

Warum Forschungsdaten nicht publiziert werden

Ben Kaden

[Vorbemerkung: Eine erste Version dieses Textes erschien am 13. März 2018 im LIBREAS. Library Ideas-Blog¹. Da das Thema der (Nicht-)Publikation von Forschungsdaten offensichtlich eines der aktuellen Trendthemen im wissenschaftlichen Publikationswesen ist, soll er im Rahmen der Rubrik *LIBREAS. Dokumentation* in einer durchgesehenen Form zweitveröffentlicht werden. Der Autor beschäftigt sich aktuell im Rahmen des DFG-Projektes eDissPlus intensiv mit dem Themenfeld der dissertationsbezogenen Forschungspublikation, was den unmittelbaren Hintergrund dieses Textes bildet.]

Eine große und vermutlich noch zu wenig systematisierte Frage aller Diskussionen um eine Offene Wissenschaft lautet zumindest für die in diesem Bereich aktiven Infrastrukturen: Was spricht eigentlich dagegen? Die Erfahrungen aus dem Open-Access-Bereich und mehr noch aus dem der Open Science beziehungsweise Open Scholarship zeigen jedenfalls, dass es nicht selten eine erhebliche Lücke zwischen Wünschen, Zielen und Vorstellungen der Forschungsinfrastruktur und den besonders engagierten fachwissenschaftlichen Vertreter*innen in diesem Bereich und einer Gruppe gibt, die hier verkürzt als „Mainstream“ der Wissenschaft bezeichnet werden soll.

Eine wichtige, wenngleich auch nicht ganz überraschende Einsicht aus den jahrelangen Auseinandersetzungen mit der Offenen Wissenschaft muss lauten, dass die meisten Forschenden vor allem forschen möchten und zwar in der ihnen vertrauten Logik der wissenschaftlichen Publikationskulturen. Defizite auch der Publikationssysteme werden durchaus erkannt, aber nur dann tiefer adressiert, wenn sie zu spürbaren Behinderungen der individuellen Forschung führen. In den meisten Fällen wollen Forschende jedoch nicht als Innovator*innen für wissenschaftskommunikative und -infrastrukturelle Lösungen in einer Weise aktiv werden, die zu einer Umwidmung der Aufmerksamkeit vom Forschungsgegenstand auf diese Metastrukturen der wissenschaftlichen Kommunikation führt. Wo also der Leidensdruck im Umgang mit bestehenden Systemen und Praxen aus Sicht der Forschenden nicht übermäßig hoch ist und tradierte Formen nach wie vor die besten Karrierewege öffnen, werden auch hochengagierte und raffiniert geschliffene Keynote-Adressen für eine Offene Wissenschaft wenig ändern. Für wissenschaftliche Bibliotheken und andere Akteure der Wissenschaftsinfrastrukturen ist es folglich unerlässlich, zu wissen, welche Ansprüche, Herausforderungen und Ziele in den einzelnen Communities existieren. Dazu zählen auch die Gründe, warum Forschungsdaten und -materialien disziplinär zwar unterschiedlich intensiv aber nach wie vor eher insgesamt selten unter den Idealvorstellungen der Offenen Wissenschaft zugänglich gemacht werden.

Auf dem am 12. März 2018 bei der Wikimedia durchgeführten Open-Science-Bar-Camp² des

¹ <https://libreas.wordpress.com/2018/03/13/forschungsdatenpublikationen/>

² <http://www.open-science-conference.eu/barcamp/>

Leibniz Forschungsverbunds Science 2.0 gab es exakt dazu eine Session mit dem Titel „Valid reasons for opting out of sharing openly“. Einer Open-Science-Einstellung angemessen wurden einige Stichpunkte freundlicherweise auch für alle die sichtbar, die nicht teilnehmen konnten, in einem Etherpad hinterlegt.³

Ich habe mir erlaubt, diese Stichpunkte zu clustern und auszuformulieren. Im Anschluss an diese Liste ergänze ich noch einige Stichpunkte aus dem eDissPlus-Projekt⁴, das sich mit den Möglichkeiten des dissertationsbegleitenden Zugänglichmachens von Forschungsdaten befasst. Zu diesen zählen in der Gesamtschau auch Daten, die man in weicheren Fächern eher als Forschungsmaterialien bezeichnen würde, also beispielsweise Quellentexte oder Bilder.

