

## **Öffentliche Konsultation des Bundesministeriums für Bildung und Forschung zum Forschungsdatengesetz**

### **Stellungnahme von Wikimedia Deutschland e. V.**

Verfasserin: Aline Blankertz, Referentin Politik und öffentlicher Sektor

#### **1. Bedarfe**

*Bitte erläutern Sie, welche Bedarfe an Daten im Kontext kommerzieller und nichtkommerzieller Forschung bestehen und gehen Sie hierbei auf folgende Aspekte ein:*

##### *a. Daten für die Forschung:*

*Zu welchen Daten besteht aus Ihrer Sicht ein Bedarf an besseren Zugangs-, Nutzungs- und Verknüpfungsmöglichkeiten z.B. lange Datenreihen, geographisch referenzierte Daten, gelabelte Daten, (nicht) personenbezogene Daten, kuratierte Daten, Rohdaten ohne Kontext, Spartendaten wie Klima, Wetter?*

Es besteht ein großer Bedarf an Daten, die von der öffentlichen Hand erhoben und aktuell nur unzureichend öffentlich verfügbar gemacht werden. Behörden erheben vielfach Daten z.B. in Form von Karten, über Mobilität, über meteorologische Verhältnisse, die prinzipiell frei verfügbar sein sollten, sowohl für die Forschung als auch für die Zivilgesellschaft und prinzipiell alle Arten von Zwecken, einschließlich wirtschaftlicher Aktivitäten.

Darüber hinaus kann Forschung auch aus aktuell privat gehaltenen Daten lernen: Das Argument, Unternehmen sollten Daten geheim halten dürfen, um z.B. Geschäftsgeheimnisse zu schützen, kommt bei vielen Formen der Forschung nicht zum Tragen, da diese in keinem wettbewerblichen Verhältnis zu Unternehmen stehen.

Es zeigt sich z.B. im Digital Services Act, dass ein großes gesellschaftliches Interesse daran besteht, die Rolle und Wirkung von wirkmächtigen Infrastrukturen von Onlinediensten zu untersuchen. Ähnliche Formen von Datenzugang, mindestens für Forschung und Zivilgesellschaft, können für ähnlich gelagerte Bereiche erwogen werden, wie z.B. weitere Onlinedienste, die aktuell nicht vom DSA erfasst werden, oder Organisationen, die an der Schnittstelle zur öffentlichen Daseinsvorsorge operieren (Mobilitätsanbieter, Energieanbieter etc.).

*In welchen wissenschaftlichen Disziplinen bzw. interdisziplinären Kontexten für welche Forschungszwecke sind Datenzugänge besonders relevant (z.B. empirische Wirtschaftsforschung, Bildungsforschung, medizinische Forschung, interdisziplinäre Forschung)?*

Um Innovationen in allen Disziplinen der Wissenschaft und durch die Zivilgesellschaft zu begünstigen, sollte anstelle von Priorisierung ein horizontaler Ansatz erwogen werden. So können konsistent mehr Daten für Forschung (durch Wissenschaft und Zivilgesellschaft) zugänglich werden.

Es liegt nahe, dass die Antworten auf diese Frage jeweils den Interessenbereich der antwortenden Organisationen widerspiegeln. Aus unserer Sicht ist es besonders wichtig, dass Forschung die Funktionsweise und die Auswirkungen von Organisationen, oft großen Plattformen, im Internet untersuchen kann. Nur so kann effektive und zielgerichtete Regulierung auf den Weg gebracht werden.

*Welche Herkunft/Quellen von Daten sind hier aus Ihrer Sicht besonders relevant (z.B. Daten aus der öffentlichen Verwaltung, Daten aus der Wirtschaft bzw. von Unternehmen, Forschungseinrichtungen oder Einzelpersonen, Maschinendaten, Daten der amtlichen Statistik, etc.)?*

Aus unserer Sicht ist die Veröffentlichung von Daten aus der öffentlichen Verwaltung besonders trivial, da bei ihnen offensichtlich ein gesellschaftliches Interesse an einer hohen Verfügbarkeit und kein Geheimhaltungsinteresse geltend gemacht werden kann. Wie oben erwähnt, gibt es zudem ein großes Interesse an Daten in Wirtschaftsbereichen, die von hohem öffentlichen Interesse sind, wie z.B. Infrastrukturen wie Energie oder Mobilität, unabhängig davon, ob diese Aktivitäten von staatlicher oder privater Seite erfolgen.

*b. Daten aus der Forschung:*

*Zu welchen Daten aus der Forschung besteht aus Ihrer Sicht Bedarf an Zugang-, Nutzungs- und Verknüpfungsmöglichkeiten für, öffentliche Stellen, Unternehmen und Zivilgesellschaft?*

Prinzipiell sollten alle Daten aus der Forschung, insbesondere wenn diese (auch teilweise) mit öffentlichen Mitteln gefördert wird, allen zur Verfügung stehen. Dies umfasst alle Daten, die zu Forschungszwecken erhoben werden - seien es Messdaten in Experimenten, Umfragedaten oder andere Methoden, die Daten für ein wissenschaftliches Erkenntnisinteresse generieren - sowie die mit diesen Daten erzielten Analyseergebnisse.

