

Bausteine Forschungsdatenmanagement  
Empfehlungen und Erfahrungsberichte für die Praxis von  
Forschungsdatenmanagerinnen und -managern

## **Discuss Data – Community-zentrierter Ansatz für das Forschungsdatenmanagement in den Geistes- und Sozialwissenschaften**

Felix Herrmann<sup>i</sup>      Daniel Kurzawe<sup>ii</sup>

2020

### **Zitiervorschlag**

Herrmann, Felix und Daniel Kurzawe. 2020. Discuss Data – Community-zentrierter Ansatz für das Forschungsdatenmanagement in den Geistes- und Sozialwissenschaften. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2020: S. 56-64. DOI: [10.17192/bfdm.2020.2.8277](https://doi.org/10.17192/bfdm.2020.2.8277).

Dieser Beitrag steht unter einer  
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

<sup>i</sup>Forschungsstelle Osteuropa an der Universität Bremen. ORCID: [0000-0002-5134-4378](https://orcid.org/0000-0002-5134-4378)

<sup>ii</sup>Niedersächsische Staats- und Universitätsbibliothek Göttingen. ORCID: [0000-0002-0121-9261](https://orcid.org/0000-0002-0121-9261)

## 1 Abstract

Dieser Artikel stellt einen Community-zentrierten Ansatz für das Forschungsdatenmanagement in den Geistes- und Sozialwissenschaften vor, der Daten explizit in einen Kontext zu ihrer Generierung, Nutzung und Kuratierung durch die entsprechenden Communities von Forschenden setzt. Discuss Data<sup>1</sup> ist eine Disziplin- und Methoden-übergreifende Plattform für die Regionalstudien zur post-sowjetischen Region und setzt diesen Community-zentrierten Ansatz in der Praxis um. Aufbauend auf vorhandene FDM-Dienste, etwa zur Authentifizierung, Datenspeicherung und -registrierung, bietet Discuss Data die Möglichkeit Datenbeschreibungen kollaborativ zu bearbeiten, unter Verwendung verschiedener Lizenzmodelle dauerhaft zu publizieren und zu diskutieren. Mittelfristiges Ziel ist die Schaffung eines Kommunikationsortes rund um Forschungsdaten, der von allen interessierten Forschungscommunities nicht nur aktiv genutzt, sondern auch redaktionell getragen und konzeptionell weiterentwickelt werden kann.<sup>2</sup>

## 2 Einleitung

Das Management digitaler Forschungsdaten hat sich in den letzten zehn Jahren grundlegend verändert. Die klassischen Anbieter wissenschaftlicher Informationsinfrastrukturen, wie etwa Universitätsbibliotheken und Archive, bauen zunehmend ihre Angebote für die digitale Forschung aus. Initial wurden dabei traditionelle Angebote in die digitale Welt überführt. Kataloge wurden digitalisiert und anstelle klassischer wurden digitale Medien bereitgestellt. Die sich aktuell noch entwickelnden Standards und die voranschreitende Digitalisierung in Forschung und Lehre verlangen jedoch ein Umdenken. Auch Endgeräte mit ihren spezifischen Paradigmen (insbesondere im Bereich User Experience) und den damit verbundenen wachsenden Fähigkeiten und Ansprüchen entwickeln sich rasant weiter.

Um dem digitalen Wandel in Forschung und Lehre gerecht zu werden, sind zudem nachhaltige Entwicklungsmodelle gefordert, auch und gerade mit Bezug auf die bislang oft ungeklärte langfristige Finanzierung digitaler Angebote. Während die Kernaufgaben des FDM durch Datenzentren und von diesen angebotene Dienste adressiert werden, ergibt sich eine weitere Dimension durch die Betrachtung der Datenqualität.<sup>3</sup> Die inhaltliche Bewertung der Datenqualität erfordert meist sehr spezifischer Kennt-

---

<sup>1</sup><https://discuss-data.net>.

<sup>2</sup>Diese Publikation ist entstanden im Rahmen des Discuss Data-Projektes, gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projekt Nummer 323616639.

