

THEMENSCHWERPUNKT

„NEUE FORMEN DER ARCHIVIERUNG, BEREITSTELLUNG UND NACHNUTZUNG VON FORSCHUNGSDATEN“

Warum sich Forschende um Archive, Zugänge und die Nutzung bildungswissenschaftlicher Forschungsdaten kümmern sollten – historische und informationswissenschaftliche Perspektiven

Sabine Reh, Lars Müller, Stefan Cramme, Bettina Reimers & Marcelo Caruso

Die Fachgesellschaften in der Bildungsforschung – Deutsche Gesellschaft für Erziehungswissenschaft (DGfE), Gesellschaft für Empirische Bildungsforschung (GEBF) und Gesellschaft für Fachdidaktik (GFD) – plädieren mit einer wichtigen Stellungnahme „Empfehlungen zur Archivierung, Bereitstellung und Nachnutzung im Kontext erziehungs- und bildungswissenschaftlicher sowie fachdidaktischer Forschung“ für den reflektierten Umgang mit Forschungsdaten und stellen in ihrem Papier nützliche Hinweise für das Datenmanagement in der empirischen Bildungsforschung zur Verfügung (DGfE/GEBF/GFD 2020). Im Sinne einer „verantwortlichen“ und „effizienten Nutzung“ von Ressourcen halten sie es für wichtig, dass Forschungsdaten so archiviert und bereitgestellt werden, dass in empirischen Studien erzielte Ergebnisse nicht nur überprüft, sondern die erzeugten Daten auch von anderen Wissenschaftlerinnen und Wissenschaftlern für ähnliche oder andere Forschungsfragen nachgenutzt werden können.

So begrüßenswert das Papier der Fachgesellschaften und die abwägende Stellungnahme ist und so sehr die pragmatischen Hinweise zum Datenmanagement hilfreich erscheinen, so deutlich stellt es doch nicht mehr als den Beginn der Diskussion um „Forschungsdaten“ in den Erziehungs- und Bildungswissenschaften dar – und das aus zwei Gründen: Zum einen wird das Thema der erziehungs- und bildungswissenschaftlichen Forschungsdaten auf Daten aus der empirischen Bildungsforschung beschränkt. Aus Sicht der historischen Bildungsforschung und mit Blick auf die Breite der bildungswissenschaftlichen Disziplinen scheint es sinnvoll, das Thema „Forschungsdaten“ weiter zu fassen und multidisziplinär anzugehen. Nicht nur in der empirischen Bildungsforschung werden „Forschungsdaten“ erzeugt, ist gefordert, solche zu archi-

vieren, und ist es sinnvoll, sie weiter zu nutzen. So können beispielsweise auch Bildungshistorikerinnen und Bildungshistoriker Daten aus der empirisch-sozialwissenschaftlichen Forschung nachnutzen. Zum anderen ist das Thema nicht ausschließlich als eines des Forschungsdatenmanagements zu behandeln.¹ Forschungsdaten sind keinesfalls nur eine Angelegenheit einiger weniger Spezialistinnen und Spezialisten, die den Forschenden durch die Forschungsförderungsbürokratie auferlegt wird, und es handelt sich schon gar nicht um eine Debatte zwischen denjenigen, die empirisch-„quantitativ“ und solchen, die empirisch-„qualitativ“ forschen. Forschungsdaten tangieren alle sozial- und kulturwissenschaftlichen wie auch die geisteswissenschaftlichen Disziplinen mit ihren Gegenständen, Methoden und Ergebnissen. Dabei geht es nicht um rein administrative Techniken, sondern um die fachliche Gestaltung der Frage, wie Wissen über Bildung erzeugt werden soll. Als Bildungshistorikerinnen und Bildungshistoriker sowie Angehörige einer bildungshistorischen Gedächtnisinstitution wollen wir in diesem Sinne den Blick auf das Thema erweitern.

Forschungsdaten werden oft mit viel Ressourcen, mit öffentlichen Geldern und persönlichem Aufwand der Untersuchenden und der Untersuchten produziert und sind doch oft nicht in all ihren Potenzialen genutzt. Einmal erzeugte Daten sollten daher auch für die Beantwortung neuer, bisher nicht gestellter Fragen zur Verfügung stehen und in diachronen Vergleichen möglicherweise Veränderungen der erforschten Wirklichkeit aufzeigen können. Die Fachgesellschaften empfehlen, auch das Erheben und Bereitstellen, d. h. das Veröffentlichenden von Daten als wissenschaftliche Leistung der Autorinnen und Autoren ebenso anzuerkennen wie das „nachnutzende“ Forschen mit schon erhobenen Daten und für das sinnvolle Archivieren und Bereitstellen von Daten entsprechende (auch finanzielle) Voraussetzungen im Wissenschaftsbetrieb und der Forschungsförderung zu schaffen.

