



Bericht zum dritten Netzwerktreffen des Verbunds Forschungsdaten Bildung

Verbund Forschungsdaten Bildung // Juli 2019

1. Einleitung

Am 11. April fand das dritte Netzwerktreffen des Verbunds Forschungsdaten Bildung (VerbundFDB) in Berlin am Institut zur Qualitätsentwicklung im Bildungswesen (IQB) statt. Der VerbundFDB ist zentraler Ansprechpartner für Forschungsdaten und Forschungsdatenmanagement in der empirischen Bildungsforschung und bietet Services in drei Schwerpunkten an: Daten finden, Daten teilen und Daten managen. Hierbei kooperieren die drei Kernpartner des VerbundFDB – GESIS, IQB und DIPF – eng mit Vertreterinnen und Vertretern anderer Forschungsdatenzentren (Apaek, BIBB, DIE, DIW Berlin, DJI, DZHW, FD-Lex, IAB, LfBi, Qualiservice Bremen und ZPID).

Gemeinsames Ziel dieses Netzwerks ist es, die Sichtbarkeit und Nutzbarkeit von Forschungsdaten zu erhöhen. Im Netzwerk gilt es, sich über gemeinsame Standards und Verfahren der Archivierung auszutauschen, um ein hohes Maß an Qualität, Effizienz und Nutzerfreundlichkeit der angebotenen Dienste zu erreichen. Hierfür treffen sich die Netzwerkpartner sowie Vertretungen der Geschäftsstelle des Rats für Sozial- und Wirtschaftsdaten (RatSWD) regelmäßig.

Beim dritten Netzwerktreffen, welches Gegenstand dieses Berichtes ist, haben die Anwesenden (Abschnitt 2) die Entwicklungen des VerbundFDB im vergangenen Jahr resümiert sowie anstehende Entwicklungen diskutiert (Abschnitt 3). Anschließend wurden die Ergebnisse der Arbeitsgruppe Metadaten (AG Metadaten; Abschnitt 4) zusammengefasst. Als Schwerpunkt des Treffens wurde das Thema Nutzungsbedingungen ausführlich besprochen (Abschnitt 5). Abschließend haben die Anwesenden Themen für die nächsten Treffen eruiert (Abschnitt 6).

2. Anwesende

(in alphabetischer Reihenfolge)

Doris Bambey, DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt
Jonathan Blumtritt, Cologne Center for eHumanities (CCeH), Köln
Marie Bormann, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin
Daniel Buck, Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Hannover
Mathias Bug, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin
Andreas Daniel, Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Hannover
Marcus Eisentraut, GESIS - Leibniz-Institut für Sozialwissenschaften, Köln
Anett Friedrich, Bundesinstitut für Berufsbildung (BIBB), Bonn
Marius Gerecht, DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt
Sabine Hachmeister, FD-Lex/Mercator Institut, Köln
Malte Jansen, Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Berlin
Nadeshda Jung, DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt
Harald Kaluza, Deutsches Institut für Erwachsenenbildung (DIE), Bonn
Helge Kminek, Archiv für pädagogische Kasuistik (Apaek), Frankfurt
Susanne Kretzer, Qualiservice, Bremen
Reiner Mauer, GESIS - Leibniz-Institut für Sozialwissenschaften, Köln
Claudia Neuendorf, Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Berlin
Lisa Pegelow, Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Berlin
Holger Quellenberg, Deutsches Jugendinstitut (DJI), München
Jessica Trixa, GESIS - Leibniz-Institut für Sozialwissenschaften, Köln
Dirk Weisbrod, DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt

3. Neues aus dem Verbund Forschungsdaten Bildung

Der VerbundFDB hat sich in seiner ersten Ausbaustufe (2016-2019) geöffnet, so dass die Services zur Datenarchivierung und Verfügbarmachung von allen Datengebenden der empirischen Bildungsforschung genutzt werden können¹. In der ersten Förderphase (2013-2016) war dies BMBF-geförderten Projekten vorbehalten. Zudem hat der VerbundFDB seine Vernetzung mit Forschungsdateninfrastrukturen weiter ausgebaut. Als neuer Partner wurde FD-Lex aufgenommen. In der zweiten Ausbauphase (2019-2021) bereitet sich der VerbundFDB auf die Aufnahme neuer Datentypen und die Verknüpfung von Daten(-beständen) vor. Zudem strebt der VerbundFDB eine stärkere Internationalisierung an. Mittelfristiges Ziel ist die Verstetigung des VerbundFDB ab 2022.