<sup>3</sup>RfII – Rat für Informationsinfrastrukturen: Herausforderung Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel, zweite Auflage, Göttingen 2019, <http://www.rfii.de/download/herausforderung-datenqualitaet-november-2019/>.

nisse, die stark von der jeweiligen Fachperspektive abhängen.<sup>4</sup> Insbesondere in den Geistes- und Sozialwissenschaften ist das Mittel der Wahl zur Beurteilung von Datenqualität die offene akademische Diskussion, die traditionell in Publikationen sowie auf Konferenzen und Workshops stattfindet.

Fachkommunikation, Begutachtung und Netzwerkbildung verlagern sich zunehmend in den digitalen Raum, u.a. auch um der immer stärker ausgeprägten globalen Vernetzung der Wissenschaft Rechnung tragen zu können. Parallel dazu ändert sich die Rolle und Stellung von Daten durch Methoden der Data Science und der Einbeziehung der Öffentlichkeit.<sup>5</sup> Dies trifft insbesondere auch auf geisteswissenschaftlichen Disziplinen zu, welche teilweise noch an der Formulierung eines passenden Datenbegriffs arbeiten. Das Verständnis von Quell- und abgeleiteten Daten ändert sich ebenfalls. Der Anspruch an die Nachvollziehbarkeit im Sinne der guten wissenschaftlichen Praxis steigt dabei. Um dies zu reflektieren, bedarf es einer weiteren Ebene des digitalen FDM, welche die informationstechnische Verwaltung und Bereitstellung von Daten mit deren Kontextualisierung verknüpft und so Diskurse über die Daten sowie deren Kontext und Methoden direkt am Objekt zulässt.

### 3 Digitale Forschungsdaten im Diskurs

Längst ermöglichen Informationstechnologien die Bereitstellung und den unbeschränkten Zugriff auf Forschungsdaten als Open Data. Entsprechende Datenrepositorien bilden die Grundlage für Debatten um Forschungsergebnisse und gelten als wichtige Säule zur Herstellung wissenschaftlicher Transparenz.<sup>6</sup> Das Potential entsprechender Systeme geht weit über den Download bereitgestellter Dateien und deren Beschreibung hinaus – doch wird dieses Potential in der gegenwärtigen Praxis noch weitgehend vernachlässigt. Repositorien sehen sich dem Datenerhalt und der Kuratierung verpflichtet, während Mehrwerte etwa durch spezifische virtuelle Forschungsumgebungen geboten werden. Dies trennt jedoch den wissenschaftlichen Diskurs um Daten von den daraus erarbeiteten Inhalten weitgehend ab. Wissenschaftliche Artikel werden zwar auf Grundlage dieser Daten verfasst, doch fließen Kommentare und Anmerkungen nur ausnahmsweise und meist nicht durch strukturelle Maßnahmen unterstützt zu den Daten zurück. Das ein Bedarf für entsprechende Systeme besteht zeigt etwa die Nutzung des Quellcoderepositories Github als Plattform zur kollaborativen

<sup>4</sup>Andreas Heinrich, Felix Herrmann und Heiko Pleines, „Transparency and Quality Assessment of Research Data in Post-Soviet Area Studies: The Potential of an Interactive Online Platform“, *Journal of Eurasian Studies* 10, Nr. 2 (Juli 2019): 136–46, <https://doi.org/10.1177/1879366519850698>.

<sup>5</sup>Gibbons, Michael, Peter Scott und Helga Nowotny. *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity Press, 2001.

<sup>6</sup>RatSWD, „Stellungnahme des RatSWD zur Archivierung und Sekundärnutzung von Daten der qualitativen Sozialforschung“ (Rat für Sozial- und Wirtschaftsdaten, 2015), [http://www.ratswd.de/dl/RatSWD\\_Stellungnahme\\_QualiDaten.pdf](http://www.ratswd.de/dl/RatSWD_Stellungnahme_QualiDaten.pdf); RfII – Rat für Informationsinfrastrukturen, „Leistung aus Vielfalt“.