Mit der Stellungnahme sind DGfE, GEBF und GFD der Aufforderung der Deutschen Forschungsgemeinschaft (DFG) an die Fachgesellschaften nachgekommen, Empfehlungen zu verfassen, in denen die Spezifika des Umganges mit Daten in den jeweils vertretenen Disziplinen Berücksichtigung finden. Als Spezifika der empirischen Bildungsforschung bestimmen die Fachgesellschaften einen in besonderer Weise Probleme aufwerfenden Forschungsgegenstand und die damit verbundene Erzeugung häufig sensibler Daten: Die zu erforschenden Bildungs- und Erziehungsprozesse seien „vulnerable Prozesse“; erforscht würden besonders schützenswerte Gruppen, zumeist Kinder und Jugendliche, die sich oft in staatlich beaufsichtigten, hierarchisch organisierten – man könnte möglicherweise sagen: machtvoll strukturierten – und für sie konsequenzenreichen Situationen befinden. Daher seien – allerdings ist das selbstverständlich und gilt auch für Daten in anderen Forschungsfeldern – bei der

1 So die Kritik des Rates für Informationsinfrastrukturen (RfII) in seinem Papier über Datenqualität, der ebenfalls dafür plädiert, die Fragen der Datenqualität als grundlegende methodische Fragen wissenschaftlichen Vorgehens zu diskutieren, RfII 2019, z. B. S. 7.

Datenerhebung datenschutzrechtliche Belange, gesetzliche Vorgaben und forschungsethische Fragen ebenso zu berücksichtigen wie hinsichtlich der Archivierung und der Bereitstellung von Daten die Regeln guter wissenschaftlicher Praxis. Gerade aufgrund der in der Bildungsforschung entstehenden sensiblen Daten müsse es Aufgabe und Angelegenheit jedes einzelnen Forschungsprojektes sein, Möglichkeiten und Bedingungen der Archivierung und Bereitstellung zu prüfen und – im Rahmen des Forschungsdatenmanagementplans – zu dokumentieren. Entsprechend dürfe zudem die Bewilligung von Fördergeldern nicht davon abhängig gemacht werden, ob in einem Projekt nachnutzbare Daten entstehen oder nicht.

Es scheint notwendig, die Fragen, die sich im Zusammenhang mit bildungswissenschaftlichen „Forschungsdaten“ stellen, weiter zu fassen und gleichzeitig präziser zu formulieren: Warum und in welcher Weise sind „Forschungsdaten“ eine Herausforderung für die Wissenschaft – und gerade auch für eine gesellschaftlich verantwortliche Wissenschaft? Und wie sind Wege zum Umgang mit Forschungsdaten in den beteiligten Communitys zu finden bzw. wie ist die Diskussion darüber zu organisieren? Archivare und Archivarinnen wie auch Bibliothekare und Bibliothekarinnen sind diejenigen, die sich professionell mit Fragen der Aufbewahrung und der Überlieferung befassen. Am Ende sind es die Historiker und Historikerinnen, die – um jetzt die historische Terminologie zu nutzen – anhand von „Dokumenten“, von „Überresten“ und sich auf diese beziehend forschen, für die diese dann eben – wiederum in der Terminologie der Historikerinnen und Historiker – zu „Quellen“ werden. Auch wenn man aus der Arbeit von Bibliothekarinnen und Bibliothekaren sowie Archivarinnen und Archivaren lernen kann, dass es keine ein für alle Mal feststehenden, allgemeingültigen Regeln für die Aufbewahrung und Sammlung, für Zugänglichkeit und Nutzung von zu Quellen werdenden „Dokumenten“ und „Überresten“ gibt, existieren aber Kriterien solcher „Überlieferungsarbeit“. Es gilt, diese Kriterien im Hinblick auf die Produktion, Speicherung und Nutzung von Forschungsdaten immer wieder zu reflektieren, zu diskutieren und Antworten auf folgende grundlegende Fragen zu finden:

- Was sollte aufbewahrt werden? Müssen alle einmal gespeicherten Forschungsdaten aufbewahrt werden – und für wie lange? Und wenn nicht alle, was genau soll dann eigentlich aufgehoben werden? Was könnten Auswahlgesichtspunkte sein?
- In welcher Weise sollen die Forschungsdaten aufgehoben werden? Welche Informationen sind nötig, um gespeicherte und aufbewahrte Daten auch außerhalb des ersten Kontextes noch verstehen zu können?
- Wer bestimmt darüber, was und wie aufbewahrt wird? Jede bzw. jeder einzelne Forschende selbst?
- Welche Ressourcen – personell und technisch – sind nötig, um Forschungsdaten in einer nutzbaren Weise zu sichern und für alle zugänglich zu machen? Und sollen sie wirklich von Beginn an für alle zugänglich sein?

