



# RDMO – Fragenkatalog für Text+

Version 1.1, 23.10.2024

Die folgenden Fragen im Abschnitt "Allgemein" dienen der generellen Projektbeschreibung, alle weiteren Abschnitte beziehen sich auf die Beschreibung der konkreten Datensätze.

## **Gliederung:**

- Allgemein
- Inhaltliche Einordnung
- Technische Einordnung
- Metadaten und Referenzierung
- Rechtliche und ethische Fragen
- Datennutzung und Veröffentlichung
- Langzeitarchivierung

# Allgemein

## Thema

Wie lautet die primäre Forschungsfrage des Projektes?

Bitte beschreiben Sie kurz das Projekt und dessen Ziele.

Bitte geben Sie einige Schlagworte zur Forschungsfrage an.

Sofern möglich, identifizieren Sie die Schlagwörter mithilfe eines standardisierten Vokabulars, wie beispielsweise der [GND](#) oder [Wikidata](#), um eine eindeutige Referenzierbarkeit zu gewährleisten. Beispiel: Schlagwort: Sophocles, Vokabular: GND, ID: 118615688

Bitte geben Sie ein Schlagwort an: \_\_\_\_\_

Auf welchem Normdatensatz basiert das vergebene Schlagwort?

GND (Gemeinsame Normdatei)

Wikidata

Sonstige: \_\_\_\_\_

## Disziplin

Welcher Disziplin / welchen Disziplinen ist das Projekt zuzuordnen?

Bitte wählen Sie aus den Hauptebenen der [DFG Fachsystematik](#) eine passende Kategorie.

## Projektablauf

Wann beginnt die Projektlaufzeit?

Wann endet die Projektlaufzeit?

# Förderung

## Wer fördert das Projekt?

Geben Sie bitte, wenn vorhanden, auch das entsprechende Förderkennzeichen an.

- Deutsche Forschungsgemeinschaft (DFG): \_\_\_\_\_
- Europäische Kommission (EC) - Horizon Europe: \_\_\_\_\_
- Bundesministerium für Bildung und Forschung (BMBF): \_\_\_\_\_
- Volkswagen-Stiftung: \_\_\_\_\_
- Österreichischer Wissenschaftsfonds (FWF): \_\_\_\_\_
- Schweizerischer Nationalfonds (SNF): \_\_\_\_\_
- Andere: \_\_\_\_\_

**Gibt es von Seiten des Forschungsförderers Vorgaben oder Richtlinien bezüglich des Umgangs mit den im Projekt erhobenen Forschungsdaten? Wenn ja, skizzieren Sie diese kurz und verweisen Sie ggf. auf weiterführende Informationen. Geben Sie bitte auch an, welchen Grad an Verbindlichkeit diese haben.**

„Grundsätzlich müssen in allen Förderprogrammen, in denen ein konkretes Arbeitsprogramm für ein Forschungsvorhaben beschrieben wird, Ausführungen zum Umgang mit Forschungsdaten beschrieben werden.“

[https://www.dfg.de/foerderung/grundlagen\\_rahmenbedingungen/forschungsdaten/forschungsfoerderung/index.html](https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/forschungsfoerderung/index.html)

Beispiele für Anforderungen an das Datenmanagement seitens der Forschungsförderer:  
Deutsche Forschungsgemeinschaft (DFG):

- [Leitlinien zum Umgang mit Forschungsdaten \(2015\)](#)
- [Umgang mit Forschungsdaten. Checkliste für Antragstellende zur Planung und zur Beschreibung des Umgangs mit Forschungsdaten in Forschungsvorhaben \(2021\)](#)

Europäische Kommission (EC):

- [Guidelines on FAIR Data Management in Horizon 2020](#)

## Zusätzliche Projektdaten

Unter welcher URL sind Informationen zum Projekt abrufbar?

## Weitere Anforderungen

Gibt es von weiteren Seiten (z.B. von der Fachcommunity) Anforderungen an das Datenmanagement, die beachtet werden müssen?

- Ja
- Nein
- Noch zu klären

Wenn ja, welche Anforderungen an das Datenmanagement sind dies?