Forschung.<sup>7</sup> Forschende haben hier in einer neuen Form der Wissenschaftskommunikation gezeigt, wie die iterative Arbeit an Daten und der Austausch über Teamgrenzen hinweg eine übergreifende Forschungsleistung unterstützen kann. Längst diskutieren Forschende auch an anderen Stellen im Netz miteinander. Mit Academia.edu und ResearchGate.net konnten sich zwei kommerziell betriebene Social Media Plattformen mit Forschenden als primärer Zielgruppe etablieren, die ihre Angebote vor allem um publizierte Forschungsergebnisse herum gruppieren. Die Nutzung wird zunehmend kritisch hinterfragt, insbesondere weil die langfristige Verfügbarkeit der eingestellten Inhalte nicht garantiert ist und einzig von der Geschäftsstrategie der beteiligten Unternehmen bestimmt wird.<sup>8</sup> Der eigenverantwortliche Betrieb passgenauer Kommunikationsräume liegt im Interesse der wissenschaftlichen Gemeinschaft, um sowohl den Einfluss auf die Gestaltung und Qualität der Angebote, als auch auf die Sicherung der dort eingestellten Inhalte und den dauerhaften Zugriff zu wahren und so den Diskussionen akademische Legitimität zu verleihen.

## 4 Discuss Data als freie Diskussionsplattform

Discuss Data stellt eine Webplattform dar, die den Anspruch hat, das Management von Forschungsdaten aus den Regionalstudien zum post-sowjetischen Raum so zu gestalten, dass ein Kommunikationsort für die community-spezifische Pflege, Annotierung und Diskussion der Forschungsdaten entsteht. Das von der DFG geförderte Kooperationsprojekt zwischen der Forschungsstelle Osteuropa an der Universität Bremen und der Niedersächsischen Staats- und Universitätsbibliothek Göttingen startete 2017 mit der Entwicklung, aktuell läuft die Plattform im geschlossenen Beta-Betrieb und der Launch ist für Herbst 2020 vorgesehen. Discuss Data ist als öffentlich finanzierte Infrastruktur frei von kommerziellen Interessen oder Lizenzen und steht der Forschung ebenfalls frei zur Verfügung.<sup>9</sup> Die Plattform basiert auf Open Source Software wie dem Webentwicklungsframework Django, Postgres als Datenbank und Elasticsearch als Suchmaschine. Die Infrastruktur wird auf Servern der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) betrieben und dort auf mehreren Ebenen gesichert. Vorhandene FDM-Dienste sind nahtlos in die Abläufe integriert, etwa die Authentifizierung durch die DARIAH-AAI oder die Speicherung von als Open Access lizenzierten Daten im DARIAH-Repository<sup>10</sup>.

Discuss Data fördert die freie Publikation von Forschungsdaten entsprechend der FAIR

<sup>7</sup>Perkel, J. (2016). Democratic databases: science on GitHub. *Nature News*, 538(7623), 127.

<sup>8</sup>Kathleen Fitzpatrick, „Academia, Not Edu“, *Planned Obsolescence*, 26. Oktober 2015, <http://www.plannedobsolescence.net/academia-not-edu/>; Mike Thelwall und Kayvan Kousha, „ResearchGate: Disseminating, Communicating, and Measuring Scholarship?“, *Journal of the Association for Information Science and Technology* 66, Nr. 5 (1. Mai 2015): 876–89, <https://doi.org/10.1002/asi.23236>.

<sup>9</sup>Der Quellcode ist unter <https://gitlab.gwdg.de/discuss-data> frei zugänglich.

<sup>10</sup>Dank des modularen Aufbaus ist auch die Anbindung anderer Repositorien durch Implementierung entsprechender Module möglich.

Prinzipien<sup>11</sup>, ermöglicht aber auch andere Publikationsmodelle als Open Access, wenn dieser z.B. aus Gründen des Informantenschutzes nicht in Frage kommt. Die Regionalstudien zur post-sowjetischen Region dienen als prototypischer Anwendungsfall. Gerade ihr Methodenpluralismus und die interdisziplinäre Ausrichtung ermöglichen auf besondere Weise die Einbeziehung aller in den Geistes- und Sozialwissenschaften etablierten Datenformen und -formaten mit Schwerpunkten auf Textdokumenten, statistischen Daten, Umfragen sowie Audio-, Video- und Bildmaterialien. Das erleichtert die vorgesehene Übertragung auf weitere Communities und die Anpassung an deren spezifische Anforderungen.