Wir werden in drei Thesen knapp erläutern, warum wir dafür plädieren, dass in Zukunft über Forschungsdaten von der Community laufend reflektiert, diskutiert und entschieden werden muss:

1. Die Frage nach Forschungsdaten ist nicht nur deshalb wichtig, weil Daten zu einem – auch ökonomischen – ‚Rohstoff‘ in der digitalen Gesellschaft geworden sind, sondern weil es um die Rolle der Wissenschaft und die Form rationaler Entscheidungen in demokratischen Gesellschaften geht und möglicherweise darum, einem Vertrauensverlust der Wissenschaft entgegenzuarbeiten.
2. Da Daten niemals einfach gegeben sind, handelt es sich in Bezug auf Forschungsdaten nicht ausschließlich um Probleme des Managements, sondern um methodologische und methodische Forschungsfragen.
3. Nachnutzung muss – durch ein angepasstes Forschungsdatenmanagement – möglich und multidisziplinär angelegt sein.

Zu 1.

Die große Bedeutung der Produktion und Verteilung von Wissen und eine wachsende Rolle von Wissenschaft und wissenschaftlichen Erkenntnissen sind Merkmale einer schon in den 1960er Jahren diagnostizierten „Wissensgesellschaft“, deren Geschichte allerdings bis weit ins 19. Jahrhundert zurückreicht, wo schon mit öffentlicher Sozialstatistik „Daten“ als Mittel der Selbstbeschreibung von Gesellschaft produziert wurden (Nassehi 2019, S. 31-35, 57-63). Die Repräsentation der Welt in Daten, die in methodisch kontrollierten Verfahren erzeugt wird, führt – wie Nassehi (2019, S. 69) schreibt – zu einem Eigenwert von Daten. Die Dynamik dieser Wissensgesellschaft hat mit methodischen Entwicklungen, mit steigenden Möglichkeiten der Formalisierung des Denkens (vgl. Krämer 1984) und mit technischen Entwicklungen zu einer „Datafizierung“, zu maschinenlesbaren Codes, automatischer Mustererkennung und -analyse geführt (Bächle 2016, S. 116-117) und schließlich mit der Einführung des Internets ungeheuer zugenommen (z. B. Stalder 2016, S. 24-39). Das zunehmende Wissen und die steigende Zahl gesellschaftlicher, wirtschaftlicher und militärischer Abläufe, in denen Wissen genutzt, verarbeitet und auch neues produziert wird, lässt sich nur mithilfe von Informationstechnologien verwalten (Stalder 2016, S. 70). Dabei werden immer mehr Daten erzeugt; sie entstehen keinesfalls mehr nur – wohl heute nicht einmal mehr hauptsächlich – in wissenschaftlichen Kontexten, sondern in vielfältigen „szientoiden“ Praktiken (vgl. Nassehi 2019, S. 68-69). Aber wo auch immer produziert: „Daten werden immer wichtiger – als Zahlungsbarriere für Plattformen, als Zahlungsmittel und als Innovationstreiber“, formuliert es EU-Kommissarin Margarethe Vestager im Hinblick auf Versuche

der EU, die global agierenden „Tech-Giganten, die das Internet weitgehend organisieren“, zu regulieren.²

Steigt also einerseits die Bedeutung von Wissenschaft, die auf der Grundlage von Daten Wissen zur politischen Steuerung bereitstellt, so ist sie gleichzeitig als Wissenschaft umstrittener denn je. Dass Datenherstellung eine kaum noch überschaubare Produktion ist, ist offensichtlich geworden. Nicht zuletzt die Auseinandersetzungen um wissenschaftliche Erkenntnisse, auf die sich Wissenschaftlerinnen und Wissenschaftler sowie Politikerinnen und Politiker bezogen, um Entscheidungen im Kontext der Pandemie zu legitimieren, hat das gezeigt. Der Bedeutsamkeit wissenschaftlichen Wissens steht ein Vertrauensverlust der Wissenschaft in der breiten Öffentlichkeit gegenüber: „Wissenschafts- und Aufklärungsfeindlichkeit, so zeigt die Coronakrise besonders deutlich, sind ein charakteristisches Merkmal von autoritärem Populismus. Gefühlte Wahrheiten, Fake News und verschwörungstheoretisch geprägtes Misstrauen gegenüber all jenen Positionen, die nicht mit der eigenen Meinung übereinstimmen, ersetzen eine an rationalen Kriterien orientierte Konfliktaushandlung. Wissenschaftsskepsis und naiver Wunderglaube liegen dabei oft nah beieinander“ (Tiedemann/Huke 2020, S. 21). Gefragt wird inzwischen von vielen: Ist die Studie, auf die sich gerade jemand bezieht, überhaupt eine qualitätsvolle Studie? Sind die Aussagen, die aufgrund ihrer Ergebnisse gemacht werden, richtig und wahr? Was sind die Grundlagen dieser Aussagen, was die Daten, die jeweils eine Aussage in welcher Weise begründen können? Wem kann man eigentlich noch trauen? Und wer kann das beurteilen?

