VolkswagenStiftung

Leitfaden für den Umgang mit Forschungsdaten

Orientierungshilfe für Antragsteller:innen, Gutachter:innen und Projektdurchführende

Forschungsdaten sind heute mehr denn je Grundlage und Ergebnis wissenschaftlichen Arbeitens. Die Vielfalt solcher Daten – Rohdaten wie Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Umfragedaten etc. sowie Sekundärdaten wie Softwareergebnisse, Simulationen, Annotationen etc. – entspricht der Vielfalt unterschiedlicher Disziplinen, Verfahren und Erkenntnisinteressen. Die Sicherung von Forschungsdaten ist dabei sowohl für die Reproduzierbarkeit von Untersuchungen als auch als Ausgangspunkt für neue Forschungsfragen wichtig. Die VolkswagenStiftung unterstützt die nachhaltige Speicherung von und den möglichst offenen Zugang zu Forschungsdaten im Sinne der FAIR-Prinzipien (Findable – Accessible – Interoperable – Reusable). Eine strukturierte Beschäftigung mit Forschungsdatenmanagement im Vorfeld eines Forschungsprojektes erhöht die Datenqualität, verbessert die Datensicherheit und steigert die Zeit- und Ressourceneffizienz des gesamten Projektes.

Dieser Leitfaden bietet Orientierung für die Beschreibung und Bewertung des Umgangs mit Forschungsdaten in einem bei der VolkswagenStiftung beantragten Projekt. Er basiert auf den Leitlinien für Forschungsdatenmanagement von Science Europe.

Es wird empfohlen, bereits in der Planungsphase eines Projekts einen Datenmanagementplan (DMP) anzulegen. Er hilft, Arbeiten und Aufwände realistisch zu planen, und sollte im Sinne eines „Living Document“ während der Projektlaufzeit bei Bedarf aktualisiert werden (siehe Abschnitt 2). Abschnitt 1 behandelt nur den Teil eines DMP, der im Antrag beschrieben und in der Begutachtung berücksichtigt wird.

1 Beschreibung und Bewertung des Umgangs mit Forschungsdaten im Projektantrag

Die folgenden Abschnitte entsprechen den Überschriften des Antragskapitels zum Datenmanagement und zeigen die Hinweise für Antragsteller:innen und die Bewertungshilfen für Gutachter:innen nebeneinander.