Auf die Entwicklungen in den drei Bereichen – Daten finden, Daten teilen, Daten managen – wird im Folgenden näher eingegangen.

Daten finden

Der Suchraum für Forschungsdaten wird erweitert, sodass künftig auch Studien mit Bezug zur Bildungsforschung aus anderen Forschungsdatenzentren (bspw. der Netzwerkpartner) in der Studiensuche des VerbundFDB recherchierbar sind. Für technische Details sei auf die Ergebnisse der AG Metadaten in diesem Bericht verwiesen.

Daten teilen

Der Datenbestand wurde durch gezielte Akquise sukzessive ausgebaut. Durch die zunehmende Verpflichtung durch Drittmittelgeber, Daten zu sichern und verfügbar zu machen, ist in den nächsten Jahren mit einer steigenden Nachfrage nach den Archivierungsangeboten der FDZ zu rechnen. Zudem muss besondere Expertise für den Umgang mit sensiblen Daten sowie mit neuen Datenformaten aufgebaut werden. Neben dem Datenbestand wurden auch die Literaturnachweise von zugehörigen Publikationen ausgebaut.

Daten managen

Der VerbundFDB unterstützt Forschende der Bildungsforschung in allen Bereichen des Forschungsdatenmanagements. Hierfür wurde neben der direkten Beratung das Schulungsangebot weiter ausgebaut, welches nun auch Webinare und spezielle Train-the-Trainer-Schulungen für Multiplikatoren umfasst. Zudem wurden weitere, anwendungsbezogene Publikationen erstellt bspw. zur informierten Einwilligung und zur Datenaufbereitung.

Diskussion

Die Diskussion der genannten Entwicklungen ergab, dass die Bewerbung der Webinare und Veranstaltungen ausgebaut werden sollte, bspw. über Forschungsnewsletter und Mailinglisten der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE) oder dem Institut für Qualitative Sozialforschung (QSf-L).

¹ Siehe hierzu die Collection Policy des VerbundFDB
<https://www.forschungsdaten-bildung.de/collectionpolicy?la=de>

4. Bericht aus der AG Metadaten

Aus dem ersten Netzwerktreffen heraus hat sich die AG Metadaten gegründet, welche sich der Harmonisierung der Metadaten von Netzwerkpartnern und VerbundFDB widmet – mit dem Ziel, über den Metadatenaustausch die Nachweise im VerbundFDB-Portal zu erweitern. Mitwirkende sind neben den VerbundFDB-Kernpartnern die Vertreterinnen und Vertreter von DZHW, LfBj, SOEP und Qualiservice. Als zentrale Ergebnisse wurden der FDZ-Editor, die Nachweiserweiterung über da|ra sowie das neue Datenmodell ausführlich vorgestellt.

Über den FDZ-Editor können Datenzentren direkt Metadaten in das Datenmanagement-System des VerbundFDB eingeben (DMM2). Für die hierfür benötigten Rechte müssen sich die Dateneingebenden zunächst authentifizieren.