## 5 Communities im Mittelpunkt

Der Community-zentrierte Ansatz in Discuss Data fokussiert auf die Kuratierung der Daten und damit vor allem auf die oberen Ebenen des research data management stacks. Die Plattform ist als Schicht zwischen verteilten Repositorien und den wissenschaftlichen Nutzern konzipiert. Ziel ist es, bestehende wissenschaftliche (Sub-)Communities zu motivieren und in die Lage zu versetzen, zur Pflege und Qualitätskontrolle der Forschungsdaten aus der eigenen knowledge domain umfassend beizutragen.

Dabei wird nicht nur davon ausgegangen, dass Forschungsdaten und deren Qualität zentral für die wissenschaftliche Forschung sind. Zum einen gestaltet sich angesichts der stetig wachsenden Datenmengen die klassische bibliothekarische Erschließung von Forschungsdaten mit Recherche-, Verzeichnungs- und Prüfprozessen zeitlich, personell und damit finanziell als sehr aufwändig. Zum anderen handelt es sich bei Forschenden um hochqualifizierte Expertinnen und Experten für die Erschließung, Kontextualisierung und Qualitätskontrolle von Forschungsdaten. Forschende sollen in Discuss Data nicht mehr nur Datengebende oder -nutzende sein, sondern sich aktiv in die Kuratierung der Forschungsdaten ihres Feldes einbringen.

Der Fokus auf Sub-Communities ist essentiell, um Einzelne zur aktiven Beteiligung zu motivieren und die virtuellen Diskussionsräume zu beleben. Erst durch deren engagierte Teilhabe können Diskussionen entstehen und damit wertvolle Inhalte generiert werden, die ihrerseits dazu beitragen, weitere Angehörige entsprechender Communities auf die jeweiligen Plattformen aufmerksam zu machen. Dabei sind enge Rahmen zielführender als breite Ansätze: Der Kreis derjenigen Forscherinnen und Forscher, die sich etwa mit Datenmaterial zu Protesten im post-sowjetischen Raum befassen, ist ob seiner Überschaubarkeit leichter zu adressieren, als ganze Disziplinen wie die Osteuropastudien oder Sozialwissenschaften. Durch den Fokus auf Community-spezifische Bedürfnisse – etwa in Bezug auf Datenbeschreibung, rechtliche Fragen und jene der

<sup>11</sup><https://www.go-fair.org/fair-principles/>.