Aber nicht nur in der breiten Öffentlichkeit wird – berechtigt oder nicht – Skepsis laut, auch Wissenschaftlerinnen und Wissenschaftler selbst sehen Probleme. Während sich die einen aber einer „Replikationskrise“ gegenübersehen,³ streiten andere die Möglichkeit und Notwendigkeit der Replikation von Ergebnissen überhaupt ab und bleiben eine systematische Antwort auf die Frage nach der Nachvollziehbarkeit berichteter Ergebnisse oft schuldig. Ob es in dieser Lage hilft bzw. welche Konsequenzen es gesellschaftlich betrachtet hat, sich darauf zurückzuziehen, dass man sich in einem vorsichtig zu begehenden Forschungsfeld bewegt und hier „sensible“ Daten generiert, die nur unter schwierigsten Bedingungen nachgenutzt werden können, muss diskutiert werden.

Vor dem Hintergrund dieser gesellschaftlichen Entwicklungen, der Tatsache, dass Wissen in Daten „Rohstoff“ (Bächle 2016, S. 113-114) der Politik so gut wie der Ökonomie ist, gibt es aus unserer Sicht keine Alternative zur Ra-

2 So äußerte sich Vestager in einem Zeitungsinterview zu den Möglichkeiten politischer und juristischer Regulierung der großen Player in diesem Feld (Meier/Müller von Blumencron/Schröder 2020).

3 Vgl. etwa die Auseinandersetzung von Gerald Wagner unter dem Titel „Gibt es die Replikationskrise? Wissenschaftliche Irrtümer in Serie“ mit der These, von einem amerikanischen Mathematiker im „Merkur“ vertreten, dass es eine „Replikationskrise“ gebe, FAZ vom 12. April 2020, F.A.Z.net.

tionalität eines in Forschungsdaten erfassten und belegbaren Wissens. Das erfordert wiederum, ein neues Verständnis des Verhältnisses von Forschung, den Institutionen der Forschung und Infrastrukturanbietern zu generieren. Aus unserer Sicht ist es daher an der Zeit, dass alle Bildungswissenschaftlerinnen, Bildungswissenschaftler und die Verantwortlichen in den Infrastruktureinrichtungen diese Herausforderung annehmen und die mit „Forschungsdaten“ zusammenhängenden Fragen gemeinsam diskutieren – sie gehen uns alle an. Eine anachronistische und schließlich auch in jeder Hinsicht – forschungs- und gesellschaftspolitisch – problematische Position scheint es zu sein, sich gegen die Überlieferung von Forschungsdaten, auch von „sensiblen“, auszusprechen. In dieser Hinsicht ist die Stellungnahme unbedingt zu begrüßen und die Diskussion darüber fortzusetzen.

Zu 2.

In der gemeinsamen Stellungnahme gehen die Fachgesellschaften von einem sehr weiten Verständnis von Forschungsdaten aus. Forschungsdaten sind demzufolge Daten, deren charakteristisches Merkmal es ist, „im Zuge wissenschaftlicher Vorhaben“ entstanden zu sein. Was „Daten“ – im Unterschied etwa zu Informationen – sind, wird nicht vorweg geklärt, hängt demzufolge weder ab von einem Trägermedium noch von dessen materialer und (technischer) Verfasstheit, sondern ist ausschließlich bestimmt als dasjenige, was in Forschung entsteht und in irgendeiner Weise als Ausgangspunkt, Untersuchungsgegenstand oder (Zwischen-)Ergebnis entsteht, genutzt und weiterverarbeitet wird – also einem Prozess unterzogen wird, in dem wiederum neue Daten anfallen.

Forschungsdaten werden in der Stellungnahme entsprechend ihrem Auftreten im Prozess des Forschens in *unbearbeitete Rohdaten*, *aufbereitete Daten* und *Kontextdaten/Kontextinformationen* unterschieden. Ein solcher – man könnte sagen – funktionaler oder pragmatischer Ansatz, um über Forschungsdaten zu sprechen, scheint vielversprechend. Er erlaubt es, die Arbeitsweisen vieler Disziplinen zu berücksichtigen und sowohl das, was sie als „Gegenstände“ ihrer Forschung verstehen, wie auch das, was sie als Aussagen über diese Gegenstände fabrizieren, als Forschungsdaten zu bezeichnen. Hinsichtlich der Nutzung der Begriffe Daten, Medien, Quellen und Information ergibt sich daraus allerdings – und das kann man auch in der Stellungnahme erkennen – oft ein munteres Durcheinander, das illustriert, wie knifflig die Materie ist, wenn man auf das gesamte Forschungsfeld der Erziehungs- und Bildungswissenschaften schaut.