Alternativ können die Daten indirekt erfasst werden. Hierfür wird auf da|ra zurückgegriffen, den Registrierungsservice für Sozial- und Wirtschaftsdaten in Deutschland für die Vergabe von Digital Object Identifiern (DOI)². Diese ermöglichen eine eindeutige und dauerhafte Identifikation von Daten(-sätzen). FDZ haben bei der DOI-Registrierung auch die Möglichkeit, die Metadaten der zu registrierenden Forschungsdaten zu erfassen. Über eine OAI-Schnittstelle (Harvester) können die bei da|ra registrierten Metadaten zukünftig an das VerbundFDB-Portal übermittelt werden, wenn sie von Datengebenden entsprechend ausgezeichnet wurden. Im VerbundFDB-Portal werden diese Datenbestände auffindbar gemacht und gesondert gekennzeichnet.

Der FDZ-Editor sowie die indirekte Erfassung über da|ra wurden bereits erfolgreich getestet. Eine direkte Datenbank-Schnittstelle zwischen VerbundFDB und den FDZ soll prototypisch mit dem DZHW entwickelt werden.

Für den Austausch der Metadaten wurde in der AG Metadaten ein gemeinsames Metadatenkernset (Kernset Bildungsforschung) definiert³. Zudem wurde das (Meta-)Datenmodell des VerbundFDB überarbeitet. Zum Zeitpunkt des Netzwerktreffens wurde dieses Modell noch getestet. Ebenfalls noch in Bearbeitung war die Anpassung der Datenbank selbst – entsprechend des neuen Datenbankmodells. Datenmodell und Datenbank sollen zukünftig eine höhere Flexibilität bei der Zuordnung der Daten zu Studien und/oder Projekten ermöglichen sowie eine direkte Schnittstelle für den Austausch von Metadaten zwischen FDZ und VerbundFDB unterstützen. Der Beginn des Testbetriebs der Datenbank wurde auf Oktober 2019 terminiert.

Diskussion

Die Diskussion zeigte den Bedarf, das Metadatenchema weiter anzupassen und zu erweitern. Konkret wurde angeregt, Metadaten zum Urheberrecht zu ergänzen, um die Transparenz zu erhöhen. Darüber hinaus wurde der Wunsch geäußert, Metadaten auch in anderen Sprachen auszutauschen. Bisher ist der Metadatenaustausch nur für deutschsprachiges Vokabular möglich. Das Thema Mehrsprachigkeit sollte zukunftsprospektiv in der AG Metadaten vertieft werden.

² <http://www.da-ra.de/home/>

³ https://www.forschungsdaten-bildung.de/files/fdbinfo_7.pdf

Überarbeitungsbedarf wurde auch bei den kontrollierten Vokabularen identifiziert. Diese seien derzeit nur eingeschränkt für die Beschreibung qualitativer Daten geeignet. DZHW, DIPF und Qualiservice haben sich bereit erklärt, die Überarbeitung gemeinsam voranzubringen. Diskutiert wurde darüber hinaus, wie kompatibel die Terminologie der kontrollierten Vokabulare zu den DDI- und CESSDA-Vokabularen sein sollte.

Beim indirekten Metadatenaustausch über da|ra wurde angeregt, eine Qualitätsprüfung zu etablieren. Zudem solle die Kennzeichnung von da|ra-Importen im VerbundFDB-Portal und damit der Unterschied zu den vom VerbundFDB selbst erschlossenen Forschungsdaten nutzerorientiert und für Forschende verständlich formuliert werden.

5. Hauptthema: Nutzungsbedingungen

Das Hauptthema Nutzungsbedingungen wurde über zwei Inputs eingeleitet. Zunächst wurden gängige Zugangsformen identifiziert und definiert: Public Use File (PUF), Campus File (CF) und Scientific Use Files (SUF). Deutlich wurde hierbei die große Heterogenität der Definitionen für SUF sowie die unscharfe Abgrenzung zwischen CF und SUF. Der zweite Input vertiefte die Inhalte von Datennutzungsbedingungen, welche in die Kategorien „allgemeine Zugangsvoraussetzungen“, „Umgang mit Daten“, „Zitierpflicht“ und „Konsequenzen bei Verstößen“ unterteilt wurden. Anschließend wurden verschiedene Formen von Nutzungsbedingungen thematisiert (bspw. Vertrag, AGB, Lizenz, (Be-) Nutzungsordnung).

