Austausch darüber erhöht und ihre Nachnutzung ermöglicht werden. Denn auch Forschungsinstrumente und ihre Weiterentwicklung sind wich-

9 <https://www.empirische-bildungsforschung-bmbf.de/>

10 <https://zis.gesis.org/>

11 <http://daqs.fachportal-paedagogik.de/>

tige Forschungsleistungen, welche die Effizienz und Leistungsfähigkeit der Bildungsforschung signifikant zu erhöhen versprechen.

Entscheidend sind die fortschreitende Vernetzung der verteilten Archivierungsstrukturen und die zentrale Auffindbarkeit ihrer Bestände über den VerbundFDB. Diese föderierte Architektur der Forschungsdatenlandschaft verspricht eine hohe Aufnahme- und Anschlussfähigkeit für neue Partner sowie eine maximale Flexibilität bei der Gewährung des Datenzugangs durch die datenvorhaltenden Institutionen. Die Zugangsmöglichkeiten variieren zwischen den Forschungsdatenzentren und auch den vorgehaltenen Datenkollektionen selbst zum Teil erheblich. Dabei verfügen die einzelnen Forschungsdatenzentren über eine höchst heterogene Ressourcenausstattung. Nicht alle Partner können aufwendige Remote-Desktop-Lösungen oder One-Site-Gastarbeitsplätze in gleichem Maße realisieren. Dennoch erlauben diese technischen Innovationen neue, abgestufte Wege der Zugänglichmachung. Durch diese granulいたe Vorgehensweise beim Datenzugang lassen sich unterschiedliche Interessen bedienen, beispielsweise Embargofristen für die Datenweitergabe oder die Bereitstellung von Scientific- bzw. Public-Use-Files. Insbesondere bei quantitativen Daten ist die Anonymisierung ein probates Mittel, um Forschungsdaten für die Nachnutzung zu öffnen.

Komplizierter gestaltet sich dies bei qualitativen Daten, die schon aufgrund geringerer Fallzahlen oder der Befragung leicht eingrenzbarer Personengruppen häufig ein erhöhtes Reidentifikationsrisiko enthalten, was eine Anonymisierung deutlich erschwert. Darüber hinaus werden bei qualitativen Erhebungen oftmals sehr sensible und sogar intime Daten gesammelt, bei denen ein hoher Vertrauensvorschuss der Probanden gegenüber der Forschung besteht, besonders auch dann, wenn beispielsweise in einem politisch umstrittenen Feld geforscht werden soll. Dies zeigt die große Bedeutung von forschungsethischen Standards auch in der qualitativen Sozialforschung (vgl. Unger 2014). Generell bieten qualitative Forschungsdaten aufgrund ihrer Reichhaltigkeit eine Vielzahl von Anknüpfungspunkten für Sekundäranalysen, besonders dann, wenn sie gut dokumentiert sind und alle notwendigen Hintergrundinformationen vorgehalten werden. Der Zugriff auf qualitative Forschungsdaten wird in der Praxis deutlich restriktiver gehandhabt. Diese Daten werden teilweise in nicht-anonymisierter oder nur teil-anonymisierter Form für Forschungszwecke zugänglich gemacht. Auch wenn die Anonymisierung sehr aufwendig und kostspielig ist, ist sie eine gute Möglichkeit, um auf der Basis von anonymisierten Transkriptionen von Audio- und Videoaufzeichnungen auch qualitative Datenbestände für Forschung und Lehre zu öffnen.