Zur ersten Gruppe der Daten, in der Stellungnahme „Rohdaten“ genannt, werden auch solche Dinge gezählt, die – in der Terminologie der Historikerinnen und Historiker – typischerweise Grundlage der bildungshistorischen Forschungsarbeit werden. Als Beispiel werden „Artefakte und Dokumente“ ge-

nannt, also Briefe oder andere Schriftstücke, aber auch Fotos und Zeichnungen. In welcher Weise aber sind dies im Zuge der Forschungsarbeit entstandene Daten? Im Falle historischer, teilweise auch kulturwissenschaftlicher Forschung handelt es sich im engeren Sinne oft um gesuchte, aufgefundene und gesammelte „Daten“. Inwiefern lässt sich dabei von „Rohdaten“ sprechen?

Die Konnotationen, die mit dem Begriff des „Rohen“ oder der Rohheit der Daten verbunden sind („unmittelbar aus der Erhebung angefallene Originaldaten, die noch in keinerlei Weise bereinigt, ausgewertet oder anderweitig nachbearbeitet wurden“, DGfE/GEbF/GFD, 2020, S. 5), erzeugen möglicherweise allerdings falsche Vorstellungen. Dieses Problem ist von Fachseite schon mehrfach diskutiert worden. Anders als das Wort „datum“ nahe legen könnte – übersetzen kann man es als „gegeben“ – sind Daten eben „keine universellen Entitäten, die darauf warten, entdeckt, aufgezeichnet und ausgewertet zu werden“ (kritisch: Bächle 2016, S. 122). Auswahlentscheidungen für Forschungsgegenstände, Datengranularität, Formate, Datenmodelle etc. gehen selbstverständlich jeder Erhebung eines Datums voraus. Forschungsdaten fallen deshalb nicht einfach „nur so“ an. Sie sind immer das Resultat von Entscheidungen der Forschenden, möglicherweise von Vorgaben der (technischen) Instrumente, mit denen Erhebungen vorgenommen werden, für die sich Forschende ebenfalls entschieden haben. Die von Historikerinnen und Historikern aufgefundenen „Rohdaten“ stehen zudem in einem Überlieferungszusammenhang, d. h. als archivierte Dokumente unterliegen sie den „Selektions-, Bewertungs-, Indexierungs- und Ordnungslogiken, welche archivalischen Arbeitsprozessen zugrunde liegen.“ (DGfE/GEbF/GFD 2020, S. 5.). Diese für die spätere Forschung wichtige Rolle von Archiven (und Bibliotheken) bei der Bereitstellung der Quellen wird auch in den historischen Disziplinen nicht immer ausreichend reflektiert. Die Überlieferung wird ebenfalls als „gegeben“ betrachtet, was sie streng genommen nur für den späteren oder den von außen darauf gerichteten Blick sind. Das dafür notwendige Vertrauen in die Bibliotheken und Archive basiert auf deren kontinuierlicher Praxis des systematischen Sammelns, Erschließens und – teilweise schon über Jahrhunderte währenden – Bewahrens von Dokumenten.

Sind nun Artefakte wie Schriftstücke und Fotos oder Zeichnungen von Historikern und Historikerinnen, die dabei zu Quellen werden, nach einer erfolgten Digitalisierung (wiederum gibt es an dieser Stelle Auswahlentscheidung, Formate, Datenmodelle, Volltexterzeugung) „aufbereitete Daten“? Die Erzeugung digitaler Abbilder von Büchern und Akten hat durch die dabei gewählten Datenmodelle, Formate und Volltexte großen Einfluss darauf, was mit erzeugten (Forschungs-)Daten gemacht werden kann. Diese Digitalisate können, in die Forschung eingeflossen, den Charakter von Forschungsdaten erhalten, die von anderen, also nicht dem/der Forschenden selbst, erzeugt wurden (Müller 2019, S. 196ff.). Damit wird hier die bewährte Arbeitsteilung von bewahrender Institution und nutzender Forschung durch die Digitalisierung ganz grundlegend hinterfragt. Und auch „Forschungsdatenzentren“ (FDZ), die selbst keine

„eigenen“ Daten bereitstellen, prägen mit ihrem Sammlungsprofilen, Sammlungszusammenhängen und Metadatenprofilen die archivierten Forschungsdaten erheblich (vgl. ebd.).