Bitte verweisen Sie ggf. auf weiterführende Informationen. Geben Sie bitte auch an, welchen Grad an Verbindlichkeit diese haben. Beispiele für fachspezifische Empfehlungen und Richtlinien aus den **Geistes- und Sozialwissenschaften** sind:

- DFG-Fachkollegium 104 ("Sprachwissenschaften"). Empfehlungen zu datentechnischen Standards und Tools bei der Erhebung von Sprachkorpora (2019)
- DFG-Fachkollegium "Literaturwissenschaften". Förderkriterien für wissenschaftliche Editionen in der Literaturwissenschaft (2015)
- DFG-Fachkollegium 106 ("Sozial- und Kulturanthropologie, Außereuropäische Kulturen, Judaistik und Religionswissenschaft") zum Umgang mit Forschungsdaten (2019)
- DFG-Fachkollegium 111 ("Sozialwissenschaften") zum Umgang mit Forschungsdaten in der Soziologie, der Politikwissenschaft und der Kommunikationswissenschaft (2020)
- DFG-Fachkollegium 112 ("Wirtschaftswissenschaften") zum Umgang mit Forschungsdaten (2019)
- DFG-Fachkollegium "Alte Kulturen" zum Umgang mit Forschungsdaten (2020)
- DFG-Fachkollegium "Erziehungswissenschaften" zum Umgang mit Forschungsdaten (2020)
- DFG Fachkollegium "Psychologie" zum Umgang mit Forschungsdaten (2020)
- DFG-Fachkollegium "Theologie". Handreichung zu Editionsprojekten (2022)
- Andere: \_\_\_\_\_

Die vollständige Liste finden Sie auf den Seiten der DFG [hier](#).

## Projektkoordination

Welches Institut / welche Institute sind für das Projekt verantwortlich?

Welche Personen (Antragsteller/-in / Sprecher/-in / Projektleitung) sind verantwortlich für die Projektkoordination?

Bitte für jede Person Vorname, Nachname, E-Mailadresse und Identifikator ([ORCID-Suche](#)) angeben, falls vorhanden.

## Projektpartner

Bitte geben Sie den ausführlichen Namen des/der institutionellen Projektpartner(s) an.

### Welche Person ist bei diesem Partner beteiligt bzw. ist Ansprechpartner/in für das Datenmanagement?

Bitte für jede Person Vorname, Nachname, E-Mailadresse und Identifikator ([ORCID-Suche](#)) angeben, falls vorhanden.

## Inhaltliche Einordnung

### Datensätze

Für die langfristige Bereitstellung und Archivierung von nachnutzbaren text- und sprachbasierten Forschungsdaten bietet Text+ eine entsprechende Infrastruktur, an der verschiedene Daten- und Kompetenzzentren mit ihrer je eigenen Expertise zu bestimmten Datentypen ([Collections](#), [Editionen](#), [Lexikalische Ressourcen](#)) beteiligt sind.

Durch die Aufnahme von Daten in eine der an Text+ beteiligten Einrichtungen wird eine nachhaltige Datenhaltung (gemäß FAIR/CARE) sichergestellt und die Nachnutzbarkeit sowie die Auffindbarkeit über zentrale Nachweissysteme erhöht (siehe auch die Abschnitte „Datennutzung“ und „Langzeitarchivierung“).

Die nächsten Fragen dienen zur Beschreibung der Datensätze, die im Projekt erzeugt und/oder verwendet werden. Sie helfen zudem, den Wert der Daten hinsichtlich der potentiellen Nachnutzung und einer späteren Archivierung einschätzen zu können.

Vor der Erzeugung von Daten empfiehlt es sich, zu prüfen, ob bereits vorhandene Daten nachgenutzt werden können. Die Vermeidung doppelter Erhebungen spart Aufwand und Kosten. Handelt es sich um personenbezogene Daten, gelten die Grundsätze der EU-Datenschutzgrundverordnung ([DSGVO](#)) wie z. B. Datenminimierung ([Art.5 Abs.1](#)) und das Bundesdatenschutzgesetz ([BDSG](#)).

Die Definition dessen, was jeweils ein Datensatz ist, ist eine wichtige konzeptionelle Entscheidung, die für jedes Vorhaben bzw. Projekt individuell getroffen und sorgfältig abgewogen werden muss. Ein Datensatz kann beispielsweise aus mehreren zusammen gruppierten Dateien verschiedener Arten (numerisch, bild- oder textbasiert) bestehen, die alle aus einer Versuchsreihe auf einem definierten Versuchsobjekt entstandene Ergebnisse enthalten.

### Welche Datentypen erwarten Sie?

Bitte wählen Sie eine Kategorie.