Aufwand

- Forschende wollen ihre Zeit lieber in die Forschung selbst als in die Organisation eines Austauschprozesses für Forschungsdaten investieren.
- In der Projektplanung sind keine zeitlichen und personellen Ressourcen für die Aufbereitung von Forschungsdaten für ein Teilen beziehungsweise eine Veröffentlichung vorgesehen.
- Die Veröffentlichung beziehungsweise Zugänglichmachung von Forschungsdaten wurden nicht bei der Projektplanung beziehungsweise beim Erstellen des Forschungsdatenmanagementplans berücksichtigt und ist nachträglich zu aufwändig umzusetzen.⁵

Datenschutzrecht

- Die Veröffentlichung beziehungsweise Zugänglichmachung von Forschungsdaten ist aus datenschutzrechtlichen Gründen ausgeschlossen.
- Für eine Zugänglichmachung oder Publikation von personenbezogenen Daten liegt keine informierte Einwilligung vor.

Institutionelle / infrastrukturelle Ausstattung

- Die eigene Einrichtung bietet keine ausreichende Unterstützung sowohl infrastrukturell als auch beratend für die Verfügbarmachung beziehungsweise Publikation von Forschungsdaten an.

Institutionelle Vorgaben

- Prüfungsordnungen untersagen Promovierenden eine Publikation von Teilen der Promotion vor Abschluss des Promotionsverfahrens.
- Es gibt keine formalen Auswahlkriterien, welche Forschungsdaten wie zugänglich gemacht werden sollten.

³ https://etherpad.wikimedia.org/p/oscibar2018_session13

⁴ <https://www2.hu-berlin.de/edissplus/>

⁵ Zum Thema Forschungsdatenmanagementpläne vergleiche auch ausführlicher: Kaden, Ben: Forschungsdatenmanagementpläne sind eine Grundbedingung guter Wissenschaft. Meint Nature. In: LIBREAS. Tumblr, 21.03.2018 Online: <http://libreas.tumblr.com/post/172103346561/>

Möglichkeiten und Kompetenzen des Teilens / Publizierens

- Wissenschaftler*innen ist nicht bekannt, wo sie ihre Daten für eine Weitergabe hinterlegen können.
- Wissenschaftler*innen sind nicht zureichend geschult, um Forschungsdaten wissenschaftlichen Publikationsstandards entsprechend zugänglich zu machen oder zu publizieren.
- Forschungsdatenpublikationen sollen ein Peer-Review-Verfahren durchlaufen, das jedoch möglicherweise noch nicht existiert. Die nicht peer-reviewte Publikation von Forschungsdaten wird abgelehnt.

Persönliche Einstellung / Datenkontrolle / Wissenschaftsethik

- Wissenschaftler*innen sind am Thema Open Science / Offene Wissenschaft nicht interessiert.
- Wissenschaftler*innen möchten gern wissen, wer ihre Forschungsdaten nachnutzt, weshalb sie diese nur auf persönliche Anfrage weitergeben würden beziehungsweise sich vorbehalten, eine Weitergabe abzulehnen.
- Kooperationspartner in einem Forschungsdaten sprechen sich gegen eine Verfügbarmachung oder Publikation der im Projekt erzeugten Forschungsdaten aus.
- Die Zugänglichmachung von Forschungsdaten wird bewusst verweigert, weil entsprechende Anregungen und Vorgaben als Eingriff in die persönliche Wissenschaftsfreiheit interpretiert werden.
- Die eigenen Forschungsdaten werden als für eine Weitergabe zu wenig relevant eingeschätzt.
- Wissenschaftler*innen möchten verhindern, dass ihre Forschungsdaten für von ihnen nicht gewünschte Zwecke nachgenutzt werden.
- Es bestehen Zweifel daran, dass Dritte die Forschungsdaten bzw. Forschungsmaterialien wissenschaftlichen Standards entsprechend nutzen können.
- Es besteht die Sorge, dass durch Zugänglichmachung von Forschungsdaten Schwächen der Datenerhebung und -analyse sichtbar werden.
- Die konkreten Forschungsdaten sind in einer Weise manipuliert, die verborgen bleiben soll.

