Open Data

# Informationen

## Was sind „Open Data“?

Open Data oder offene Daten sind Daten, *die von jedermann frei benutzt, weiterverwendet und geteilt werden können - die einzige Einschränkung betrifft die Verpflichtung zur Nennung des Urhebers/der Urheberin[[1]](#footnote-1).* Folgende drei Punkte gelten hierbei als Kernpunkte:

* Verfügbarkeit und freier Zugang
* Möglichkeit zur Wiederverwendung und Weitergabe
* Universelle Beteiligung

Soweit es möglich ist, sollten sowohl die Rohdaten, als auch die Datensätze, auf deren Basis die finalen Analysen durchgeführt wurden, auf einem sicheren Repositorium hochgeladen und öffentlich verfügbar gemacht werden (über einen Link im Manuskript). Wenn die Daten oder Teile der Daten nicht offen zugänglich gemacht werden können, kann dies im Manuskript erwähnt und erläutert werden. Vor der Veröffentlichung der Daten muss eine Anonymisierung der Daten vorgenommen werden, die eine Rückführbarkeit von Daten auf Einzelpersonen nicht möglich macht.

## Publikationen

Hartig, K. & Soßna, V. (2016). Forschungsdatenmanagement in DFG-Anträgen: Was kann, was soll, was muss beschrieben werden*? Institutionelles Repositorium der Leibniz Universität Hannover*. http://dx.doi.org/10.15488/262

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Poline, J. B., Breeze, J. L., Ghosh, S. S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., & Ashburner, J. (2012). Data sharing in neuroimaging research. *Frontiers in neuroinformatics*, *6*, 9.

[Schönbrodt, F., Gollwitzer, M. & Abele-Brehm, A. (2017). Der Umgang mit Forschungsdaten im Fach Psychologie: Konkretisierung der DFG-Leitlinien. *Psychologische Rundschau,* *68*, 20–35.](http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/empfehlungen_forschungsdaten_psychologie.pdf)

Spindler, G., & Hillegeist, T. (2011). Rechtliche Probleme der elektronischen Langzeitarchivierung von Forschungsdaten. In S. Büttner, H.-C. Hobohm & L. Müller (Hrsg.), *Handbuch Forschungsdatenmanagement* (S. 63-69). Bad Honnef: Bock und Herchen Verlag.

Stodden, V. (2011). Trust your science? Open your data and code. *Amstat News*, 21-22.

## Links zu weiterführenden Informationen

Datenmanagement und Data Sharing in der Psychologie: Einführung und Manual: <https://www.psychdata.de/downloads/PsychData-Handbuch_2013.pdf>

Data Wiz - Ein Assistenzsystem für Forschungsdatenmanagement:

<https://datawiz.zpid.de/>

ICPSR. (n.d.). Guide to social science data preparation and archiving: Phase 3: Data collection and file creation. Retrieved from <http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter3quant.html>

Infos der zu Datenmanagementplänen: <https://www.uni-muenster.de/Forschungsdaten/organisieren/datenmanagementplan/> und <http://www.forschungsdaten.org/index.php/Data_Management_Pl%C3%A4ne>

Infos über Repositorien: <http://www.forschungsdaten.org/index.php/Repositorium> und <https://open-access.net/informationen-zu-open-access/repositorien/>

## Links zu bereits existierenden Workshops und Vorträgen

MANTRA - Research Data Management Training: <http://datalib.edina.ac.uk/mantra/>

Consortium of European Social Science Data Archives: Data Management ‑ Archive & Publish: <https://www.cessda.eu/Research-Infrastructure/Training/Expert-tour-guide-on-Data-Management/6.-Archive-Publish>

Open Data 101 introduction crash course: <http://www.opencon2017.org/open_data_101>

Rechtliche Rahmenbedingungen bei der Anonymisierung von Daten: <https://osf.io/4e8bn/>