## Datentyp

- Tabellarische Daten (z.B. Messreihen)
- Text (z.B. digitale Edition, Interview-Transkription)
- Datenbank
- Audiodatei(en) (z.B. Interview, Sprachaufzeichnung)
- Visuelle Datei(en) (z.B. Film, Fotografie, Abbildung)
- Geografische Information
- Software (z.B. Eigenentwicklung innerhalb des Projekts)
- 3D-Modellierung (z.B. digitale Rekonstruktion einer steinzeitlichen Siedlung)
- Quantitative Online-Befragung
- Sonstiges: \_\_\_\_\_

**Geben Sie, wenn möglich, an, zu welcher der drei Text+ Datendomänen der jeweilige Datensatz zuzuordnen ist.**

- Collections (individuelle Textkorpora, Textannotationen, Interviews, Transkriptionen, Sensordaten, Umfragen, usw.)
- Editionen (OCR-Dateien, Handschriften und deren Transkriptionen, Textapparate, usw.)
- Lexikalische Ressourcen (Übersetzungen, Terminologien, Wortnetze, Wortlisten, usw.)

## Datenursprung

**Wird der Datensatz selbst erzeugt oder nachgenutzt?**

Wenn der Datensatz nachgenutzt wird, beachten Sie bitte auch die Hinweise unter „Urheber- oder verwandte Schutzrechte“.

- Erzeugt
- Nachgenutzt

**(a) Wenn nachgenutzt, wer ist Urheber des Datensatzes?**

Benennen Sie bitte die betreffenden Personen bzw. Institutionen.

**(b) Wenn nachgenutzt, unter welcher Adresse, PID oder URL ist der Datensatz verfügbar?**

**(c) Wenn nachgenutzt, welcher Lizenz (z.B. Creative Commons) bzw. welchen Nutzungsbedingungen (Urheber-, Patent- oder andere Eigentumsrechte) unterliegt die Nachnutzung des Datensatzes?**

(Hinweis: Wenn die Lizenz/Nachnutzungsbedingungen noch nicht klar sind: Welche Schritte planen Sie zur Klärung zu unternehmen?)

## Welche (analogen) Quellen liegen dem Datensatz zugrunde?

(z.B. Manuskript, Sprachatlas, archivalische Quelle; sofern vorhanden, bitte die Archivsignatur angeben! (siehe hierzu beispielsweise [Bundesarchiv: Quellennachweise](#))

## Datenerhebung

### Wann beginnt/begann die Erhebung bzw. Erstellung der Daten?

### Wann endet/endete die Erhebung bzw. Erstellung der Daten?

### Wie wird der Datensatz erzeugt?

Bitte beschreiben Sie, mit welcher Methode die Daten erhoben oder erstellt werden. Auch Zusatzinformationen, wie z.B., ob Audioaufzeichnungen grammatikalisch oder phonetisch analysiert werden, sind dabei zu berücksichtigen. Bitte erläutern Sie ergänzend, ob es sich um einen festen, zu einem bestimmten Zeitpunkt gespeicherten Datensatz handelt oder ob dieser dynamisch (also über einen bestimmten Zeitraum laufend) entsteht.

## Technische Einordnung

### Formate

#### In welchen Formaten werden die Daten vorliegen?

Bei der Wahl des Dateiformats sollten auch die Konsequenzen für die kollaborative Nutzung, die Langzeitarchivierung sowie die Nachnutzung beachtet werden. Es empfiehlt sich, möglichst standardisierte, nicht-proprietäre, verlustfreie und allgemein bzw. in der spezifischen Community verbreitete Formate zu nutzen. Eine gute Übersicht hierzu liefert [forschungsdaten.info](#) oder [WissGrid-Leitfaden](#), S. 22 f.

- Tagged Image File Format (TIFF, TIF) für Bilder
- Waveform Audio File Format (WAV) für Audiodateien
- Plain Text Document (TXT, ASC) für Dokumente
- Portable Document Format/A (PDF/A) für Dokumente

- Extensible Markup Language (XML) für Dokumente
- Comma-Separated Values (CSV) für tabellarische Daten
- JavaScript Object Notation (JSON) für Datenaustausch
- Sonstige: \_\_\_\_\_

## Datengröße

### Was ist die tatsächliche oder erwartete Größe des Datensatzes?

Wenn große Datenmengen anfallen, sollten finanzielle Mittel zur Bereitstellung der Infrastruktur berücksichtigt werden (siehe auch Abschnitt „Förderung“).

- weniger als 1 GB
- 1 GB bis 10 GB
- 10 GB bis 100 GB
- zwischen 100 GB und 1 TB
- zwischen 1 TB und 10 TB
- > 10 TB
- Genau: \_\_\_\_\_
- Noch nicht bestimmt

## Werkzeuge

### Welche Software, Technologien oder Verfahren werden zur Erzeugung oder Erfassung der Daten genutzt?

Diese Informationen sind für alle Arten von Daten relevant, um ihre Genese (z.B. via MAXQDA, Excel, MySQL-Datenbank, R, QGIS, Gephi, usw.) nachvollziehen zu können. Bei reproduzierbaren Daten kommt ein weiterer Aspekt hinzu. Diese müssen nicht notwendigerweise aufbewahrt werden - allerdings müssen alle Geräte, Software, Softwareversionen und auch Informationen über die Vorgehensweise erhalten bleiben, die notwendig sind, um die Daten erneut erstellen zu können.