Verlags-, Urheber- und Nutzungsrecht

- Wissenschaftler*innen haben die Nutzungs- und Verwertungsrechte im Zuge einer Copyright-Vereinbarung an einen Wissenschaftsverlag übertragen und besitzen daher keine Verfügungsmöglichkeiten zum Teilen und Veröffentlichenden von Forschungsdaten.
- Promovierende, deren Forschungsprojekt in Kooperation mit Dritten stattfindet, haben nur begrenzt Verfügungsrechte über ihre Forschungsdaten. Dies betrifft insbesondere Kooperationen mit kommerziellen Partnern.

- Es ist nicht bekannt, wer die rechtliche Eigentümerschaft zu den jeweiligen Forschungsdaten besitzt.

Weitere Rechtsgebiete / Wissenschaftsethik

- Das Forschungsthema ist zu sensibel als dass die Forschungsmaterialien und Forschungsdaten frei und international verfügbar gemacht werden können.
- Es ist unklar, wer langfristig die Verantwortung für die jeweiligen Forschungsdaten und -materialien übernimmt.

Wissenschaftssoziologie

- Forschungsdaten und -materialien gelten als wissenschaftliches Kapital und werden (noch) zurückgehalten, weil sie in einem späteren Projekt weiter ausgewertet werden sollen.
- Forschungsdaten und Forschungsmaterialien sollen als exklusives Asset für einen Antrag auf Projektförderung angeführt werden. Sind sie frei verfügbar, sinkt, so die Wahrnehmung, die Chance auf Förderung.
- Forschungsdaten sollen zunächst exklusiv weiter beforscht werden, weshalb eine Publikation beziehungsweise Zugänglichmachung bestenfalls nach einem Embargo in Frage kommt.
- Die Publikation oder das Teilen Forschungsdaten wird nicht ausreichend als wissenschaftliche Leistung gewürdigt.

Wissenschaftsfreiheit

- Das Prinzip der Open Science / Offenen Wissenschaft sollte nicht als Druck wirken – im Sinne der Wissenschaftsfreiheit sollten Wissenschaftler*innen selbst entscheiden ob beziehungsweise wie und wann sie Forschungsdaten zugänglich machen.

Aus den Erfahrungen des eDissPlus-Projektes, das Einstellungsmuster von Promovierenden zum Publizieren von Forschungsdaten untersuchte, lassen sich, wie angekündigt, noch einige weitere Hürden benennen bzw. genannte Aspekte weiter differenzieren. Dies sind unter anderem:

Aufwand

- Der Aufwand für eine dissertationsbegleitende Forschungsdatenpublikation wird, wie auch schon bei der Forschungsplanung erwähnt, nur sehr selten in der Dissertationsplanung und – sofern überhaupt vorhanden – in Forschungsdatenmanagementplänen berücksichtigt.

Institutionelle und disziplinäre Vorgaben / Rahmenbedingungen

- In vielen Bereichen fehlen für Forschungsdatenmanagement und das Publizieren von Forschungsdaten Standards, die eine Orientierung geben können.
- Forschungsdatenpolicies werden im Einzelfall häufig als untauglich empfunden, unter anderem da sie zum Beispiel datenschutzrechtliche sowie weitere rechtliche Einschränkungen einer möglichen Forschungsdatenpublikation in keiner Weise würdigen.