## Data Wiz ‑ Assistenzsystem für Forschungsdatenmanagement

Das Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) ist gerade dabei, ein Assistenzsystem für das Management psychologischer Forschungsdaten mit dem Namen Data Wiz zu entwickeln (<https://datawiz.zpid.de/>). Data Wiz soll bei der systematischen Dokumentation von Forschungsdaten helfen. Die so systematisch gesicherten und dokumentierten Daten können dann über ein Repositorium wie OSF zur Verfügung gestellt werden. Das Assistenzsystem befindet sich momentan in der öffentlichen Testphase und kann hier getestet werden: <https://datawizdemo.zpid.de/>

Das ZPID richtet sich momentan strategisch generell neu aus als Universalinfrastruktureinrichtung im Sinne eines Open Science Instituts. Hierbei ist mittelfristig geplant, in ähnlicher Weise wie das OSF verschiedene Open Science Angebote der Projektplanung, -registrierung, -dokumentation, -durchführung, -archivierung und -veröffentlichung zu integrieren (siehe <http://leibniz-psychology.org/>).

# Lösungen

Im Folgenden sollen konkrete Möglichkeiten und Schritte vorgestellt werden, mit denen Open Data Praktiken in die Realität umgesetzt werden können. Die erste Anleitung soll einen niedrigschwelligen Einstieg geben, während im zweiten Teil einige Themen und Hinweise vertieft dargestellt werden. Zusätzlich zu der in diesem Modul beschriebenen offenen Bereitstellung der Daten sollten alle weiteren notwendigen Informationen zur Verfügung gestellt werden, die notwendig sind, um die in einem Paper berichteten Ergebnisse reproduzieren zu können. Dazu gehören insbesondere eine Erläuterung der Variablen (z.B. in einem Codebook; vgl. Modul „Open Materials“) und die Beschreibung der mit den Daten durchgeführten statistischen Analysen (z.B. durch die Bereitstellung eines gut dokumentierten Codes; vgl. Modul „Reproducible Code“)

## Open Data KURZ UND KNAPP

### Suchen Sie die aktuellste Datei Ihres Datensatzes heraus (z.B. .csv, .sav-file)

### Prüfen Sie, ob Ihr Datensatz vollständig anonymisiert ist. Ist dies nicht der Fall, anonymisieren Sie ihren Datensatz.

### Laden Sie Ihren Datensatz auf einem Repositorium, beispielsweise auf einem für ihr Manuskript kreierten projekt auf [www.osf.io](http://www.osf.io) hoch.

### Verlinken Sie den Datensatz in Ihrem Manuskript.

## Open data DETAILLIERT

### Ankündigung in Ethik- und Drittmittelanträgen

Bereits im Ethikanträgen/Drittmittelanträgen sollte die geplante offene Bereitstellung der Daten in anonymisierter Form angekündigt werden.

### Passende Einverständnis bei Versuchspersonen einholen

Die Teilnehmer und Teilnehmerinnen einer Studie sollten im Rahmen der Einwilligung zur Teilnahme darauf aufmerksam gemacht werden, dass ihre anonymisierten Daten ggf. für eine Nachnutzung durch Dritte zur Verfügung gestellt werden. Wenn Daten vollständig anonymisierbar sind, muss eine entsprechende Einwilligungserklärung nicht notwendigerweise eingeholt werden, da keine individuellen Zuordnungen mehr möglich sind. Ist dies nicht der Fall, und die Daten sind nicht vollständig anonymisierbar, so muss eine explizite, schriftliche Einwilligung zur Datennachnutzung eingeholt werden. Eine ausformulierte Einwilligungserklärung, sowie weitere Empfehlungen in Form eines Entscheidungsbaums finden sich in den Zusatzinformationen weiter unten in diesem Dokument.