Diskussion

In der Diskussion zu den Zugangsformen gaben einige Netzwerkpartner Einblicke in ihren Umgang mit den Zugangsformen, wobei deutlich wurde, dass es sehr heterogene Ansätze bei der Definition und Anwendung von Zugangsformen auch zwischen den beteiligten FDZ gibt. Einige Netzwerkpartner gaben an, gar nicht mit den Begriffen SUF, CF und PUF zu arbeiten. Es stellte sich heraus, dass eine überinstitutionelle Harmonisierung der Zugangsbedingungen wegen der unterschiedlichen institutionellen Rahmenbedingungen der FDZ schwierig ist. Der Austausch über die Nutzungsbedingungen wurde jedoch als hilfreich für die interne Diskussion und Überarbeitung gesehen. Verwiesen wurde auch auf die Arbeitsgruppe „Gemeinsame Richtlinien in den Forschungsdatenzentren“ des RatSWD⁴. Die Arbeitsgruppe hat unter anderem verschiedene Nutzungsverträge vergleichend gegenübergestellt. Zudem sei dort geplant, Vertragstemplates zu erstellen. Es wurde ein Erfahrungsaustausch zwischen der Arbeitsgruppe und dem VerbundFDB-Netzwerk angeregt.

In der Diskussion zu den Datennutzungsbedingungen stellte sich heraus, dass die Nutzungsbedingungen der FDZ alle in der Einführung benannten Kategorien umfassen: „allgemeine Zugangsvoraussetzungen“, „Umgang mit Daten“, „Zitierpflicht“ und „Konsequenzen bei Verstößen“. Dennoch finden sich Unterschiede in den konkreten Ausgestaltungen. Vertiefend wurde der Umgang mit Verstößen gegen vertragliche Vereinbarungen diskutiert, wobei das De-Anonymisierungsverbot und die Zitierpflicht als besonders bedeutsam betont wurden. Das De-Anonymisierungsverbot sei

⁴ <https://www.ratswd.de/arbeitsgruppen/fdz-richtlinien>

eine Bedingung, gegen die eher selten verstoßen werde. Die Zitationspflicht und die Anzeigepflicht von Publikationen würden von Forschenden hingegen häufig nicht eingehalten. Beides diene aber sowohl der Sichtbarmachung der Arbeit an den FDZ (Messung des Impacts) als auch der Sicherung guter wissenschaftlicher Praxis.

Die Diskussion der Konsequenzen bei Nichteinhaltung der Nutzungsbedingungen bzw. der vertraglichen Vereinbarungen verdeutlichte Unterschiede in der Praxis. Einige FDZ waren zum Zeitpunkt des Netzwerktreffens noch mit der Erarbeitung eines Sanktionskatalogs beschäftigt. Die FDZ waren sich in der Diskussion jedoch einig, dass Vertragsverletzungen bei der Zitierpflicht und Meldepflicht von Publikationen in der Regel nicht auf einen Vorsatz der Forschenden beruhten und auch mit höheren Vertragsstrafen nicht minimiert werden könnten. Vielmehr sei es wichtig, für diese Themen zu sensibilisieren (bspw. bei der Beratung und Betreuung von Arbeiten/Projekten oder auf Tagungen und Workshops) und Hilfestellungen für Forschende anzubieten. Als Ideen wurden zum einen die Bereitstellung praktischer Hinweise für Datennutzende in Form von Publikationen, FAQs, Webinaren, kurzen Clips oder Links in Verträgen genannt. Zum anderen wurde die Bereitstellung niedrigschwelliger Tools als hilfreich angesehen (bspw. die Möglichkeit eines Imports der Datenzitation in das Literaturverwaltungssystem über die Website oder der Einsatz von Crypto-Containern beim Dateiaustausch). Als zentrales Element wurde von den Anwesenden die universitäre Ausbildung genannt, in der bereits Studierende im Forschungsdatenmanagement und dem richtigen Umgang mit Forschungsdaten geschult werden sollten.