- In vielen Disziplinen gibt es keinen nachhaltigen und systematischen Austausch darüber, welchen Stellenwert und welche Form Forschungsdatenpublikationen für wissenschaftliche Kommunikation haben sollten.
- Prüfungsordnungen treffen in der Regel keine Aussagen zu Forschungsdatenpublikationen und bieten daher auch keine Orientierung.
- Für den Titelerwerb sind Forschungsdatenpublikationen in den meisten Fällen nicht erforderlich.

Kompetenzen und Kompetenzvermittlung

- Bereits für das generelle Forschungsdatenmanagement werden häufig Vermittlungsdefizite benannt: Das Thema findet in Lehre und Methodenausbildung kaum statt. Die Frage der Forschungsdatenpublikation wird in der Regel überhaupt nicht angesprochen.
- Da in vielen Disziplinen Forschungsdatenpublikationen unüblich sind, kommen Promovierende auch bei der Literatursuche nicht mit dieser Gattung in Kontakt. Die Idee und Möglichkeit einer Forschungsdatenpublikationen ist ihnen daher häufig nicht bekannt.

Persönliche Einstellung / Datenkontrolle / Wissenschaftsethik

- Forschungsdaten werden im Rahmen von Promotionsprojekten häufig mehr als Mittel zum Zweck als als eigene publikationswürdige Größe angesehen.
- Promovierende sehen sich angesichts der geringen Etablierung von Forschungsdatenpublikationen in vielen Bereichen überfordert und nicht in der Lage, zusätzlich zu ihrer Promotion entsprechende Pionierarbeit für die Publikationskulturen ihrer Fächer zu leisten.

Rechtliche Aspekte

- Promovierende können oft die urheber- beziehungsweise erheber-rechtlichen Folgen einer Forschungsdatenpublikation nur unzureichend abschätzen. So scheint die Bedeutung der Creative-Commons-Lizenzen für konkrete Szenarien oft wenig eindeutig.
- Das Datenschutzrecht steht einer Publikation von dissertationsbegleitenden Forschungsdaten mit Personenbezug aktuell in fast allen Fällen im Weg, unter anderem da selten Forschungsdatenpublikationen von vornherein eingeplant und in den jeweiligen informierten Einwilligungen nicht vorkommen. Das nachträgliche Einholen eine Publikationserlaubnis ist oft nicht möglich oder wird als deutlich zu aufwändig eingeschätzt.

Wissenschaftssoziologie / Forschungsdatenkontrolle

- Forschungsdatenpublikationen versprechen in den meisten Fällen keinen zusätzlichen Reputationsgewinn. Teilweise wird von den Gutachter*innen eine dissertationsbegleitende Forschungsdatenpublikation sogar als potentiell schädlich eingeschätzt.
- Forschungsdaten gelten bei vielen Promovierenden als wissenschaftliches Kapital. Besteht die Bereitschaft zur Weitergabe, wird eine selektive Zugänglichmachung auf Anfrage deutlich gegenüber einer allgemeinen Zugänglichmachung als Publikation bevorzugt.

Beide Auflistungen sind sicher keinesfalls erschöpfend. Zu vielen Aspekten wären auch vertiefende Einzeluntersuchungen sinnvoll und notwendig. Deutlich wird jedoch bereits an dieser losen Reihung, dass individuelle Einstellungsmuster zwar einen wichtigen Aspekt darstellen und entsprechend Lobbyarbeit für Open Access und weitere Elemente der Offenen Wissenschaft sicher sinnvoll ist. Nachhaltig wirksam werden sie aber nur sein können, wenn auch entsprechende Rahmenbedingungen existieren und zwar sowohl infrastrukturell als auch fachkulturell.