### Daten erheben

### Daten anonymisieren

Damit personenbezogene Daten offen zugänglich gemacht werden können, müssen diese anonymisiert werden. Die Daten gelten dann als anonymisiert, wenn sie so abgeändert wurden, dass eine direkte oder indirekte Identifizierung der einzelnen Merkmalsträger unmöglich ist. Anonymisierte Daten unterliegen nicht dem Datenschutz. Möglichkeiten, die eigenen Daten zu anonymisieren, gibt es viele. Hier einige häufig verwendete Verfahren:

* Vollständige Entfernung entsprechender Variablen aus dem Datensatz: Sowohl direkte Identifikatoren (z.B. Telefonnummer) als auch indirekte Identifikatoren (z.B. die Kombination von Angaben zum Geburtsjahr & Arbeitsstelle) sollten entfernt werden.
* Zusammenfassung von oder zu Kategorien: z.B. Alterskategorien, Regionen / Bundesländer
* Zusammenfassung verschiedener Variablen zu einem abgeleiteten Wert: z.B. Ersetzung der expliziten Angaben von Wohnort und Arbeitsplatz durch eine Angabe zur Distanz zwischen beiden Orten.

Es ist zu beachten, dass nicht aus der Kombination verschiedener erhobener Merkmale (u.a. auch über unterschiedliche Studien mit den gleichen Teilnehmer\*innen hinweg) Personen identifizierbar werden (z.B. Geburtsort, Geschlecht, Semesterzahl, Alter). Mögliche Maßnahmen sind u.a. die Löschung der entsprechenden Merkmale oder die Zusammenfassung zu Kategorien. Daten, die prinzipiell nicht vollständig anonymisierbar sind (z.B. Videos, Bilder, Interview-Daten), sollten nicht veröffentlicht, sondern stattdessen intern (sicher) archiviert werden. In diesem Fall eignen sich meist weiter verarbeitete Varianten der Daten (z.B. Teile von Transkripten, Kategorisierungen, Verhaltenskodierungen) zur anonymisierten Veröffentlichung.

Eine weitere Möglichkeit des Data Sharing bei datenschutzrechtlichen Bedenken bieten die so genannten „Scientific use files“ (SUF), die eine abgestufte Einschränkung des Nutzerkreises bieten. SUF setzen eine faktische Anonymisierung voraus (d.h. eine Deanonymisierung ist nur mit einem unverhältnismäßig hohen Aufwand an Zeit, Kosten und Arbeitskraft möglich). Solche Datensätze werden auf Antrag nur Forscherinnen und Forschern von wissenschaftlichen Einrichtungen zu wissenschaftlichen Zwecken zur Verfügung gestellt. Diese Datensätze dürfen üblicherweise die Räume der nachnutzenden Institution nicht verlassen, Nachnutzende müssen eine Geheimhaltungs- und Nichtweitergabeerklärung unterzeichnen und die Rohdaten müssen nach einer definierten Frist gelöscht werden. Insbesondere bei besonders schützenswerten Daten wie etwa Selbstauskünften über die ethnische Herkunft, politische, religiöse oder philosophische Überzeugungen sowie Daten über Gesundheit oder Sexualleben sollte man das Identifizierungsrisiko genau prüfen und die Verwendung eines SUF in Betracht ziehen. Die Anwendung einer SUF-Lizenz kann in Einzelfällen auch dann angemessen sein, wenn begründbar anzunehmen ist, dass Datensätze für missbräuchliche Zwecke genutzt werden. An dieser Stelle sollte jedoch betont werden, dass der Default beim Data Sharing immer der sein sollte, die Daten für alle frei verfügbar zu machen, sofern sich die Daten anonymisieren lassen.

### Nutzungsrechte klären

Um dafür zu sorgen, dass Andere die eigenen Daten ohne jeweilige Nachfrage nutzen können, bieten sich so genannte Creative-Commons Lizenzen an. Dabei handelt es sich um Standard-Lizenzverträge, die es den Urhebern auf einfache Weise erlauben, der Allgemeinheit Nutzungsrechte einzuräumen. Je nach Variante können hierbei mehrere Einschränkungen erhoben werden, beispielsweise eine Zitierungspflicht oder eine Verpflichtung auf nichtkommerzielle Nutzung. Auf der Website [www.creativecommons.org](http://www.creativecommons.org) lassen sich mit wenigen Klicks die Lizenzeigenschaften auswählen, die Namensgebung anpassen, woraufhin ein Code generiert wird, der sich problemlos auf Websites oder in den Datensatz einfügen lässt.