Eine breite Verfügbarkeit leitet sich daraus ab, dass ein gesellschaftliches Interesse an der Nachnutzung besteht, z.B. um die Ergebnisse zu replizieren, Metastudien zu erstellen oder weitere Analysen mit den Daten durchzuführen. Demgegenüber steht, wie bereits erwähnt, kein Anlass zur Geheimhaltung, da durch die öffentliche Finanzierung keine Ausschließlichkeitsrechte (die in der Privatwirtschaft teils zur Refinanzierung genutzt werden) nötig sind.

Bei sensiblen Forschungsdaten ist ein verantwortungsvoller und ethisch angemessener Umgang essentiell. Hier sind die FAIR-Prinzipien um Aspekte möglicher kollektiver Rechte beforschter Gruppen und Communities zu ergänzen, wie dies etwa in den CARE-Prinzipien der Fall ist.

*Zu welchen Zwecken sind Datenzugänge für Unternehmen, Zivilgesellschaft oder öffentlichen Stellen besonders relevant?*

Ebenso wenig wie die Breite der Forschung in einer abschließenden Nennung von Zwecken abgebildet werden kann, ist es schwierig, die Nachnutzungsmöglichkeiten von Daten umfassend abzubilden. Es besteht ein allgemeines gesellschaftliches Interesse daran, dass auch die interessierte Öffentlichkeit wissenschaftlichen Diskursen folgen und sie nachvollziehen kann, einschließlich aktuell nicht veröffentlichter und somit nicht verfügbarer "non-findings", in denen Studien nicht zu statistisch signifikanten Ergebnissen kommen (was aber mindestens für zukünftige Forschungsdesigns wertvolle Erkenntnisse sind).

Dazu müssen mindestens Daten, doch darüber hinaus auch Methoden und Erkenntnisse möglichst frei verfügbar sein. Das heißt, offene Forschungsdaten sollten in einen Kontext von Open Access und von Open Science eingebettet sein. Auf diesem Weg kann Forschung besser aufeinander aufbauen (durch das Teilen von Daten aktuell nicht veröffentlichter Forschung), transdisziplinäre Forschung vereinfachen (in denen Daten und Erkenntnisse übergreifend verfügbar sind) und Hürden in Richtung Gesellschaft und Politik abbauen (die sich aktuell keine teuren Zugänge zu Datenbanken etc. leisten können).

*Welche Herkunft/Quellen von Forschungsdaten sind hier aus Ihrer Sicht besonders relevant?*

Das freie Verfügbarmachen von Forschungsdaten ist bei öffentlich geförderter Forschung (auch bei Mischfinanzierung) in besonderem Maße geboten, da Steuergelder zur Generierung der Daten genutzt werden. Das umfasst zudem alle Forschungsinstitute, die eine Grundförderung von öffentlicher Seite erhalten, selbst wenn einzelne Projekte möglicherweise durch privatwirtschaftliche Gelder getragen werden.

*c. Was würde die Bereitschaft zum Datenteilen auf Seiten von*

- Unternehmen bzw. der Wirtschaft*
- Wissenschaft*
- Zivilgesellschaft*
- öffentlichen Stellen*

*erhöhen? Welche Anreize könnten gesetzt werden?*

Öffentliche Aufträge, insbesondere an die Forschung, sollten an die Bedingung geknüpft werden, dass die im Projekt anfallenden Daten frei verfügbar zu machen sind. In der Förderung selbst muss

dann gegebenenfalls ein leicht erhöhter Dokumentationsaufwand Berücksichtigung finden, denn damit offene Forschungsdaten einen Mehrwert stiften, müssen sie mit aussagekräftigen Metadaten und Dokumentation versehen sein.

*d. Wie schätzen sie die Bedeutung von Forschungsk Kooperationen (z.B. mit Partnern aus der Wissenschaft und der Wirtschaft) in Bezug auf das sektorübergreifende Teilen von Daten ein?*

Damit haben wir keine direkte Erfahrung.

*e. Welche Rolle und welche Aufgaben sollten Datenmittlerstrukturen beim Datenteilen einnehmen?*

Datenmittler können dabei helfen, Friktionen zwischen einzelnen datenteilenden Organisationen abzubauen, indem z.B. vertraulich zu behandelnde Daten verschlüsselt oder aggregiert werden. Es ist jedoch fraglich, inwiefern solche Strukturen nötig sind, wenn Forschungsdaten nicht nur für einen beschränkten Kreis, sondern für die allgemeine Öffentlichkeit nachnutzbar werden sollen. Viele Forschungsdaten sind nicht sensibel und sollten somit möglichst nah an der Datenquelle zur Verfügung stehen. Ein höherer Schutzbedarf besteht z.B. bei Gesundheitsforschungsdaten, sodass in diesem und ähnlichen Kontexten Datenmittler helfen können, Vertraulichkeit und Verfügbarkeit zusammenzubringen.