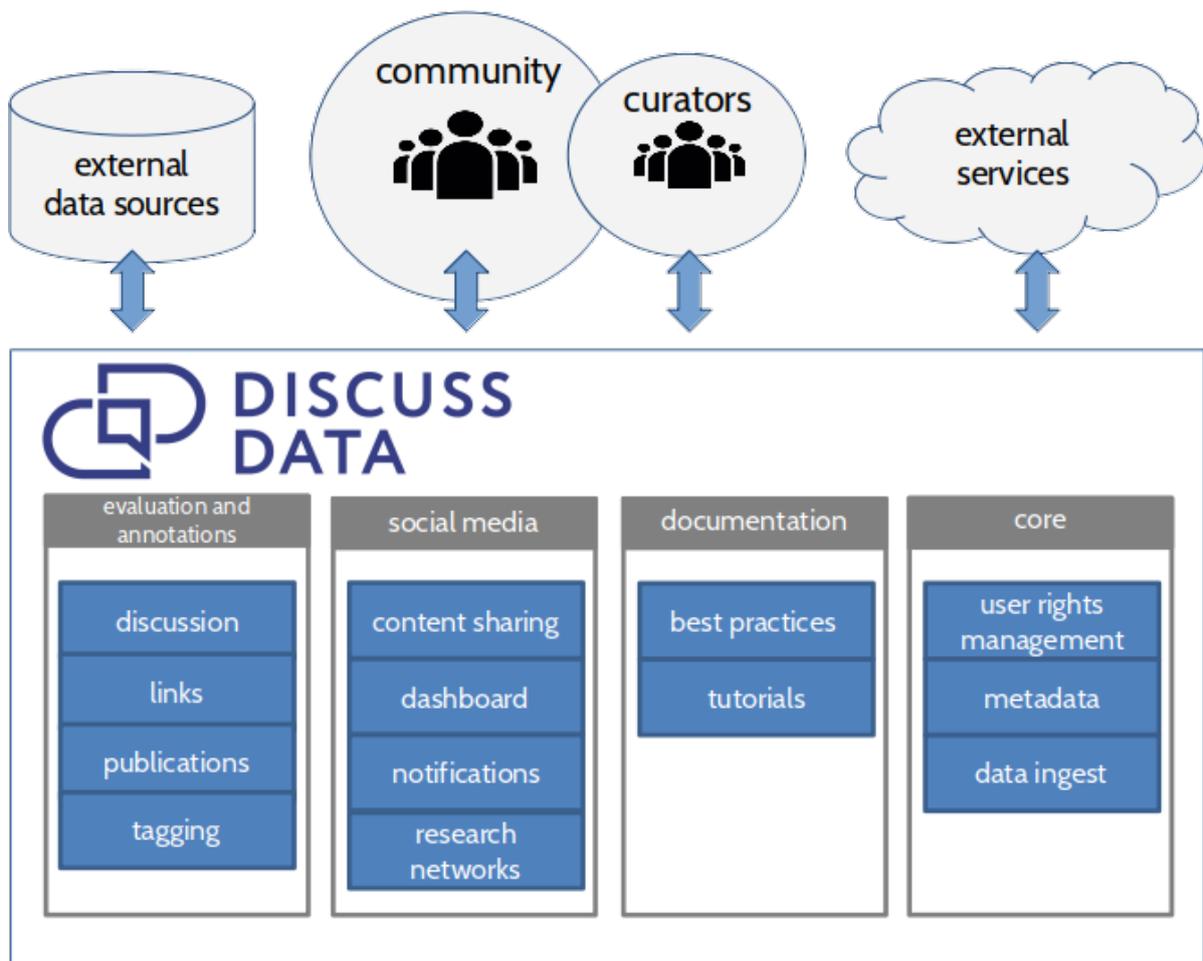


Abbildung 1: Struktur von Discuss Data.

Datensicherheit – die sich aus Überschneidungen in Methodik, Datenmaterial und Fragestellungen ergeben, lassen sich zielgenaue Angebote ableiten, die Forschenden zusätzliche Anreize bieten, FDM Plattformen zu nutzen und stärker als bisher in ihre Tätigkeit zu integrieren. Discuss Data wird mit den Kategorien Protesten, Korruption, Bildung und Sozialpolitik starten, Feldern, auf denen die Forschungsstelle Osteuropa gut vernetzt und damit in der Lage ist, die entsprechenden Sub-Communities gut zu erreichen. Die Einbindung weiterer Communities wird für die Etablierung der Plattform von existenzieller Bedeutung sein.

## 6 Social Media like User Experience

Das Design der Oberfläche ist in Discuss Data bewusst an eine von Social Media Plattformen bekannte User Experience angelehnt, um auch weniger Technik-affinen Nutzerinnen und Nutzern intuitiv zugänglich zu sein. Forschende können Nutzerprofile

erzeugen und sich über die auf Discuss Data bereitgestellten Daten mittels Kommentaren austauschen. Datasets und Profile können zur Beobachtung markiert werden und so neue Publikationen, Ergänzungen oder Kommentare verfolgt werden. All diese Funktionalitäten sind von anderen Netzwerken bekannt und bieten gerade deshalb im Kontext von digitalen Forschungsdaten einen bisher nicht gegebenen Mehrwert.