Bezüglich der Form von Nutzungsbedingungen wurden zunächst die Vor- und Nachteile von Verträgen und Lizenzen gegenübergestellt. Als Ergebnis wurde festgestellt, dass eine Lizenz – im Gegensatz zum Vertrag – in der gleichen Form für alle Nutzenden gelte. Sie habe damit den Vorteil, dass sie allgemein verbindlich und öffentlich einsehbar sei sowie universelle Gültigkeit besitze – auch für Dritte, die ohne formellen Vertrag in den Besitz der Daten gelangten. Der Vorteil eines Vertrags bestehe hingegen darin, dass dieser durch die persönliche Unterschrift mit größerer Wahrscheinlichkeit gelesen werde. Dadurch würden die Nutzenden sensibilisiert, dass der Umgang mit personenbezogenen Daten keineswegs unbedenklich ist. Zusätzlich könnten mit Verträgen sowohl den Datengebenden als auch den Studienteilnehmenden der hohe professionelle Anspruch der FDZ und die Vertrauenswürdigkeit der Datennutzung vermittelt werden. Hinsichtlich der rechtlichen Konsequenzen gebe es keine Unterschiede zwischen den zwei Formen.

Von den FDZ wurde am Ende der Diskussion eine zweistufige Form der Nutzungsbedingungen favorisiert: (1.) Allgemeine, standardisierte Regeln sollten in einer Nutzungsordnung zusammengefasst werden. (2.) Besondere Inhalte, die von den Nutzenden zwingend beachtet werden müssen (z. B. De-Anonymisierungsverbot), sollten in den Vertrag aufgenommen werden, um diese hervorzuheben. Auch eine flexible Gestaltung der Verträge wurde diskutiert. Als Vorteil wurde die flexible Reaktion auf die unterschiedlichen Datenlagen betont (bspw. abgestuft nach Sensibilität der beantragten Daten oder der Art des Forschungsprojekts). Ein Nachteil sei, dass eine größere Modularität die Datenweitergabe durch die zusätzlichen Verwaltungsaufwände verteuere.

Daran anschließend wurde allgemein über Geschäfts- bzw. Refinanzierungsmodelle gesprochen. Aufwände und Kosten könnten an Datengebende und/oder Datennutzende weitergegeben werden, insbesondere für kostenintensive Prozesse. Dazu gehörten v. a. manuelle Prozesse, Betreuung, Beratung, Vertragsgestaltung (Personalkosten) oder On-site-Nutzung (Safe-Rooms). Bei der

Kosten-Nutzen-Analyse müsse jedoch berücksichtigt werden, dass durch die Gebührenerhebung wiederum Kosten entstehen (bspw. für die Rechnungsstellung oder das Mahnwesen).

In der weiterführenden Diskussion tauschten sich die FDZ über das Erfassung von Publikationen mit nachgenutzten Daten aus. Idealerweise melden die Datennutzenden ihre Publikationen an das jeweilige FDZ und zitieren pflichtgemäß. Werden Publikationen nicht gemeldet, könne dies jedoch nur schwer kontrolliert werden. Die FDZ berichteten unterschiedliche Strategien beim Umgang mit dieser Problematik: Nutzende werden aktiv angefragt (bspw. nach Ablauf der Vertragslaufzeit), automatische Suchtools werden eingesetzt (bspw. Google Scholar Alerts) und/oder Mitarbeitende suchen gezielt nach Publikationen.