## 2. Hindernisse

*Bitte erläutern Sie, welche Hindernisse beim Zugang und im Umgang (Verknüpfung, Nutzung etc.) mit Daten für und aus der Forschung derzeit bestehen.*

*Bitte gliedern Sie der Übersichtlichkeit halber gern in z.B. rechtliche, technische, organisatorisch Hindernisse und erläutern Sie (wenn möglich) jeweils anhand von Beispielen.*

rechtlich:

- Eine wesentliche Hürde sind (urheber)rechtliche Einschränkungen, insbesondere wenn die Rechte an Forschungsergebnissen an geschlossene Formate, wie von den meisten Wissenschaftsverlagen verwendet, übergehen. Die Regelungen für Text und Data Mining greifen somit leider nur für Veröffentlichungen, die Open Access verfügbar sind.
- In manchen Bereichen, wie z.B. Gesundheit, gibt es eine überzeichnete Rolle des Datenschutzes. Statt komplexe Abwägungen umfassend auf Einwilligungen Einzelner abzuwälzen, sollte es klarere gesetzliche Vorgaben geben dafür, welche Daten für Forschung verwendet werden dürfen (und wozu sie nicht verwendet werden dürfen).

organisatorisch:

- der zusätzliche Aufwand, der damit einhergeht, andere Daten verfügbar zu machen (mindestens Dokumentation), lohnt sich oftmals nicht für Einzelne, auch wenn alle davon profitieren, wenn mehr Daten für und aus der Forschung zur Verfügung stehen.

### 3. Lösungsvorschläge

*Mit Blick auf die beschriebenen Bedarfe und Hindernisse, welche legislativen Regelungen sollten aus Ihrer Sicht für ein Forschungsdatengesetz geprüft werden (konkrete Formulierungsvorschläge erwünscht)?*

*a. (zu den nicht-legislativen Lösungen s. 3 c)*

*b. In welchen Gesetzen (deutschen/ggf. auch europäischen) besteht aus Ihrer Sicht Änderungs-/Ergänzungsbedarf mit Blick auf die oben beschriebenen Bedarfe und Hindernisse (Formulierungsvorschläge erwünscht)?*

Ein wichtiger Hebel sind die Bedingungen, die an die Vergabe öffentlicher Gelder (im vorliegenden Kontext für Forschung) geknüpft sind: Diese sollten dem Ansatz "öffentliches Geld - öffentliches Gut" folgen, sodass die im Kontext öffentlich geförderter Forschung erhobenen Daten frei verfügbar sind. Darüber hinaus könnte man erwägen, die vertragsrechtliche Einschränkung von privatwirtschaftlichen Forschungsdaten zu beschränken, da diese voraussichtlich von der besseren Verfügbarkeit öffentlich geförderter Forschungsdaten profitieren.

*c. Welche weiteren, nicht-legislativen Lösungsvorschläge sehen Sie mit Blick auf die beschriebenen Bedarfe und Hindernisse?*

Es bedarf einer breiten Umstellung weg von einer geschlossenen Wissenschaft, die Verlagen lukrative Geschäftsmodelle ermöglicht auf Kosten einer langsameren, intransparenteren und weniger zugänglichen Forschung, hin zu einer Offenen Wissenschaft. Die Ansätze von Open Science umfassen nicht nur Daten, sondern alle Ebenen von Wissenschaft: Theorien, Methoden und Ergebnisse, aber auch den Umgang mit Rohdaten und Zwischenergebnissen. Open Science macht diese (und auch die "non-findings") einer größeren Zahl von Menschen zugänglich – ohne die Öffnung von bestimmten Anwendungsfällen abhängig zu machen. Dabei ist Open Science kein Selbstzweck. Sie erleichtert den wissenschaftlichen Erkenntnisgewinn in einer Disziplin und den produktiven interdisziplinären Austausch. Zur Einbettung der bestehenden Open-Data- und Open-Access-Strategien ist es wünschenswert, auch eine übergeordnete Open-Science-Strategie auf nationaler Ebene zu entwickeln.



#### 4. Sonstiges

*Was möchten Sie uns sonst noch zum Thema Forschungsdatengesetz/Forschungsklauseln mitgeben?*

Open Science ist keine Randbewegung mehr. Der Rat der EU befürwortet eine stärkere Ausrichtung an Open Science und betont, dass bestehende Anreizsysteme in der Forschung hierfür angepasst werden müssen. Die Deutsche Forschungsgemeinschaft hat kürzlich eine überwiegend positive Positionierung zu Open Access oder Open Code veröffentlicht. Die Voraussetzung für Open Science sind offene Daten, offene Software und ggf. offene Hardware. Daher sollte ein größerer Fokus darauf liegen, diese Themen übergreifend auszubauen.