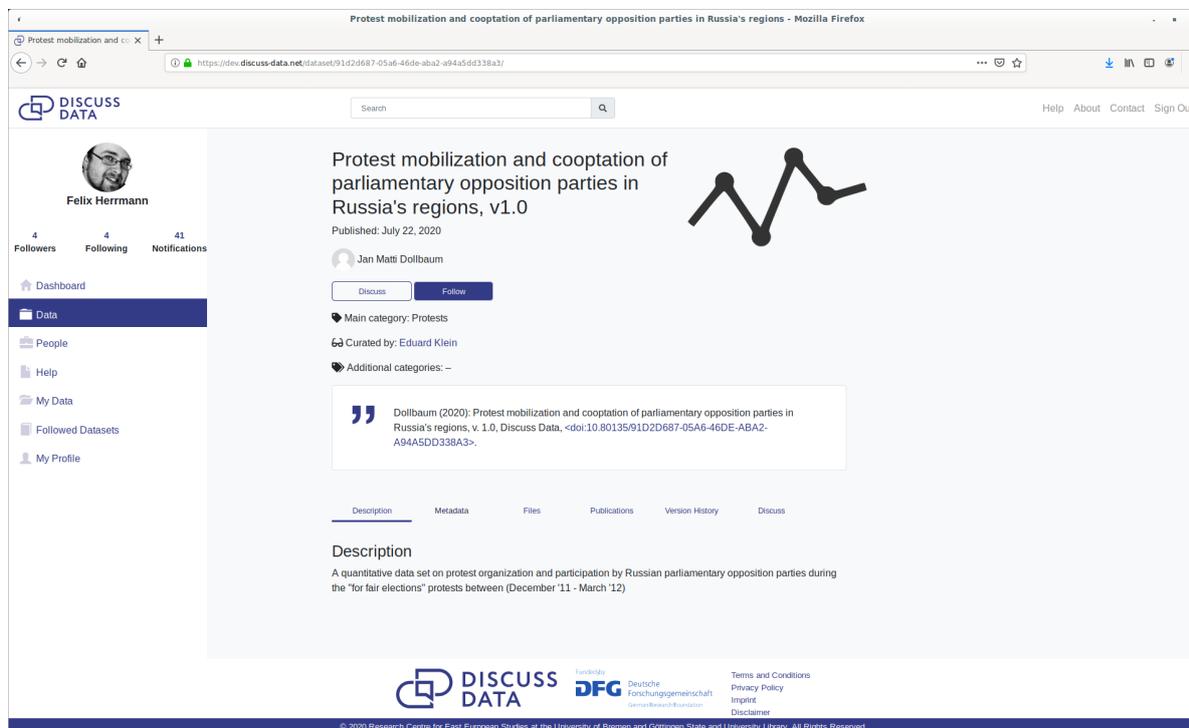


Abbildung 2: Datensatz in Discuss Data.

Das Rechtemanagement erfolgt per Data Set und erlaubt neben der Auswahl aus etablierten Datenlizenzen auch die Nutzung individuell erstellter oder externer Lizenzen. Sowohl im Vorbereitungsprozess als auch nach der Publikation können Leserechte übersichtlich durch Datengebende oder deren Bevollmächtigte verwaltet werden. Mit der Kommentarfunktion können Datensammlungen nicht-öffentlich vor der Publikation (etwa durch Arbeitsteams) oder im Anschluss daran öffentlich diskutiert werden. Abbildung 2 zeigt die Darstellung eines Datensatzes mit den grundlegenden Informationen sowie Links zu Metadaten, Dateien, der Kommentarfunktion etc. Auf der linken Seite sind die einzelnen Module zu sehen, welche es erlauben Daten zu recherchieren und hochzuladen, mit Forschenden in Kontakt zu treten und das eigene Profil zu bearbeiten. Alle auf Discuss Data publizierten Daten und Kommentare sind mittels Digital Object Identifier (DOI) referenzierbar. Dies ermöglicht es, dass Publikationen, welche auf anderen Plattformen veröffentlicht werden, auf die Daten und die dafür freigegebenen Kommentare dauerhaft verweisen können. Es sollen zudem Prozesse und Best Practices etabliert werden, welche schrittweise die inhaltliche und redaktionelle Verantwortung für Discuss Data ganz an die Community übertragen. Für Redaktion

und Moderation notwendige Funktionen lassen sich Usern zuweisen, die dadurch zu data curators einer bestimmten thematischen Kategorie werden. Diese stehen für Anfragen zur Verfügung und nehmen eine technische Überprüfung von zur Publikation eingereichten Datensammlungen vor.