5 Fazit

Die Unsicherheit hinsichtlich juristischer Fragen stellt sowohl für die Datengeber als auch für das aufnehmende Forschungsdatenzentrum häufig ein entscheidendes Hindernis bei der Weitergabe von Forschungsdaten an ein Forschungsdatenzentrum dar. Die Datenweitergabe außerhalb eines Forschungsdatenzentrums erfolgt dann meist in einer Grauzone informeller wissenschaftlicher Kooperationen und Netzwerke. Um dieser Praxis entgegenzuwirken, ist gerade die Sicherung und Weitergabe von Daten durch ein Forschungsdatenzentrum ein guter Weg, um das Vertrauen zwischen den Probanden und den Datenproduzenten sowie diesen und den Forschungsdatenzentren nachhaltig zu stärken. Die Offenheit von Forschungsdaten ist, wie dargestellt, nicht nur eine forschungsethische, sondern auch eine juristische Angelegenheit sowie eine Frage des Datenmanagements im Vorlauf und im Nachgang der Erhebung. Um einen fairen Zugang zu den Daten zu fördern, ist ein geregeltes und transparentes Verfahren des Zugriffs auf Forschungsdaten unverzichtbar. Forschungsbefunde zum Nachnutzungsverhalten von Forschenden zeigen, dass die wahrgenommene Nützlichkeit (Qualität der Daten, Passung der Daten zu den Forschungsfragen sowie Kosten und Aufwand des Datenzugriffs) ein entscheidender Faktor für die Entscheidung einer Sekundärnutzung von Daten ist (Kim/Yoon 2017). Darüber hinaus spielen Möglichkeiten zur Beantragung zusätzlicher Ressourcen und die Unterstützung sowie die Verfügbarkeit einer Dateninfrastruktur in Form von Forschungsdatenzentren eine wichtige Rolle bei der Entscheidung für eine Datennachnutzung (vgl. Fear/Donaldson 2012; Yakel et al. 2013). Nachdem diese Vorgaben aus der Forschung nun von den Infrastrukturangeboten umgesetzt und implementiert worden sind, setzt langsam ein Umdenken gegenüber Sekundäranalysen in der empirischen Bildungsforschung ein. Dies hängt nicht nur mit der gestiegenen Verfügbarkeit und Zugänglichkeit von Forschungsdaten in diesem Feld zusammen, sondern ist auch eine Folge der praktizierten Sensibilität im Hinblick auf die juristischen und forschungsethischen Fragen. Es liegt nun an der wissenschaftlichen Gemeinschaft, die notwendigen Mittel für das Datenmanagement zu beantragen sowie die Angebote in Zusammenarbeit mit den Forschungsdatenzentren weiterzuentwickeln und letztendlich auch zu nutzen, um Forschungsdaten für ihre Zwecke weiter zu öffnen.

Marius Gerecht, Dr., ist Wissenschaftlicher Mitarbeiter am DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Abteilung: Informationszentrum Bildung, Arbeitsbereich: Forschungsdaten Bildung.

Helge Kminek, Dr., ist Wissenschaftlicher Mitarbeiter an der Johann Wolfgang Goethe-Universität Frankfurt am Main, Institut für Pädagogik der Sekundarstufe, Administrator ApaeK.

Literatur

- Fear, Kathleen/Donaldson, Devan Ray (2012): Provenance and credibility in scientific data repositories. In: *Archival Science* 12, 3, S. 319-339. <https://doi.org/10.1007/s10502-012-9172-7>.
- Kim, Youngseek/Yoon, Ayoung (2017): Scientists' data reuse behaviors: A multi-level analysis. In: *Journal of the Association for Information Science and Technology* 68, 12, S. 2709-2719. <https://doi.org/10.1002/asi.23892>.
- Kminek, Helge (2018): Möglichkeiten und Grenzen der Archivierung und Nachnutzung qualitativer Forschungsdaten aus Sicht der Objektiven Hermeneutik. In: Richter, C./Mojescik, K. (Hrsg.): *Vom Geben und Nehmen. Die Praxis der Aufbereitung und sekundäranalytischen Nutzung von qualitativen Daten in der Sozialwissenschaft und ihren Nachbardisziplinen*. Wiesbaden: Springer VS. (im Erscheinen).
- Unger, Hella von (2014): Forschungsethik in der qualitativen Forschung: Grundsätze, Debatten und offene Fragen. In: Unger, H. v./Narimani, P./M'Bayo, R. (Hrsg.): *Forschungsethik in der qualitativen Forschung – Reflexivität, Perspektiven, Positionen*. Wiesbaden: Springer VS 2014, S. 15-40. https://doi.org/10.1007/978-3-658-04289-9_2.
- Yakel, Elizabeth/Faniel, Ixchel M./Kriesberg, Adam/Yoon, Ayoung (2003): Trust in digital repositories. *International Journal of Digital Curation*, 8, 1, S. 143-156. <https://doi.org/10.2218/ijdc.v8i1.251>.