Informationen über die Entscheidungen, die zur Erhebung oder zur Sammlung der „Rohdaten“ getroffen wurden, „Kontextdaten/-information“, wie es in der Stellungnahme heißt, manche nennen Teile dieser Daten auch „Metadaten“ und „Paradaten“, werden von Historikerinnen und Historikern für ihren forschenden Umgang mit den „Rohdaten“, den Dokumenten und Überresten, wo immer es möglich ist, mit herangezogen und ebenso berücksichtigt wie der jeweilige Sammlungszusammenhang bei der Interpretation von Quellen – das ist ein Teil dessen, was in der historischen Forschungsarbeit Quellenkritik genannt wird. Werden in historischen Forschungen z. B. Forschungsdaten der empirischen Sozial- und Bildungsforschung analysiert, so fließen sie in das untersuchte Korpus – unabhängig davon, ob sie physisch oder digital vorliegen – als „Quellen“ der Forschung ein und werden in ihrem historischen Gesamtzusammenhang analysiert. In dieser Sichtweise haben Überreste und Dokumente, die in der historischen Forschung, also ihrer spezifischen Nutzung zu Quellen werden, immer einen „doppelten Verweischarakter: einmal auf das in ihnen zum Ausdruck Kommende, [man könnte sagen: den propositionalen Gehalt, d. V.] zum anderen auf den Entstehungs- und Produktionskontext.“ (Fickers 2020, S. 160). Bei der digitalen Aufbereitung der Quellen und deren computergestützten Auswertung (automatische Themenerkennung, statistische Inhaltsanalysen, Netzwerkvisualisierungen etc.) entstehen zudem weitere, spezifische Forschungsdaten, die sich in Bezug auf Fragen des Forschungsdatenmanagements nicht wesentlich von denen der empirischen Forschung unterscheiden.

In den „Praktischen Empfehlungen“ der Stellungnahme erhalten die Datenzentren und -archive eine passive Rolle zugewiesen (DGfE/GEBF/GFD 2020, S. 8), als geeignete Ablageorte für Forschungsdaten, die nicht weiter reflektiert wird. Genau dies aber müsste geschehen, denn „[d]ie Qualität von Forschungsdaten wird auch vom *state of the art* der Forschungsinfrastrukturen bestimmt, mit deren Hilfe (namentlich digitale) Daten analysiert, gespeichert und geteilt werden.“ (RfII 2019, S. 82). Genauso wenig wie das historische Archiv ist das Forschungsdatenarchiv allerdings ein passives Depot. Es ist ein Akteur im Forschungsdatenzyklus (vgl. Müller 2019, S. 199f.). Diese aktive Rolle muss reflektiert, ausgefüllt und kommuniziert werden.

Vor dem Hintergrund des breiten Datenbegriffs der Stellungnahme und im Brennglas der historischen Bildungsforschung, ihres – so möchten wir sagen – erprobten Umganges mit Gegebenem als Produziertem und ihrem Bewusstsein dafür, dass die Funktion im Forschungsprozess und die Art des Umganges etwas erst zu etwas Bestimmtem macht, Daten entstehen lässt, wird sehr schnell deutlich, was geschehen muss, damit Wissenschaft und das von ihr produzierte Wissen glaubwürdig bleibt. So wird in der Forschungsarbeit des Historikers oder der Historikerin erst der Überrest zur Quelle, und Literatur, die als For-

schungsliteratur genutzt wird, ist nicht „an sich“ etwas anderes, sondern wird erst in ihrem Gebrauch zu etwas anderem, das dann entsprechend auch in einem Extra-Verzeichnis und getrennt von dem aufgeführt wird, was als Quelle genutzt wurde. In gemeinsamen und sensiblen Debatten muss geklärt und festgelegt werden, dass und wie genau entsprechende Informationen zur Rekonstruktion dieser Prozesse bei Entstehung und Überlieferung von Daten mitgeliefert werden müssen. Im Bewusstsein muss bleiben, dass sich über größere Zeiträume dabei auch ändern kann, welche Metadaten und Paradata für grundlegend gehalten werden. Diese Festlegungen finden also nicht einmal statt, sondern es müssen laufend darüber fachintern und vielleicht auch in abgestimmten Verfahren Debatten geführt werden. Diese sind in dem Sinne, wie hier Daten definiert wurden, Debatten über wissenschaftliche Methoden. Diskutiert werden muss etwa auch, ob die Überlieferung von Forschungsdaten von derjenigen abgetrennt werden muss, die traditionell in historischen Archiven stattfindet. Gerade begleitende, administrative (physische wie digitale) Unterlagen sind wichtig für das, was hier infrage steht. Sammlungsprofile, Schutz- und Aufbewahrungsfristen, Kassationsentscheidungen etc., die festen Regeln und Vorschriften folgen, müssen nochmals gezielt abgestimmt und gestaltet werden. Das ist im Sinne einer qualitätsgesicherten Datenüberlieferung und im Sinne der Qualitätssicherung der Wissenschaft selbst.

Zu 3.