## 7 Erste Erfahrungen

Die Erfahrungen aus der geschlossenen Beta-Phase lassen eine optimistische Prognose dahingehend zu, dass auch Fachwissenschaftler/innen ohne informationswissenschaftlichen Hintergrund dazu in der Lage sein werden, Daten zu erschließen und zu annotieren. Dazu sollen neben der Unterstützung durch die Kurator/innen insbesondere die Bereitstellung einer ausführlichen Anleitung zur Plattform (sowie deren Integration als interaktive Tooltips auf den entsprechenden Seiten) und detaillierte Best Practices zur Datenpublikation beitragen. Das Feedback der Nutzer/innen soll zudem in der Kategorie "Discuss Discuss Data", in persönlicher Kommunikation und einem öffentlichen Issue Tracker<sup>12</sup> gesammelt und bei der weiteren Entwicklung berücksichtigt werden.

Im Laufe der Entwicklung und der ersten Test-Phase ist deutlich geworden, dass sich die Gewinnung von Gutachtern für klassische Peer Review Verfahren<sup>13</sup> angesichts der damit verbundenen Arbeitsbelastung und dem bei einem neuen Projekt bislang nicht vorhandenen Renomé als schwierig gestalten wird. Stattdessen setzt Discuss Data auf die Begutachtung durch die Community insgesamt: Alle registrierten Nutzer/innen werden in die Lage versetzt, Kritik und Vorschläge so einzubringen wie etwa während eines Workshops. Vor allem die dauerhafte Referenzierbarkeit mittels DOIs wird dabei von Forschenden als motivierender Faktor zur persönlichen Mitarbeit genannt.

## 8 Fazit

Die Veränderungen im Zuge der Digitalisierung der Forschung stellen die Anbieter von Informationsinfrastrukturen vor neue Herausforderungen, doch bieten sich dabei auch neue Möglichkeiten. Discuss Data verfolgt einen neuen und offenen Ansatz für Forschungscommunities und setzt dabei auf bestehenden Basisinfrastrukturen auf. Als Mediator zwischen Repositorien, Datendiensten und den Forschenden ermöglicht Discuss Data letzteren einen einfachen und nachhaltigen Umgang mit digitalen Forschungsdaten und bezieht sie umfassend in das Forschungsdatenmanagement mit

<sup>12</sup><https://gitlab.gwdg.de/discuss-data/discuss-data/-/issues>.

<sup>13</sup>Zur Kritik an klassischen Peer Review Verfahren siehe: Fröhlich, Gerhard. „Anonyme Kritik: Peer Review auf dem Prüfstand der Wissenschaftsforschung“. In: *medizin-bibliothek-information* 2003, Vol.3 Nr. 2: S. 33-39, <https://media02.culturebase.org/data/docs-agmb/mbi-2003-2.pdf>.

ein. Dazu wird zunächst die Osteuropaforschung angesprochen und Discuss Data spezifisch auf die dort vorhandenen Bedürfnisse zugeschnitten. Es ist angedacht, diesen Ansatz weiter zu verbreitern und Discuss Data Instanzen auch für weitere Forschungsfelder bereit zu stellen. Ziel ist die Bildung eines übergreifenden Netzwerks von jeweils fachspezifischen Communities. Wir halten die enge Verbindung von Daten und Diskursen zu diesen für einen wichtigen Ansatz, um die Qualität und Transparenz der Forschung zu stärken. Die klare Ausrichtung auf FAIR Data soll der Nachprüfbarkeit, Nachnutzbarkeit und Zugänglichkeit von Forschungsergebnissen und -daten dienen. Nicht zuletzt soll so auch die Publikation von Daten als wissenschaftliche Leistung gestärkt werden.