Weiterer Diskussionspunkt war das Spannungsverhältnis zwischen der freien Nachnutzung (FAIR-Prinzipien) von Forschungsdaten und den Rechten der Forschenden an den von ihnen erhobenen Daten. Die Anwesenden waren sich einig, dass Forschungsdaten möglichst zeitnah und umfassend zur Nachnutzung bereitgestellt werden sollten, da Forschungsprojekte in der Regel steuerfinanziert seien. Zudem solle eine Überregulierung des Zugangs vermieden werden, welche die Nachnutzung teilweise zu stark einschränke. Eine einfache Registrierung hingegen erhöhe die Nutzungszahlen. Andererseits könne ohne Verträge häufig nicht die entsprechende rechtliche Absicherung gewährleistet werden. Darauf aufbauend wurden Angebote der FDZ erörtert, um die Rechte der Datenerhebenden zu wahren und zugleich Vorbehalte zu minimieren. Hierfür bieten einige Netzwerkpartner Beratung an. Alle Netzwerkpartner bieten den Datenerhebenden die Möglichkeit, den Datenzugang zu restringieren, bspw. über Sperrvermerke für Datensätze, bzw. einzelne Variablen oder über Einspruchsmöglichkeiten bei Nachnutzung durch Dritte. Das konkrete Vorgehen variiert jedoch zwischen den FDZ.

Bei den Datennutzungsverträgen wurden von den FDZ unterschiedliche Regelungen für die Vertragspartner bzw. Unterschriftsberechtigte berichtet: (1.) Datennutzende und deren Vorgesetzte unterschreiben, (2.) Unterscheidung zwischen Hauptnutzenden und Mitnutzenden, (3.) Alle Personen, die mit den Daten oder auch nur mit Teildaten arbeiten, müssen einen eigenen Vertrag unterzeichnen. In der Diskussion stellte sich heraus, dass neben inhaltlichen Gründen für die unterschiedlichen Regelungen, institutionelle Anforderungen oder Eigeninteressen die Ausgestaltung der Verträge mitbestimmen.

Daran anschließend wurde die Anonymisierung der Datenherkunft diskutiert. Auf der einen Seite wurde es als berechtigte Forderung der Forschenden angesehen, die Datenherkunft zu anonymisieren, wenn im Rahmen der Metaforschung der oder die betreffende Forschende selbst zum Gegenstand einer Studie wird. Dem entgegen stünden allerdings der Anspruch auf Transparenz und Nachvollziehbarkeit bei öffentlich finanzierten Forschungsprojekten. Für konkrete Einzelfälle wurde auf Ethikkommissionen verwiesen, die über die Freigabe der Daten entscheiden könnten.

Zuletzt wurde der Umgang mit fremdsprachigen Daten besprochen: Diese könnten aufgrund von Sprachbarrieren oftmals nicht angemessen bewertet werden, um sie zur Nachnutzung zur Verfügung zu stellen.

6. Abschlussdiskussion

In der Abschlussdiskussion benannten die Netzwerkpartner Themen und Bedarfe, die bei weiteren Treffen diskutiert werden sollten. Folgende Themen wurden genannt:

1. Datenüberlassungsverträge
 - a. Abgleich und Diskussion der Datenüberlassungsverträge und Nutzungsbedingungen
 - b. Inhaltliche Bearbeitung von Schwerpunktthemen, bspw. Langzeitarchivierung personenbezogener Daten oder die nicht-wissenschaftliche Nutzung von Forschungsdaten
 - c. Konsequenzen bei einer Schließung eines Datenzentrums, bspw. bezogen auf den Verbleib der Daten oder die vertraglichen Verpflichtungen
2. Nationale und internationale Zusammenarbeit:
 - a. Gestaltung, Hürden und Potentiale der Zusammenarbeit mit Forschungsverbänden, internationalen Projekten oder internationalen Daten
 - b. Internationales Datenmanagement (bspw. Anonymisierungsstrategien) und internationales Vertragsmanagement
3. Betriebs- und Geschäftsmodelle
 - a. Möglichkeiten der Kostenkompensation bzw. Refinanzierung
 - b. Austausch über Erfahrungswerte

Nach dem allgemeinen Meinungsbild sollte das Kernthema des nächsten Netzwerktreffens entweder Datenüberlassungsverträge sowie Ergebnisse der AG-Nutzungsordnung oder Ideen zum Geschäftsmodell sein.