

# Öffentliche Konsultation zum Forschungsdatengesetz - Fragebogen

Frank Löffler (frank.loeffler@uni-jena.de)  
Harald von Waldow (harald.vonwaldow@thuenen.de)  
Michael Schlottke-Lakemper (m.schlottke-lakemper@acom.rwth-aachen.de)

2023-04-11

## Vorwort

Das folgende Dokument besteht aus dem vom Bundesministerium für Bildung und Forschung am 10.03.2023 veröffentlichten Fragebogen zur öffentlichen Konsultation zum Forschungsdatengesetz und den von der Gesellschaft für Forschungssoftware in Deutschland - de-RSE e.V. am 11.4.2023 eingereichten Antworten.

Lediglich das Format wurde geändert, weil die vorgegebene Datei im pdf-Format teilweise nicht genügend Platz für Antworten bereitstellte. Da die Antworten im Markdownformat von der Gemeinschaft der RSEs eingesammelt wurden, liegen sie nun sowohl in Markdown als auch als PDF-Datei vor.

## Interessengruppe

Bitte ordnen Sie sich einer Gruppe zu.

- ☒ Wissenschaft
- ☐ Unternehmen (groß, KMU)
- ☐ Zivilgesellschaft
- ☐ Öffentliche Stelle

## Verfasserin oder Verfasser (Institution/Person)

Gesellschaft für Forschungssoftware - de-RSE e.V.

## 1. Bedarfe

Bitte erläutern Sie, welche Bedarfe an Daten im Kontext kommerzieller und nicht-kommerzieller 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?

Software, die zur Datenverarbeitung (u.a. Extraktion, Umwandlung, Analyse, Visualisierung) sonstiger Daten im Einflussbereich des Forschungsdatengesetzes notwendig ist oder üblicherweise verwendet wird. Grund ist, dass in der Forschung die Nachvollziehbarkeit, Reproduzierbarkeit und Provenienz von Daten essentiell sind. Diese Informationen werden vollständig und zuverlässig nur durch die begleitende Veröffentlichung des Quellcodes der verwendeten Software transportiert und nur durch für Nachnutzung geeignete Lizenzen leicht nutzbar. Lücken in der durch Quellcode dokumentierten Datenverarbeitung, z.B. das Fehlen oder die anderweitige Nichtbenutzbarkeit der nötiger Software, gefährden den wissenschaftlichen Wert von Forschungsdaten in fundamentaler Weise.

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

Die quelloffene Verfügbarkeit der zur Datenerzeugung, -bearbeitung und -analyse verwendeten Software ist für alle wissenschaftlichen Disziplinen von sehr großer Relevanz.

- 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.)?

Die quelloffene Verfügbarkeit der zur Datenerzeugung, -bearbeitung und -analyse verwendeten Software ist für alle Datenquellen von sehr großer Relevanz.

- **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?

Datensätze, die wissenschaftliche Ergebnisse stützen und in wissenschaftlichen Publikationen interpretiert werden, sind in der Regel das Produkt einer Reihe von softwaregestützten Reinigungs-, Auswahl- und Analyseschritten.

Derartige Daten bedürfen deshalb in besonderem Maße der Dokumentation durch die zur Verarbeitung benutzte quelloffene Software, um Transparenz und korrekte Interpretation durch öffentliche Stellen, Unternehmen und die Zivilgesellschaft zu gewährleisten.

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

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

Daten aus groß angelegten Projekten, die eine möglichst große Untermenge ihrer Dimensionen möglichst granular abbilden, z.B. langjährige Zeitreihen deutschlandweiter Monitoringprogramme, können von vielen Akteuren in vielfältiger Weise nachgenutzt werden und haben dadurch potentiell einen besonders großen gesellschaftlichen Wert.

Voraussetzung für die Inwertsetzung solcher Daten ist die Dokumentation der Datenaufbereitung durch begleitend veröffentlichte quelloffene Software.

- c. Was würde die **Bereitschaft zum Datenteilen** auf Seiten von
  - Unternehmen bzw. der Wirtschaft
  - Wissenschaft
  - Zivilgesellschaft
  - öffentlichen Stellen erhöhen? Welche Anreize könnten gesetzt werden?