### Dateiformat für die Archivierung festlegen

Es bietet sich natürlich an, das Dateiformat für die Veröffentlichung zu verwenden, mit dem auch selbst gerechnet wurde. Wenn Daten für die Öffentlichkeit bestimmt sind, sollte allerdings darauf geachtet werden, dass ein interoperables Format (ein Format, dass von unterschiedlichen Systemen gleichermaßen genutzt werden kann) verwendet wird. Als solches bieten sich Plaintext-Dateien beispielsweise im CSV (Comma separated values)-Format an, die von allen Statistikprogrammen gelesen werden können. Weitere Informationen finden sich weiter unten unter „Zusatzinformationen - Dateiformate für die Langzeitarchivierung“.

### Datensätze auf ein Repositorium hochladen

Repositorien sind Informationsinfrastrukturen, die digitale Forschungsdaten dauerhaft speichern und organisieren, um ihre Auffindbarkeit und Zugänglichkeit zu sichern. Ein Beispiel für ein solches Repositorium ist das Open Science Framework ([www.osf.io](http://www.osf.io)). Nach kostenloser Registrierung lassen sich hier Forschungsdaten ganz einfach hochladen. Folgende Schritte sind hierzu vom Dashboard aus notwendig:

1. Oben rechts auf den Button: „Create new project“ klicken.
2. In die auftauchende Maske den Titel des Projekts eingeben, dann auf „Go to new project“ klicken.
3. In der oberen Leiste auf die Registerkarte „Files“ klicken.
4. Auf den Unterordner „OSF storage“ klicken.
5. In der Leiste oben auf den Button „Upload“ klicken.
6. Die Datensätze auswählen und hochladen. Auf eine sinnvolle Benennung achten.
7. In der Leiste links oben auf den Projektnamen klicken.
8. Oben rechts auf die Auswahl „Make public“ klicken und auf Nachfrage mit „Confirm“ bestätigen.

Neben dem Open Science Framework gibt es auch weitere Repositorien, wie das ZPID Psych Data (<http://psychdata.zpid.de/>) oder das GESIS datorium (<https://datorium.gesis.org/xmlui/>). Weitere Übersichten der verschiedenen Repositorien finden sich im Registry of Research Data Repositories (<https://www.re3data.org/>) oder bei der WWU (<https://www.uni-muenster.de/Forschungsdaten/publizieren/publikationswege/>).

### Datensatz zitierbar machen ‑ Persistent identifiers

Im Zuge der Digitalisierung und den Entwicklungen des Internets bieten sich heutzutage neue Möglichkeiten. Durch neue Technologien ist das ehemals bestehende Platzproblem von Forschungsdaten quasi aufgehoben. Das Internet bietet unbegrenzt viel Platz, um Forschungsergebnisse zu publizieren. Weiterhin können digitalisierte Datensätze einer großen Anzahl von Leuten ohne großen Aufwand in Repositorien zugänglich gemacht werden. Im Zuge dessen gewinnen veröffentlichte Datensätze eine zunehmend eigenständige Bedeutung. Sie können weitere Forschungsvorhaben anregen oder beschleunigen Durch **persistent identifiers** (z.B. doi) können sie leicht zitierbar gemacht werden und somit das Ansehen einer Forschungsarbeit beträchtlich steigern. Langfristig erfolgt somit potentiell eine Verschiebung von einer publikationszentrierten hin zu einer datenzentrierten Perspektive.

Persistent identifiers ermöglichen eine stabile, weltweit eindeutige und ortsunabhängige Referenzierung elektronischer Dokumente und bieten so Zitationssicherheit. Ein häufig verwendeter Identifier-Typ ist der DOI (Digital Object Identifier).

Sobald ein Projekt im OSF öffentlich gemacht wurde, kann für das Projekt ein DOI vergeben werden. Folgende Schritte sind hierfür notwendig:

1. Das Projekt öffentlich machen, falls nicht bereits geschehen.
2. Die Seite des Projekts aufrufen.
3. In der dritten Zeile oben links auf die Option „Create DOI/ARK“ klicken und bestätigen.