## Metadaten und Referenzierung

### Metadaten

Welche Normdatenvokabulare werden zur eindeutigen Identifikation von Entitäten innerhalb des Datensatzes verwendet?

(z.B. [GND](#) für Personen, [Wikidata](#) für Objekte und Konzepte, [GeoNames](#) für Geographika)

### Welche Standards, Ontologien, Klassifikationen usw. werden zur Beschreibung der Daten und Kontextinformation genutzt?

(z.B. [CIDOC CRM](#), [Europeana Data Model](#), [TEI-XML](#), [MEI-XML](#))

Bedenken Sie auch bereits an dieser Stelle, dass fachspezifische Repositorien zur langfristigen Speicherung und Veröffentlichung der Daten (siehe Abschnitt „Weitergabe und Veröffentlichung“) jeweils eigene Anforderungen an das Metadatenschema haben.

- Es werden disziplinspezifische Standards, Klassifikationen, usw. genutzt.
- Es wird ein eigenes Beschreibungssystem genutzt (bitte beschreiben Sie dieses kurz und geben, wenn nötig, an, wo es ausführlicher dokumentiert ist):

Sonstiges: \_\_\_\_\_

- Es wurde noch nicht entschieden, mit welchem System die Metadaten und Kontextinformationen beschrieben werden.
- Es wird kein festgelegtes System zur Beschreibung genutzt.

### Wie werden die Metadaten erhoben?

- Automatische Erfassung (z.B. Zeitstempel eines Interviews, Aufnahmedatum eines Fotos):

- Semi-automatische Erfassung (Metadaten, die durch den Computer/die Software automatisch erhoben werden, jedoch überprüft werden müssen, z.B. HTR/OCR Verfahren mit Named-Entity-Recognition): \_\_\_\_\_
- Manuelle Erfassung (z.B. Orts-, Datumsangaben oder Personendaten, die manuell aus historischen Dokumenten extrahiert werden) Dafür bietet sich ein Metadateneditor an.):

### Persistente Identifikatoren (PIDs)

Persistente Identifikatoren (PIDs) sollen die dauerhafte Referenzierung von (insbesondere) digitalen Objekten wie Publikationen oder Forschungsdaten ermöglichen. Statt, wie i.d.R. bei der Angabe eines Hyperlinks als Referenz der Fall, auf den Speicherort des Objektes zu verweisen, fungiert die PID als eine Zwischeninstanz, von der aus zum Objekt weitergeleitet wird (dies nennt man "Auflösen" der PID). Die PID bleibt gleich, auch wenn sich der Speicherort des Objektes ändert. Während ein Hyperlink in diesem Fall ins Nichts führen würde, ist das Objekt über die PID weiterhin erreichbar. Weitere Informationen zur Funktionsweise, Verwendung und den verschiedenen Arten von PIDs findet man z.B. im [Handbook der Digital Preservation Coalition](#) oder auf der Website des [PID Network Deutschland](#) bzw. von [PID4NFDI](#).

### Sollen für diesen Datensatz persistente Identifikatoren (PIDs) genutzt werden?

- Ja
- Nein

**Welches System von persistenten Identifikatoren soll genutzt werden?**

- Handle / DOI
- PURL
- ARK
- URN
- ISLRN
- Anderes: \_\_\_\_\_
- Commit-Nummer auf einem Code-Repositorium: \_\_\_\_\_
- Keins

**Welche Untereinheiten / Teildatensätze sollten sinnvollerweise eigene Identifikatoren erhalten? Welche dieser Identifikatoren sollten dauerhaft und zitierfähig sein?**

**Wer ist verantwortlich für die Pflege der PIDs und die Objektpflege (d.h. die Langzeitarchivierung des Objekts und somit dafür, dem PID-Service einen Objektumzug und die neue Adresse mitzuteilen)?**

Die Voraussetzung dafür, dass persistente Identifikatoren auch wirklich halten, was sie versprechen, ist, dass sie - wie auch die Objekte, auf die referenziert wird - dauerhaft und verlässlich gepflegt werden. D. h. beispielsweise, dass bei einer Änderung des Speicherortes des Objekts die Information über den neuen Speicherort entsprechend aktualisiert werden muss.

Werden die Daten in einem Datenzentrum oder -repositorium gespeichert, wird diese Verantwortung i.d.R. von diesem übernommen. Um sicher zu sein, sollten die Verantwortlichkeiten jedoch in jedem Fall geprüft bzw. geklärt werden.