Im Papier der Fachgesellschaften ist mehrfach die Rede von der Bedeutung der Nachnutzung von Forschungsdaten. Dennoch entsteht der Eindruck, das Forschungsdatenarchiv ist der Endpunkt eines Datenlebens und es gäbe eine qualitative Hierarchie von Primär- und Sekundärnutzung von Daten wie auch von Primär- und Sekundärforschenden (DGfE/GEBF/GFD 2020, S. 3). Für die historische Forschung ist das Archiv der Startpunkt. Die Archivrecherche, Dokumentensammlung und die darauf folgende sorgfältige Zusammenstellung eines für die jeweilige Forschungsfrage angemessenen Quellen- bzw. Dokumentenkörpus (wir könnten auch sagen: Datenkorpus) ist eine oft hochkomplexe Tätigkeit. Sie steht am Beginn und ist integraler Bestandteil von neuer („primärer“) Forschung. Dies ist die Perspektive, aus der die Anforderungen an die Archivierung von Forschungsdaten bewertet werden sollten. Aus den Anforderungen leiten sich die Qualitätsmerkmale der Forschungsdaten ab: „Eine pragmatische Kurzdefinition für Datenqualität ist die Formel ‚fit for purpose‘, also eine Orientierung am Zweck oder der Absicht der Verwendung. [...] Datenqualität wird auf diese Weise umfassend, aber bestimmungsoffen definiert als Gesamtheit von Eigenschaften und Merkmalen von Daten hinsichtlich ihrer Eignung, einen bestimmten Zweck zu erfüllen.“ (RfII 2019, S. 25). Der Zweck der Datenarchivierung sollte in einer breiten fachlichen Debatte genauer be-

stimmt und mit Leben gefüllt werden. Der Horizont sollte dabei umfassend von der Dokumentation aktueller Forschungsergebnisse über Re- und Sekundäranalysen bis zu den archivwissenschaftlichen und geschichtswissenschaftlichen Anforderungen reichen.

Daraus folgt, dass die Forschungsdatenarchivierung zur Nachnutzung denjenigen, die die Daten „im Feld“ erhoben haben, eine große Verantwortung auferlegt (nicht nur, weil jeder unmittelbare Feldzugang eine kostbare Ressource ist). Sie müssen Qualitätskriterien vor einem breiten fachlichen Horizont erfüllen. Denn jede denkbare Nachnutzung von Daten fußt auf deren Datenqualität und leitet aus den vorgefundenen Datensätzen neue ab und bringt sie erneut in den Research-Data-Lifecycle ein. Der Respekt vor der Leistung derjenigen, die Datensätze erstellt haben, ist dabei so selbstverständlich wie das korrekte Zitieren.

Forschungsdaten werden künftig noch verstärkt in der historischen Bildungsforschung neben allen bisher verwendeten physischen und digitalen Dokumenten eine weitere Quellengattung für die historische Analyse sein. Die Gleichzeitigkeit und das Ineinandergreifen von neuen Erhebungen und Nachnutzung von Forschungsdaten und das Angewiesensein auf lückenlose Nachweis Ketten bis zurück zum originären Forschungsgegenstand wird das wissenschaftliche, methodische Arbeiten auch in den anderen Disziplinen der Bildungsforschung mindestens so stark bestimmen wie die Auswahl einer Befragungsmethode oder die Entscheidung für eine Analysetechnik. Wir brauchen also einen lückenlos belegten Bezug der Daten zu ihrem Ursprung in der (physischen) Welt. Ihr Entstehungskontext muss ebenfalls rekonstruierbar sein. Eine gute Leitlinie zur Entwicklung von Kriterien und einer umfassenden Praxis für das Forschungsdatenmanagement bieten die FAIR Guiding Principles (Wilkinson et al. 2016). FAIR steht für:

Findable: Die (Meta)daten sind im Netz dauerhaft auffindbar und über eindeutige Identifier (wie z. B. Digital Object Identifier (DOI)) referenzierbar.

Accessible: Die (Meta)daten sind über standardisierte, offene Computerschnittstellen zugänglich.

Interoperable: Die (Meta)daten basieren auf formalen, offenen und verbreiteten Standards und Vokabularen zur Wissensrepräsentation.

Reusable: Die Beschreibung der (Meta)daten ist hinsichtlich ihrer Nutzungsbedingungen, Herkunft und fachlicher Anforderungen so vollständig, dass eine wissenschaftliche Nachnutzung möglich ist.

In Analogie zur Bereitstellung von Archivadokumenten oder Buch- und Zeitschriftenpublikationen durch Archive und Bibliotheken unterstützt die Beachtung der FAIR-Prinzipien die Nachvollziehbarkeit und Nachprüfbarkeit einer Forschungsarbeit, wenn darin bspw. große Datenmengen, automatische Analysen, Visualisierungen oder Modellierungen eine tragende Rolle spielen. Die

Ausrichtung der FAIR Principles auf maschinelle Weiterverarbeitung der Metadaten und Daten zielt zudem darauf, Nachnutzung einmal erzeugter Forschungsdaten zu ermöglichen.