### Verlinkung im Manuskript

Die Information dazu, dass, unter welchen Bedingungen und welche Daten öffentlich zugänglich sind, sollte gemeinsam mit dem entsprechenden Link zu den Daten im main text des Manuskripts gegeben werden. Der übliche Hinweis dazu, dass die Daten auf Anfrage geschickt werden, hat sich in vielen Fällen als nicht funktional herausgestellt, da nur die wenigsten Autorinnen und Autoren auf Anfragen reagieren und ihre Daten zur Verfügung stellen. Daher lässt sich dieser Hinweis nicht als „Open Data“-Praxis verstehen.

Können (Teile der) Daten (bspw. aus rechtlichen Gründen oder wegen der fehlenden Möglichkeit der Anonymisierung) nicht öffentlich zugänglich gemacht werden, so sollte im main text des Manuskripts erläutert werden, warum dies so ist.

In den meisten Fällen ist Open Data keine entweder/oder-Frage, denn oft kann zumindest etwas Open Data erreicht werden: Durch effektive Anonymisierung, Auswahl anonymisierbarer Daten aus dem Gesamtdatensatz, Auswahl bereits aggregierter Daten, die Verwendung von scientific use files etc. Ein Vorgehen nach dem Motto „So viel Open Data wie möglich“ ist empfehlenswert. In jedem Fall sollte das genaue Vorgehen der Datenauswahl im Manuskript erklärt und begründet werden.

Die Daten sollten optimalerweise bereits bei der Einreichung eines Manuskriptes offen verfügbar sein, so dass Reviewer für etwaige Re-, Kontroll- und Alternativanalysen Zugang zu den Daten haben. Um trotzdem im Falle eines blind-review Verfahrens die Anonymität der Autoren aufrechtzuerhalten, können anonymisierte view-only links angegeben werden. Hier eine genaue Anleitung: <http://help.osf.io/m/sharing/l/524049-view-only-links>.

## FAQ

**Warum besteht Bedarf, die Regel des Teilens von Forschungsdaten nur auf Anfrage zu verändern?**

 Wird zu wenig umgesetzt, Datensätze werden oft auch auf Anfrage nicht geteilt.

 Häufig sind Daten älterer Projekte nicht mehr vollständig auffindbar oder sind in einem schlechten Aufbereitungsstand. Daten von Beginn an öffentlich zugänglich zu machen sorgt nicht zuletzt auch dafür, dass Daten nicht verschwinden und deutlich besser aufbereitet und weniger fehlerbehaftet sind.

**Warum sollte ich meine Daten überhaupt veröffentlichen?**

 Mit Open Data Praktiken lassen sich mehrere Ziele für die wissenschaftliche Gemeinschaft erreichen, zum Beispiel:

* Unterstützung von Transparenz in der Forschung im Sinne der Open Science Praktiken
* Verbesserung der Qualität der Datenaufbereitung und –kontrolle
* Bessere Möglichkeiten der Qualitätskontrolle im Reviewprozess
* Ermöglichung der Reanalyse und des Testens von neuen oder alternativen Hypothesen und Methoden der Analyse mit den Originaldaten
* Unterstützung und Ermöglichung von Meta- und Megaanalysen
* Unterstützung von methodischen Studien zu Datensammlungsmethoden und statistischen Analysen (Daten dienen als Beispieldatensätze)
* Ermöglichung der Erforschung neuer Fragestellungen, die von den ursprünglichen Autoren nicht verfolgt wurden
* Ermöglichung der Erstellung von neuen Datensätzen durch die Kombination verschiedener Quellen

**Was, wenn jemand meine Daten verwendet und mir somit meine Forschungsfrage klaut?**

 Die bisherigen Erfahrungen zeigen, dass dieser Fall in der Regel nicht auftritt. Das Potential für effektivere und neue Kooperationen ist deutlich höher als die Gefahr eines Gebrauchs für eine Publikation ohne weitere Absprache. Die von der DGPs herausgegebenen [Richtlinien zum Umgang mit Forschungsdaten im Fach Psychologie](http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/empfehlungen_forschungsdaten_psychologie.pdf) gehen in sehr differenzierter Weise auf die Rechte und Pflichten von Datenbereitstellern und Datennachnutzenden ein. An diese Richtlinien sollten sich alle halten.