1.1 Datenbeschreibung

Fragen zum Datenmanagement und Hinweise für Antragsteller:innen		**Bewertungshilfe**	
		Ausreichend berücksichtigt: Im Kapitel zum Datenmanagement wird/sind…	**Unzureichend berücksichtigt: Im Kapitel zum Datenmanagement wird/sind…**
Welche für das Projekt relevanten Daten existieren bereits? Welche davon sollen genutzt werden (Herkunft, Nutzungsbedingungen)? Begründen Sie, falls vorhandene Daten nicht genutzt werden können.	• Recherchieren Sie, ob für das Projekt nutzbare Daten existieren. • Wenn Sie Daten nachnutzen, geben Sie die Herkunft der Daten, Zugriffsmöglichkeit und die Lizenz bzw. sonstige Nutzungsbedingungen an. • Falls die Nutzung bestehender Daten nicht in Betracht kommt, begründen Sie dies kurz.	• skizziert, welche relevanten Daten existieren. • klar angegeben, woher die vorhandenen Daten stammen. • angegeben, wenn vorhandene Daten genutzt werden, wie auf diese Daten zugegriffen wird und welche Einschränkungen für ihre Nutzung bestehen. • klar erläutert, falls zutreffend, warum neue Daten erhoben werden müssen.	• keine Angabe, ob relevante Daten für das Projekt zur Nachnutzung zur Verfügung stehen. • nur wenige oder keine Angaben dazu enthalten, welche Daten genutzt werden und woher die Daten stammen. • falls zutreffend, keine ausreichende Begründung für die Erhebung oder Erzeugung neuer Daten geliefert.
Welche Daten (inkl. Datentypen und -formate) sollen im Projekt erhoben oder erzeugt werden? Welches Datenvolumen ist zu erwarten?	• Geben Sie an, welche Methoden oder Software verwendet werden, um neue Daten zu erheben oder zu erzeugen. • Geben Sie Einzelheiten zum Typ der Daten an: z. B. numerisch (Datenbanken, Tabellen), textuell (Dokumente), Bild, Audio, Video und benennen Sie die Dateiformate, z. B. pdf, csv, txt, xlsx, docx. • Bevorzugen Sie offene und standardisierte Formate, da diese den Austausch und die langfristige Nachnutzung von Daten erleichtern, und begründen Sie die Verwendung abweichender Formate. • Machen Sie Angaben zum Umfang in erforderlichem Speicherplatz (Bytes) und/oder in der Anzahl der Objekte, Dateien, Zeilen und Spalten.	• klare Angaben dazu enthalten, wie neue Daten erzeugt werden. Die verwendeten Methoden und Software werden klar benannt. • klar beschrieben oder aufgelistet, welche Datentypen erzeugt werden und welche Dateiformate damit verbunden sind. • gut begründet, warum abweichende Dateiformate gewählt werden, falls es sich nicht um offene und standardisierte Formate handelt. • das geschätzte Datenvolumen angegeben. • ggf. klar angegeben, dass im Rahmen des Projekts keine neuen Daten erzeugt werden.	• nur wenige oder keine Angaben dazu enthalten, welche Daten erhoben werden. • die im Projekt erzeugten Datentypen nicht oder nur unzureichend beschrieben. • Datentypen gelistet/beschrieben, ohne deren Formate anzugeben. • Dateiformate aufgelistet, ohne die Datentypen anzugeben. • keine Schätzung des Datenvolumens angegeben.

1.2 Dokumentation, Recht und Ethik

Fragen zum Datenmanagement und Hinweise für Antragsteller:innen		**Bewertungshilfe**	
		Ausreichend berücksichtigt: Im Kapitel zum Datenmanagement wird/sind…	**Unzureichend berücksichtigt: Im Kapitel zum Datenmanagement wird/sind…**
Welche Dokumentation und Metadaten werden für die erhobenen oder erzeugten Daten erstellt? Wie werden dabei fachspezifische Standards berücksichtigt?	• Überlegen Sie, welche Dokumentation für die Nachnutzung erforderlich sind. Dazu können Informationen über die Datenerhebung, Analyse, Definitionen von Variablen, Maßeinheiten usw. gehören. • Überlegen Sie, wie diese Informationen erfasst und wo sie gespeichert werden sollen (z. B. Readme-Datei, Codebook). • Geben Sie an, welche Metadaten (ggf. in welchem Standard, Schema) erstellt werden, um die Daten später auffindbar zu machen. • Gehen Sie auf fachspezifische Praktiken und ggf. Standards bei der Organisation und Ablage der Daten ein. Bei interdisziplinären Projekten können unterschiedliche Standards für die jeweiligen Datentypen genannt werden.	• die Dokumentation klar beschrieben, die für die Nachnutzung der Daten erforderlich ist, und angegeben, wo die Informationen gespeichert werden (z. B. Readme-Datei, Codebook, Laborbuch, Datenbank mit Verweisen zu den einzelnen Elementen). • die Metadaten klar skizziert, unter Bezugnahme auf bewährte Verfahren (z. B. Metadatenstandards).	• keine Informationen oder nur sehr vage Hinweise auf eine Dokumentation enthalten, ohne Erläuterungen zu liefern. • nur wenige oder gar keine Informationen zu den Metadaten angegeben, die die Daten begleiten werden.
Welche ethischen oder rechtlichen Aspekte sind bei der Erhebung und Verarbeitung der Daten zu berücksichtigen? Wie wird ggf. die Vertraulichkeit personenbezogener oder anderweitig sensibler Daten gewährleistet?	• Gehen Sie auf fachspezifische Praktiken und ggf. Standards bei der Erhebung personenbezogener oder anderer sensibler Daten ein (z. B. informierte Einwilligung, Schutz der Anonymität). • Stellen Sie dar, mit welchen technischen und organisatorischen Maßnahmen Sie sensible Daten (z. B. personenbezogene Daten oder sonstige sensible Informationen) während und nach dem Projekt schützen. • Beachten Sie Rechte Dritter und Löschfristen. • Geben Sie an, ob für Ihr Projekt ein Votum einer Ethikkommission erforderlich ist und wie dieses ggf. eingeholt wird.	• nachvollziehbar dargelegt, wie mit personenbezogenen und anderen sensiblen Daten verantwortungsvoll umgegangen wird. • Schutzmaßnahmen, wie Zugriffs- und Nutzungsverwaltung und/oder gesondert gesicherte Bereiche auf eigener Infrastruktur, klar beschrieben. • angegeben, ob ein Votum einer Ethikkommission erforderlich ist und wie es ggf. eingeholt wird.	• keine Informationen dazu enthalten, ob ethisch sensible Daten erhoben und verarbeitet werden. • keine oder unzureichende Informationen zum Schutz enthalten, obwohl absehbar sensible Daten erhoben und/oder verarbeitet werden. • keine oder unzureichende Informationen zur Notwendigkeit eines Ethikvotums enthalten.