In Bezug auf die ersten drei FAIR-Prinzipien ist das Fach mit seinen bestehenden FDZen bereits gut aufgestellt. Diskussionsbedarf besteht insbesondere in Bezug auf das Prinzip der Nachnutzbarkeit, die noch nicht umfassend genug implementiert ist. Die FAIR-Prinzipien sehen wir, wie ursprünglich intendiert, „as a guide to data publishers and stewards to assist them in evaluating whether their particular implementation choices are rendering their digital research artefacts Findable, Accessible, Interoperable, and Reusable“ (Wilkinson et al. 2016, S. 5).

Fazit

Die Gestaltung von Forschungsdateninfrastrukturen, Empfehlungen zum Forschungsdatenmanagement, eine Kultur der Datenpublikation und Archivierung sind zutiefst fachliche Aufgaben. Wir plädieren dafür, dass auch die bildungswissenschaftlich Forschenden – und zwar aus allen Subdisziplinen – sich des Themas der Forschungsdaten als eines methodischen und methodologischen annehmen, das weit über die Debatten zwischen quantitativ und qualitativ empirisch Forschenden zu methodologischen Fragen hinaus geht. Wir könnten uns vorstellen, dass dafür organisatorische und zeitliche Orte geschaffen werden – etwa durch die einzelnen Fachgesellschaften –, in denen die anstehenden Fragen und Probleme diskutiert und laufend bearbeitet werden können.

Sabine Reh, Prof. Dr., ist Stellvertretende Geschäftsführende Direktorin des DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation und Direktorin der Bibliothek für Bildungsgeschichtliche Forschung des DIPF und Hochschullehrerin für Historische Bildungsforschung an der Humboldt-Universität zu Berlin

Lars Müller, MA, ist Wissenschaftlicher Bibliothekar an der Bibliothek für Bildungsgeschichtliche Forschung des DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation in Berlin.

Stefan Cramme, Dr., ist Stellvertretender Leiter der Bibliothek für Bildungsgeschichtliche Forschung des DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation in Berlin.

Bettina Reimers, Dr., ist Leiterin des Archivs der Bibliothek für Bildungsgeschichtliche Forschung des DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation in Berlin.

Marcelo Caruso, Prof. Dr., ist Hochschullehrer für Historische Bildungsforschung im Institut für Erziehungswissenschaften der Humboldt-Universität zu Berlin.

Literatur

- Bächle, Thomas Christian (2016): *Digitales Wissen, Daten und Überwachungen zur Einführung*. Hamburg: Junius.
- DGfE/GEbF/GFD (2020): *Empfehlungen zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten im Kontext erziehungs- und bildungswissenschaftlicher sowie fachdidaktischer Forschung*. Gemeinsame Stellungnahme der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE), der Gesellschaft für Empirische Bildungsforschung (GEbF) und der Gesellschaft für Fachdidaktik (GFD) zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten in den Erziehungs- und Bildungswissenschaften und Fachdidaktiken. https://www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2020.03_Forschungsdatenmanagement.pdf [Zugriff: 7. Oktober 2020].
- Fickers, Andreas (2020): Update für die Hermeneutik. *Geschichtswissenschaft auf dem Weg zur digitalen Forensik?* In: *Zeithistorische Forschungen/Studies in Contemporary History* 17,1, S. 157-168, <https://doi.org/10.14765/zzf.dok-1765>.
- Meier, Albrecht/Müller von Blumencron, Matthias/Schröder, Miriam (2020): EU-Digitalkommissarin Margrethe Vestager im Interview „Chinesische Firmen müssen nach unseren Regeln spielen“. In: *Der Tagesspiegel* vom 23. August 2020, S. 3.
- Müller, Lars (2019): *Kooperatives Management geisteswissenschaftlicher Forschungsdaten*. In: *ABI Technik* 39, 3, S. 194-201. <https://doi.org/10.1515/abitech-2019-3003>.
- Nassehi, Armin (2019): *Muster. Theorie der digitalen Gesellschaft*. München: Beck. <https://doi.org/10.17104/9783406740251>.
- RfII – Rat für Informationsinfrastrukturen (2019): *Herausforderung Datenqualität. Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel*. <http://www.rfii.de/?wpdmdl=4043> [Zugriff: 7. Oktober 2020].
- Stalder, Felix (2016): *Kultur der Digitalität*. Frankfurt am Main: Suhrkamp.
- Tiedemann, Norma/Huke, Nikolai (2020): *Die tödlichen Folgen der Realitätsverweigerung*. In: *Der Tagesspiegel* vom 22. August 2020, S. 21.
- Wagner, Gerad (2020): *Gibt es die Replikationskrise? Wissenschaftliche Irrtümer in Serie*. In: *FAZ* vom 12. April 2020, [FAZ.net](https://www.faz.net).
- Wilkinson, Mark D. et al. (2016): *The FAIR Guiding Principles for scientific data management and stewardship*. In: *Scientific Data* 3, 1, <https://doi.org/10.1038/sdata.2016.18>.