## Rechtliche und ethische Fragen

### Recht allgemein

Einen allgemeinen Überblick zu rechtlichen Aspekten von Forschungsdaten bietet die Website [forschungsdaten.info](http://forschungsdaten.info). Auch der [Text+ Helpdesk](#) informiert zu rechtlichen und ethischen Belangen in der text- und sprachbasierten Forschung, wie z.B. dem Urheberrecht, bietet allerdings keine Rechtsberatung an.

**Muss die rechtliche Situation verschiedener Länder berücksichtigt werden?**

Falls Sie diese Frage mit "Ja" beantworten, setzen Sie sich bitte mit der Rechtsabteilung bzw. einem/r entsprechenden Ansprechpartner/in an Ihrer Institution in Verbindung, um zu klären, ob sich daraus Konsequenzen für Ihr Projekt ergeben und wenn ja, welche dies sind.

- Ja
- Nein

## Personenbezogene Daten

**Enthält dieser Datensatz personenbezogene Daten?**

- Ja
- Nein

Wenn ja, setzen Sie sich bitte mit der Rechtsabteilung bzw. einem/r entsprechenden Ansprechpartner/in an Ihrer Institution in Verbindung, um zu klären, ob sich daraus Konsequenzen für Ihr Projekt ergeben und wenn ja, welche dies sind.

Bitte prüfen Sie auch, ob Sie die Abschnitte "Datennutzung" und "Datenspeicherung und -sicherheit" mit den entsprechenden Maßnahmen zum Schutz der Daten gemäß den geltenden Datenschutzgesetzen ausgefüllt haben.

Grundsätzlich müssen bei personenbezogenen Daten folgende Punkte dokumentiert werden:

- **Zweck der Verarbeitung / Einwilligungsmanagement**
- **Befristung der Aufbewahrung**
- **Rechtsgrundlage**

Der Umgang und die Verarbeitung personenbezogener Daten ist gesetzlich geregelt. Die EU-weit einheitliche Anwendung erfolgt nach der [EU-Datenschutz-Grundverordnung \(DSGVO\)](#). Sie gestattet Regelungsspielräume auf nationaler Ebene. In Deutschland regelt dies das [Bundesdatenschutzgesetz \(BDSG\)](#). Für Hochschulen gelten zum großen Teil die individuellen Datenschutzgesetze der Länder, z. B. das [Datenschutzgesetz Nordrhein-Westfalen \(DSG NRW\)](#). Die europäische DSGVO definiert personenbezogene Daten als alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen (Art. 4 DSGVO, Abs. 1). Identifiziert ist eine Person, wenn eindeutig erkennbar ist, zu welcher Person die Daten gehören. Identifizierbar wird eine Person, wenn sie mittels Zusatzinformationen identifiziert werden kann. Mehr Informationen zum Thema finden Sie u.a. auf der Website von [forschungsdaten.info](http://forschungsdaten.info).

## Datenschutz

**Welche Gesetze sind neben der europäischen DSGVO bezüglich der Fragen des Datenschutzes für das Projekt zu beachten?**

Die [EU Datenschutzgrundverordnung \(DSGVO\)](#) hat Vorrang vor nationalem Recht. Sie findet seit dem 25.05.2018 Anwendung. Für Hochschulen finden weiterhin die Landesdatenschutzgesetze Anwendung, die bestimmte Regelungen der DSGVO ausgestalten können, z.B. [Datenschutzgesetz Nordrhein-Westfalen \(DSG NRW\)](#).

In bestimmten Bereichen gelten spezifische Gesetze, die über den Landesdatenschutzgesetzen bzw. [Bundesdatenschutzgesetz \(BDSG\)](#) stehen.

- Für Daten des Mikrozensus gilt das Bundesstatistikgesetz (BStatG).
- Für Forschung in Schulen in NRW gilt das Schulgesetz NRW (SchulG NRW)

Demzufolge können je nach Forschungsprojekt auch mehrere Gesetze betroffen sein.