1.3 Archivierung und Publikation

Fragen zum Datenmanagement und Hinweise für Antragsteller:innen		**Bewertungshilfe**	
		Ausreichend berücksichtigt: Im Kapitel zum Datenmanagement wird/sind…	**Unzureichend berücksichtigt: Im Kapitel zum Datenmanagement wird/sind…**
Wo und wie lange sollen die im Projekt erhobenen oder erzeugten Daten und Metadaten nach Projektende archiviert werden?	• Skizzieren Sie den Plan für die Datenarchivierung im Sinne der guten wissenschaftlichen Praxis (eigene Infrastruktur, Archiv, Repositorium o.Ä.) und geben Sie an, wie lange die Daten archiviert werden sollen. • Nennen Sie ggf. abweichende Archivierungsfristen für personenbezogene und sensible Daten.	• Informationen darüber enthalten, wie lange die Daten archiviert werden (i.d.R. 10 Jahre). • Ort der Archivierung (eigene Infrastruktur, Archiv, Repositorium) genannt, ggf. mit dem Namen des Dienstes, in dem die Daten bzw. Metadaten archiviert sein werden.	• nur wenige oder keine Angaben, wie, wann, wo und wie lange die Daten archiviert werden.
Ab wann sollen die Daten der Wissenschaft unter welcher Lizenz zur Verfügung stehen? In welchem Repositorium sollen die Daten veröffentlicht werden? Begründen Sie ggf. Einschränkungen oder den Verzicht auf eine Veröffentlichung.	• Erläutern Sie, wann die Daten verfügbar gemacht werden. • Nennen Sie die geplante Lizenz, z. B. CC BY. • Erläutern Sie, wie die Daten auffindbar und zugänglich werden (z. B. durch Speicherung in einem vertrauenswürdigen Repositorium, Indizierung in einem Datenkatalog oder einen anderen Mechanismus). • Bevorzugen Sie ein disziplinspezifisches Fachrepositorium. Wenn kein Fachrepositorium genutzt wird, erläutern Sie, warum nicht. • Achten Sie bei der Wahl des Repositoriums darauf, dass persistente Identifikatoren (PIDs) vergeben werden, mit denen die Daten nachhaltig referenzierbar und zitierfähig sind. • Falls Sie die Daten nicht veröffentlichen oder nur eingeschränkt bzw. erst nach einer Sperrfrist (Embargo) zugänglich machen wollen, erläutern Sie die Gründe (z. B. um Veröffentlichungen zu ermöglichen, sensible Daten zu schützen oder Patente anzumelden) und den vorgesehenen Zeitraum.	• angegeben, wann Daten zugänglich gemacht werden (voraussichtliches Veröffentlichungsdatum) und unter welcher Lizenz. • der Name des Repositoriums und/oder des Datenkatalogs genannt, in dem die Daten voraussichtlich gespeichert oder verzeichnet werden sollen. • erläutert, ob und wie persistente Identifikatoren erlangt werden. In der Regel vergeben anerkannte Repositorien PIDs. • ggf. klar erläutert, warum die Veröffentlichung eingeschränkt oder nicht möglich ist und wer unter welchen Bedingungen auf die Daten zugreifen kann. • erläutert, soweit möglich, welche Maßnahmen ergriffen werden, um Beschränkungen der Veröffentlichung zu überwinden oder zu minimieren.	• macht keine oder nur wenige Angaben wann Daten zugänglich gemacht werden, oder die Erläuterungen sind unzureichend oder technisch nicht realisierbar. • nur wenige oder keine Angaben enthalten, in welchem Repositorium die Daten gespeichert werden. • keine Angabe, ob die Daten PIDs erhalten (und welche), und kein triftiger Grund für deren Nichtverwendung genannt. • keine klaren Informationen enthalten, ob einzelne Datensätze und/oder Sammlungen mit PIDs versehen werden.