Ein offensichtliches Haupthindernis liegt sicher im aktuell in vielen Fällen deutlichen Missverhältnis von Aufwand und Nutzen. Eine wissenschaftlichen Standards entsprechende Forschungsdatenpublikation erfordert gerade angesichts des Mangels an Best-Practice-Beispielen und auch im Einzelfall passenden Leitlinien eine vergleichsweise hohe zusätzliche Arbeitsbelastung, der jedoch kein erwartbarer Reputationsgewinn entgegensteht. Je niedrigschwelliger hier Infrastrukturen Beratung und andere Dienste anbieten können, desto besser. Die Universitätsbibliothek wurde von vielen der befragten Promovierenden im eDissPlus-Projekt als natürliche Ansprechpartnerin für alle Fragen zum Thema Forschungsdaten angesehen und zwar auch für Aspekte, die man gemeinhin eher den Instituten und der dortigen Ausbildung zugeschrieben hätte. Man wünscht sich von der Bibliothek idealerweise ein umfassendes Spektrum an Dienstleistungen von der Beratung über Cloud-Dienste bis zur Langzeitarchivierung für komplexe Datenstrukturen. Was davon wie tatsächlich angeboten werden kann, ist allerdings eine andere Diskussion. In der Erwartungshaltung der Promovierenden, die mit hoher Kompetenz an Forschungsdaten gehen, ist Github ein Benchmark. Bei den anderen eher Dropbox. Im Ergebnis weist der Wunsch in eine Richtung, die beide Dienste mit umfassenden Beratungsangeboten verbindet und überschätzt nebenbei deutlich die Entwicklungskapazitäten, die die öffentliche Hand an dieser Stelle bereitzustellen vermag.

Die zweite große Herausforderung liegt in einer unklaren Rechtslage in Bezug auf Forschungsdaten. Hierzu gab es im Januar 2018 einen Workshop an der Viadrina in Frankfurt/Oder,⁶ der erwartungsgemäß wenige Antworten dafür aber noch tiefere Einblicke in die Komplexität der Gemengelage bot.

Und schließlich fehlen für viele Disziplinen tatsächlich praktikable Infrastrukturangebote, auch übrigens von Verlagen oder anderen kommerzieller aufgestellten Anbietern, für eine zeitgemäße und dauerhafte Zugänglichmachung von Forschungsdatenpublikationen. Das Druckparadigma, dass sich im PDF-Format vergleichsweise angenehm spiegeln ließ, funktioniert für digitale Forschungsdaten endgültig nicht mehr. Will man sie zum festen Teil der wissenschaftlichen Kommunikation werden lassen, benötigt man oftmals überhaupt erst einmal adäquate mediale Präsentationsformen – eine Debatte übrigens, die in zahlreichen Bereichen bestenfalls nebenbei geführt wird. Implizit lässt sich hier auch aus den eDissPlus-Befragungen ein weiteres sehr großes Hindernis für Forschungsdatenpublikationen ermitteln: Die Daten sind unter Umständen mit den bestehenden Möglichkeiten gar nicht sinnvoll als Publikation darstellbar.

Zu all diesen Problemen existiert selbstverständlich engagierte Arbeit hinsichtlich möglicher Lösungen, auch wenn die Digitalisierung der Wissenschaft gerade im Infrastrukturbereich noch ganz anders auch von den Träger- und Förderinstitutionen adressiert werden könnte, als dies bislang geschieht. So ermöglicht beispielsweise der edoc-Server seit diesem Jahr Forschungsdatenpublikationen der Humboldt-Universität zu Berlin.⁷ Auch Zenodo kann als gelungenes Bei-

⁶ http://www.forschungsdaten.org/index.php/Rechtliche_Aspekte_bei_digitalen_Forschungsdaten

⁷ https://www.hu-berlin.de/de/pr/nachrichten/maerz-2018/nr_180309_00

spiel für einen zeitgemäßen Publikationsserver für eine Vielzahl denkbarer Materialien gelten. Dass Forschungsdaten auf den Publikationsservern mit Metadaten erschlossen, mit DOIs versehen werden und wenigstens teilweise sogar in Bibliothekskatalogen bibliografiert erscheinen, mag ebenfalls ein früher Schritt in Richtung Anerkennung als ordentliche wissenschaftliche Publikation sein. Aber damit endet in den meisten Fällen die Reichweite dessen, was Bibliotheken und Infrastrukturen zu leisten in der Lage sind.