- Bundesdatenschutzgesetz (BDSG)
- Landesdatenschutzgesetz Baden-Württemberg
- Landesdatenschutzgesetz Bayern
- Landesdatenschutzgesetz Berlin
- Landesdatenschutzgesetz Bremen
- Landesdatenschutzgesetz Brandenburg
- Landesdatenschutzgesetz Hamburg
- Landesdatenschutzgesetz Mecklenburg-Vorpommern
- Landesdatenschutzgesetz Hessen
- Landesdatenschutzgesetz Nordrhein-Westfalen
- Landesdatenschutzgesetz Rheinland-Pfalz
- Landesdatenschutzgesetz Niedersachsen
- Landesdatenschutzgesetz Saarland
- Landesdatenschutzgesetz Sachsen
- Landesdatenschutzgesetz Sachsen-Anhalt
- Landesdatenschutzgesetz Schleswig-Holstein
- Landesdatenschutzgesetz Thüringen
- Sozialgesetzbuch X (z.B. für medizinische Daten)
- Bundesstatistikgesetz (z.B. für Zensusdaten)
- Sonstiges: \_\_\_\_\_

## Sensible Daten

### Enthält der Datensatz personenbezogene Daten besonderer Kategorien?

Hierunter fallen laut DSGVO personenbezogene "Daten, aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person" ([Art. 9 DSGVO, Abs. 1](#)). Solche Daten gelten als besonders sensibel und erfordern noch strengere Schutzmaßnahmen als bei personenbezogenen Daten generell bereits erforderlich ist.

Wenn Sie diese Frage mit "Ja" beantworten, informieren Sie sich bei der/dem Datenschutzbeauftragten Ihrer Institution, welche zusätzlichen Schutzmaßnahmen notwendig sind.

- Ja
- Nein

### Werden die Daten anonymisiert oder pseudonymisiert?

- Ja, während der Erhebung
- Ja, vor / zu Beginn der Datenanalyse
- Ja, nach der Datenanalyse / vor der Publikation
- Nein

### In welchem Umfang wird die "informierte Einwilligung" der Betroffenen eingeholt?

Grundsätzlich gilt, dass eine Erhebung, Verarbeitung, Archivierung und Veröffentlichung personenbezogener Daten nur zulässig ist, wenn eine entsprechende "informierte Einwilligung" der Betroffenen vorliegt. Bitte informieren Sie sich ggf. bei der/dem Datenschutzbeauftragten Ihrer Institution über die notwendigen Bedingungen für die Einwilligung.

- Zur Analyse/Nutzung der Daten im Rahmen des Projektes sowie zur Nachnutzung
- Nur zur Analyse/Nutzung der Daten im Rahmen des Projektes
- Zur Analyse/Nutzung der Daten im Rahmen des Projektes sowie zur langfristigen Aufbewahrung

### Wo und wie sind die "informierten Einwilligungen" abgelegt?

### Löschmanagement: Bis wann werden die (unanonymisierten bzw. unpseudonymisierten) Originaldaten spätestens sicher vernichtet? Wer ist dafür zuständig?

## Weitere sensible Daten

### Enthält dieser Datensatz nicht-personenbezogene sensible Daten?

Beispiele hierfür sind etwa Daten, die Betriebs- oder Geschäftsgeheimnisse oder Ortsangaben zu bedrohten Tier- oder Pflanzenarten enthalten.

- Ja
- Nein

### Wenn ja, um welche nicht personenbezogenen sensiblen Daten handelt es sich und welche Schutzmaßnahmen werden dafür getroffen?

## Offizielle Genehmigung

Bestimmte Bereiche der Forschung erfordern zum Schutz von Menschen, Tieren und der Umwelt eine Genehmigung. Eine Zusammenstellung von Behördenkontakten und Informationen zum Thema Forschung findet sich z.B. [hier](#).

### Wurde das Forschungsvorhaben von einer Ethikkommission begutachtet?

- Ja, positiv begutachtet von folgender Kommission: \_\_\_\_\_
- Ja, begutachtet mit Auflagen, die folgendermaßen erfüllt werden: \_\_\_\_\_
- Noch nicht, es befindet sich aber bereits in Begutachtung.
- Noch nicht, es wird zur Begutachtung eingereicht bis spätestens: \_\_\_\_\_
- Nein, eine Begutachtung ist nicht notwendig, weil: \_\_\_\_\_

### Wird für das Forschungsvorhaben eine offizielle Genehmigung benötigt?

- Nein
- Ja. Die Genehmigung liegt bereits vor.
- Ja. Die Genehmigung wurde beantragt am/im: \_\_\_\_\_
- Ja. Die Genehmigung wird beantragt bis spätestens: \_\_\_\_\_

### Wenn ja, um welche Genehmigung handelt es sich?

### Wenn ja, welche ist die ausstellende Behörde?

Wird eine Datenzugangs-Kommission benötigt, die bei Zugriffsanfragen auf vom Projekt veröffentlichte Daten entscheidet, ob Zugang gewährt wird oder nicht?

- Ja
- Nein

## Urheber- oder verwandte Schutzrechte I

Werden Daten genutzt und/oder erstellt, die durch Urheber- oder verwandte Schutzrechte geschützt sind?