2 Empfehlungen zum Umgang mit Forschungsdaten im Projektverlauf

Ein gutes Forschungsdatenmanagement ist im gesamten Projektverlauf eine wichtige Grundlage für den Projekterfolg und für die Nachnutzbarkeit der Daten. Es wird empfohlen, bereits in der Planungsphase einen vollständigen Datenmanagementplan (DMP) anzulegen und ihn während der Projektlaufzeit bei Bedarf zu aktualisieren. Das Antragskapitel zum Datenmanagement greift nur einen Ausschnitt daraus auf. Die folgende Übersicht enthält die Fragen, die in einem vollständigen DMP adressiert werden sollten:

- **Datenbeschreibung:**
  Welche bestehenden Daten oder Sammlungen werden nachgenutzt?
  Welche Daten (Typen, Formate, Volumen) werden im Projekt erhoben oder erzeugt, und wie werden sie weiterverarbeitet?
- **Dokumentation und Datenqualität:**
  Welche Dokumentation und Metadaten werden nach welchen Standards erstellt?
  Wie werden die Daten organisiert (z. B. Ordnerstruktur, Versionsverwaltung), und wie wird ihre Qualität gesichert?
  Welche Software ist zur Nutzung erforderlich?
- **Speicherung und Sicherung während des Projekts:**
  Wo werden die Daten gespeichert, wie werden sie gesichert und wiederhergestellt?
  Wer hat Zugriff, und wie werden sensible Daten geschützt?
- **Rechtliche und ethische Aspekte:**
  Welche datenschutzrechtlichen und ethischen Anforderungen gelten?
  Wie sind Urheber-, Persönlichkeits-, Verwertungs- und Eigentumsrechte geregelt, auch an Daten Dritter und zwischen Projektpartner:innen?
- **Archivierung und dauerhafte Zugänglichkeit:**
  Wo und wie lange werden die Daten archiviert?
  Welche Daten werden ab wann, unter welcher Lizenz und in welchem Repositorium mit persistentem Identifikator (PID) zugänglich gemacht?
- **Verantwortlichkeiten und Ressourcen:**
  Wer ist während und nach dem Projekt für die Daten und die Aktualisierung des DMP zuständig?
  Welche Ressourcen (z. B. Personalzeit, Speicher, Repositoriumskosten) werden benötigt?

Für geförderte Projekte gilt zudem:

Die Stiftung erwartet, dass die Nutzung der im Projekt erzeugten oder verwendeten Daten sowie die Dateneigentums-, Zugangs- und Verwertungsrechte während und nach der Projektlaufzeit zwischen den Projektpartner:innen, z. B. in einer Vereinbarung, geregelt sind.