Die Selbstorganisation der Wissenschaft macht es erforderlich, dass sich die Fachkulturen darüber verständigen, welchen Stellenwert in welcher Form die Publikation von Forschungsdaten und anderen Forschungsmaterialien für sie einnehmen kann und soll. Sie müssen selbst diskutieren, prüfen und entscheiden, ob sie zum Beispiel ein Peer Review wollen, ob komplexe Forschungsdatenpublikationen auch berufsrelevant sein können, welche Formate sie bevorzugen und welche Metadaten sie brauchen. Die Infrastrukturseite kann aufzeigen, was möglich ist, kann Erfahrungen, Erkenntnisse und Überblickswissen vermitteln. Dafür brauchen wir Veranstaltungen wie das Open-Science-Bar-Camp und Wissenschaftsforschung, wie sie im eDissPlus-Projekt stattfinden konnte. Die Absicherung eines Wissenstands auf der jeweiligen Höhe der Zeit zu den Praxen und Wünschen der Fachkulturen einerseits und den technischen Möglichkeiten andererseits ist bereits für sich eine enorme Herausforderung und zugleich Minimalbedingung jeder zielorientierten Infrastrukturentwicklung. Bereits dafür benötigt man, wenn man so will, Brückenakteure, die sowohl Fach- und Publikationskulturen als auch Ziele, Möglichkeiten, Grenzen von Wissenschaftsinfrastruktur und -organisation kennen. Man braucht solche Akteure aber noch mehr, wenn es darum geht, den eigentlichen Schritt einer digitalen Wissenschaft zu gehen, nämlich die Infrastruktur mit der wissenschaftlichen Kommunikation und an bestimmten Stellen direkt mit der Forschung zu verzahnen. Wir können auf Barcamps und in Workshops umfassend darüber diskutieren, warum Forschende ihre Daten nicht publizieren. Greifbare und praktikable Lösungen werden sich jedoch erst dann daraus ableiten lassen, wenn diese Diskussionen auch mit den Wissenschaftler*innen geführt werden. Dazu ist es notwendig, beide Seiten nicht nur zu kennen, sondern in einem stetigen Dialog zu halten. Ich habe eingangs bemerkt, dass Forschende vor allem forschen und sich möglichst wenig mit Infrastrukturfragen befassen wollen. Dies ändert sich bei der digitalen Wissenschaft natürlich dann, wenn Infrastruktur und Forschung zusammenfallen. Ein gutes Beispiel unter anderem auch für die Schwierigkeiten dieser Entwicklung sind die Digital Humanities.

Wir, als Vertreter zum Beispiel der Universitätsbibliotheken, bemühen uns unter anderem in Projekten wie eDissPlus intensiv darum, zu verstehen, was die Forschenden als Zielgruppen umtreibt. Konsequenterweise könnte sich das Konzept der Zielgruppe allerdings an nicht wenigen Stellen zunehmend relativieren und das Gewicht deutlich in Richtung einer Partnerschaft verschieben. Ein unmittelbares Desiderat ist aktuell ein Forum oder eine Form, das beziehungsweise die es uns ermöglicht, Erkenntnisse wie die oben zusammengetragenen in einen übergreifenden und gestaltungsorientierten Dialog mit allen Stakeholdern einzubringen. Ein zweites ist häufig eine stabile und ein idealerweise unkomplizierte Struktur, die es nach Ende von Projekten von eDissPlus erlaubt, über die, wenn man so will, Anamnese hinaus, zu konkretisieren, wie, mit welchen Mitteln und an welchen Stellen die in diesem Fall identifizierten Hürden abgebaut werden können. Diese Situation steht dabei exemplarisch für etwas sehr Generelles: Die Ansprüche einer Offenen Digitalen Wissenschaft werden sich nur als Projekt des Gesamtsystems Wissenschaft realisieren lassen.

Ben Kaden arbeitet an der Universitätsbibliothek der Humboldt-Universität zu Berlin und ist Mitherausgeber der LIBREAS.