Nehmen Sie in diesem Fall frühzeitig Kontakt mit entsprechenden Beratungsstellen (z.B. der eigenen Rechtsabteilung und/oder Beratungsstelle von Bibliotheken u.a. im eigenen Haus) auf.

Daten oder Software können Urheber- oder anderen Schutzrechten unterliegen. Die Rechtslage kann selbst in der EU von Land zu Land erheblich abweichen. In Deutschland sind nach dem Urheberrechtsgesetz (UrhG) u.a. Werke der Literatur, Wissenschaft und Kunst, die eine

„persönliche geistige Schöpfung“ darstellen, urheberrechtlich geschützt. Der urheberrechtliche Schutz erlischt 70 Jahre nach dem Tod der bzw. des Urheberin/s.

Reine Daten, z. B. Messdaten oder Surveydaten, aber auch Metadaten (bis auf ggf. „beschreibende“ Metadaten) sind hingegen nicht schutzfähig.

In § 2 nennt das UrhG folgende geschützte Werkarten, wobei die Aufzählung nicht abschließend ist:

- Sprachwerke, wie Schriftwerke, Reden und Computerprogramme
- Werke der Musik
- pantomimische Werke einschließlich Werke der Tanzkunst
- Werke der bildenden Künste einschließlich der Werke der Baukunst und der angewandten Kunst und Entwürfe solcher Werke
- Lichtbildwerke einschließlich der Werke, die ähnlich wie Lichtbildwerke geschaffen werden
- Darstellungen wissenschaftlicher oder technischer Art wie Zeichnungen, Pläne, Karten, Skizzen, Tabellen und plastische Darstellungen.

Nach § 3 sind auch „Übersetzungen und andere Bearbeitungen von Werken geschützt, die persönliche geistige Schöpfungen des Bearbeiters sind“.

Schließlich sind nach § 4 auch Sammelwerke und Datenbankwerke geschützt, was im Bereich Forschungsdaten durchaus relevant sein kann. Sammelwerke werden dabei definiert als „Sammlungen von Werken, Daten oder anderen unabhängigen Elementen, die aufgrund der Auswahl oder Anordnung der Elemente eine persönliche geistige Schöpfung sind“.

Bei einem „Datenbankwerk im Sinne des Gesetzes“ handelt es sich um ein „Sammelwerk, dessen Elemente systematisch oder methodisch angeordnet und einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sind“.

Weitere relevante Schutzrechte können gewerbliche Schutzrechte wie Patente, Gebrauchsmuster, Sortenschutz [bei Pflanzenzüchtungen], Halbleiterschutz, Marken, geographische Herkunftsangaben, eingetragene Designs oder geschäftliche Bezeichnungen sein.

- Ja
- Nein
- Unbekannt

## Urheber- oder verwandte Schutzrechte II

### Be- oder entstehen an diesem Datensatz Urheberrechte?

Beachten Sie hierbei auch die Hinweise unter den Abschnitten "Inhaltliche Einordnung/Datenursprung" sowie "Datennutzung und Veröffentlichung/Weitergabe und Veröffentlichung".

- Werk der Literatur, Wissenschaft oder Kunst
- Übersetzung oder andere Bearbeitung eines Werkes
- Sammelwerk oder Datenbankwerk
- Andere: \_\_\_\_\_

Nein

**Wurde der/die Rechteinhaber/in recherchiert?**

Ja

Nein

## Datennutzung und Veröffentlichung

### Datenspeicherung und -sicherheit

**Wer darf während der Projektlaufzeit auf den Datensatz zugreifen?**

Projektmitglieder

Projektpartner

Nur Interne

Externe Partner

**Welche Maßnahmen zur Gewährleistung der Datensicherheit werden getroffen?**

Schutz vor unbefugtem Zugriff auf den Institutsserver

Datenwiederherstellung / Backup

**Wer ist verantwortlich für die Erstellung der Backups?**

Die Frage bezieht sich auf Backups während der Zeit, in denen mit den Daten gearbeitet wird. Fragen der Langzeitarchivierung werden gesondert im entsprechenden Abschnitt behandelt.