Insbesondere in der Wissenschaft gibt es außerhalb von Verbundprojekten bisher wenig Anreize zum Teilen von Daten oder Software. So ist es für eine Karriere in der Wissenschaft (z.B. bei Berufungen) immer noch oft unerheblich, ob Daten/Software geteilt bzw. sinnvoll aufbereitet zu Verfügung gestellt wurden. Eine mögliche Maßnahme wäre daher, bei öffentlich geförderten Ausschreibungen die Anträge von daten-/softwarelastigen Projekten nicht nur nach den wissenschaftlichen Vorarbeiten in Form von Veröffentlichungen, sondern auch auf Basis einer dokumentierten Historie von geteilten Daten/Software zu bewerten.

- 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?

Für die Weiterentwicklung der Gesellschaft ist das Anwenden von Forschungsergebnissen auch außerhalb der Wissenschaft enorm von Vorteil. Dabei ist sowohl das Anwenden von neuen Forschungsmethoden auf Daten anderer Sektoren gemeint (u.a. durch Nachnutzung von Software), als auch die Nutzung entstandener Forschungsdaten in anderen Sektoren.

Sektorübergreifende Forschungs Kooperationen können also sowohl für die Wissenschaft als auch für andere Sektoren von großem Nutzen sein.

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

Eine Rolle könnten Datenmittler beim kontrollierten Einhalten des Datenschutzes dienen und hier z.B. entsprechend der Anfrage anonymisierten Zugriff auf Teile von Datensätzen erlauben. Abgesehen vom Datenschutz könnten Datenmittlerstrukturen bei der Etablierung und Einhaltung von Datenaustauschstandards behilflich sein. Dies muss auch die Koordination mit entsprechenden internationalen Standards einschließen, weil nationale Grenzen in der Wissenschaft oft keine große Bedeutung haben.

## 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.

Rechtliche Hindernisse: - Daten werden oft entweder ohne oder mit ungeeigneten Lizenzen zur Verfügung gestellt und sind so für die Forschung nicht verwendbar. Dies gilt unabhängig vom Sektor, der Daten zur Verfügung stellt. - Software, die zur Aufbereitung von publizierten Daten benutzt wurde, ist oft nicht quelloffen. Wenn der Quellcode nicht eingesehen werden kann, ist die Interpretation der publizierten Daten erschwert und die Korrektheit kann nicht, oder nur mit hohem Aufwand, überprüft werden.

Technische Hindernisse:

Daten werden oft in einer Form angeboten, deren Interoperabilität und Maschinenlesbarkeit nicht dem Stand der Technik entsprechen.

Sowohl Daten für die Forschung, als auch Daten aus der Forschung sollten

- in offenen, wohldokumentierten, robusten und verbreiteten Formaten angeboten werden und
- wenn Nachfrage und Updatezyklen das nahelegen, über wohldokumentierte webbasierte Schnittstellen ("APIs") angeboten werden.

## 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)

- Daten aus dem öffentlichen Sektor sollen analog zu Beispielen anderer Länder grundsätzlich frei und unentgeltlich verfügbar zu machen sein (mit nötigen Einschränkungen wie dem Datenschutz) und mit allgemein kompatibler Lizenz zu versehen sein, wie zum Beispiel der Lizenz “CC0”.
  - Software aus dem öffentlichen Sektor soll grundsätzlich unter einer anerkannten Open-Source-Lizenz veröffentlicht werden.
  - Die durch öffentliche Stellen zur Erzeugung, Bearbeitung und Bereitstellung von Daten verwendete Software soll grundsätzlich quelloffen sein.
- **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)?

- **c.** Welche weiteren, nicht-legislativen Lösungsvorschläge sehen Sie mit Blick auf die beschriebenen Bedarfe und Hindernisse?
- Die Etablierung einer anerkannten Plattform für die Publikation quelloffener Software von Seiten der öffentlichen Stellen würde ein aktuelles Hindernis aus dem Weg räumen. <https://opencode.de> ist dabei ein Anfang, aber wichtig ist ein gelebter Wechsel zur quelloffenen Software als Normalfall.

#### 4. Sonstiges

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