### Weitergabe und Veröffentlichung

**Sollen die Daten, die im Projekt entstehen, veröffentlicht oder geteilt werden?**

Ja, extern für alle

Ja, extern in begrenztem Umfang mit individueller Freigabe

Nein

**Wenn nicht: Begründen Sie dies bitte und unterscheiden Sie dabei zwischen rechtlichen und/oder vertraglichen Gründen und freiwilligen Einschränkungen.**

**Wenn nicht: Können die Projektmitarbeiter nach Projektende noch weiter auf die Daten zugreifen (wenn sie die Institution verlassen), wenn ja über welche Regelung?**

**Unter welchen Nutzungsbedingungen bzw. welcher Lizenz werden die Daten veröffentlicht?**

Die Auswahlmöglichkeiten orientieren sich an Lizenzen der [Creative-Commons-Familie](#). Mehrfachauswahl in folgender Auflistung sinnvoll, z.B. BY SA)

- Namensnennung (BY)
- Keine kommerzielle Nutzung (NC)
- Keine Bearbeitungen (ND)
- Weitergabe unter gleichen Bedingungen (SA)
- Gemeinfrei (CC0)
- MIT-Lizenz für Software
- Andere, bzw. genaue Bezeichnung - falls bekannt: \_\_\_\_\_

**Wo werden die Daten (einschließlich Metadaten, Dokumentation und ggf. relevantem Code bzw. relevanter Software) veröffentlicht?**

Nennen Sie die Datenbank/das Repositorium, ggf. mit Link.

Bei der Entscheidung für ein geeignetes Repositorium zur Veröffentlichung der Daten sind folgende Merkmale von zentraler Bedeutung:

- Öffentlich und entgeltfrei zugängliche Infrastruktur
- Dauerhafte Bestandsgarantie des Angebotes
- Möglichkeit zur offenen Lizenzierung
- Eindeutige Referenzierung der Ressourcen mittels Persistenter Identifikatoren (PID)
- Unterstützung einschlägiger Metadatenschemata
- Einbindung in übergeordnete Suchinfrastrukturen

Eine Auswahl an Repositoriums-Registern finden Sie hier: [Re3data](#), [DataCite](#), [RIsources \(DFG\)](#), [OpenDOAR](#).

Beispiele für generische Repositorien sind u.a.

- [Zenodo](#) (bietet u.a. eine Verbindung zu github und ist in Communitys organisiert, z.B. Text+)
- [DARIAH-DE Repository](#) (bietet durch den [DARIAH-DE Publikator](#) einen eigenen Bereich für die Verwaltung, Organisation und Veröffentlichung der Daten als Kollektionen)
- [RADAR4Culture](#) (Formatunabhängiges Repositorium für die Fächer Architektur-, Kunst- und Musik-, Theater-, Tanz-, Film- und Medienwissenschaften)

Beispiele für disziplinspezifische Repositorien sind u.a.

- [TextGrid Repository](#) (Fokus auf Textdaten in TEI XML mit literaturwissenschaftlichem Hintergrund)
- [CLARIN-D Zentren](#) (akzeptiert werden linguistische Text- und Sprachdaten auf Anfrage; alle CLARIN-D Zentren sind auch Teil von Text+; erster Kontakt: [Text+ Helpdesk](#))

Insbesondere die geisteswissenschaftlichen NFDI-Konsortien führen aktuelle kuratierte Listen von Repositorien oder bieten selbst passende FD-Repositorien zur freien und kostenlosen Nutzung für ihre Partner an (vgl. dazu auch diese [Übersicht von Text+](#) sowie die [kuratierte Liste von 4Culture](#)).

Der [Text+ Helpdesk](#) bietet allen Forschenden eine spezifische Beratung an und wählt gemeinsam mit Ihnen entlang der festgelegten Kriterien ein geeignetes Repository aus. Siehe hierzu auch

- [Hug et al. \(2023\): Wohin damit? Storing and reusing my language data](#)
- [Draxler et al. \(2023a\): Wohin damit? So kommen Ihre Forschungsdaten in die Text+ Infrastruktur](#)
- [Draxler et al. \(2023b\): Wohin damit? Wir geben Ihren Daten ein Zuhause](#)
- [Leitlinie für das Integrieren von Daten in Text+/NFDI](#)

---

## Langzeitarchivierung

### Speicherung

**Wo werden die Daten (einschließlich Metadaten, Dokumentation und ggf. relevanter Software) nach Projektende gespeichert bzw. archiviert?**

Anders als Repositorien, die ihre Daten für die sofortige Nutzung bereithalten, dient ein Langzeitarchiv der langfristigen Aufbewahrung von Daten, die möglicherweise nicht häufig abgerufen oder aktiv genutzt werden. Eine Lösung für Langzeitarchivierung mit Bitstream-Preservation wird derzeit beispielsweise im Rahmen des NFDI-Konsortiums Text+ entwickelt, siehe hierzu auch [Dogaru, G. \(2023\). Das Text+ Langzeitarchiv.](#)

- Eigene Institution
- Partnerinstitution: \_\_\_\_\_
- Sonstiges: \_\_\_\_\_
- Wird noch entschieden