 Es kann hilfreich sein, den gewünschten Umgang mit den Daten und die gewünschte Kommunikation und ggf. Kooperation mit Datennachnutzenden auf der Seite, auf der die Daten zur Verfügung gestellt werden, explizit auszuführen.

 Man kann Sperrfristen zur Veröffentlichung von Daten einrichten, oder nach und nach Teile des Datensatzes veröffentlichen. So werden bspw. häufig nicht direkt alle Daten eines größeren Forschungsprojekts offen zugänglich gemacht, sondern zunächst nur die Daten, die zur Reproduktion der Ergebnisse einzelner Paper notwendig sind.

 Alternativ können Daten auch als scientific use files gespeichert werden und nur auf Anfrage herausgegeben werden – dies sollte aber die Ausnahme darstellen (mehr Informationen oben im Abschnitt *4. Daten anonymisieren*)

**Wie gehe ich sicher, dass mein Datensatz anonymisiert ist?**

 Ein Datensatz ist anonymisiert, wenn die direkte oder indirekte Identifizierung der einzelnen Merkmalsträger unmöglich ist. Um dies zu erreichen, können Variablen beispielsweise gelöscht werden oder zu Kategorien oder zu einem abgeleiteten Wert zusammengefasst werden. Sollte all dies nicht möglich sein, so gibt es immer noch die Möglichkeit, die Daten als Scientific Use Files zu Verfügung zu stellen (mehr Informationen oben im Abschnitt *4. Daten anonymisieren*). In den meisten Fällen wird es so möglich, zumindest Teile der Daten für die Öffentlichkeit zugänglich zu machen.

**Open Data kommt für mich nicht in Frage, weil ich empfindliche klinische Daten/Videodaten/ … erhoben habe, richtig?**

 In den allermeisten Fällen ist es möglich durch geeignete Methoden der Anonymisierung (siehe oben) zumindest Teile der Daten offen bereitzustellen. Aspekte, die zur Identifikation einzelner Personen führen können und die nicht anonymisierbar sind, sollten selbstverständlich nicht offen zur Verfügung gestellt werden. Das heißt aber nicht, dass für das entsprechende Projekt Open Data gar nicht in Frage kommt.

**Open Data kommt für mich nicht in Frage, weil ich bei dem Projekt keine entsprechenden Einverständniserklärungen eingesetzt habe, richtig?**

 Die offene Bereitstellung anonymisierter Daten bedarf rechtlich keines expliziten Einverständnisses der Versuchspersonen. Trotzdem sollte man bei neu zu erhebenden Daten eine entsprechende Information in den Einverständniserklärungen ergänzen (siehe oben). Das Fehlen solcher Hinweise bei bereits abgeschlossenen Datenerhebungen ist aber kein Grund, die anonymisierten Daten nicht offen bereitzustellen.

**Open Data bei Meta-Analysen, ergibt das Sinn?**

 Ja, auch bei Meta-Analysen können und sollten die Daten, die metaanalysiert wurden offen zur Verfügung gestellt werden.

**Ich mache mir Sorgen über Data Sharing in Bezug auf Big Data**.

 Bei gelungener Anonymisierung sollte keine Identifikation der Teilnehmenden möglich sein.

 Bei der Analyse von Smartphonedaten, Online Social Network-Daten und anderer Online-Daten ist allerdings besondere Vorsicht geboten, da eine Anonymisierung häufig schwer zu realisieren ist.

**Unter welchen Bedingungen bekommt meine Publikation den „Open Data“ Stempel?**

* Das Journal in dem das Paper erscheint muss diese Möglichkeit anbieten.
* Im Paper wird ein dauerhafter Link zu den Daten eingefügt, die in einem öffentlichen Open-Access Repositorium abgelegt sind. Die Daten müssen einen persistent identifier haben und in einem zeitgestempelten, unveränderlichen und permanenten Format bereitgestellt sein.
* Unabhängige Forscher haben alle Informationen, um mit den Daten die im Paper berichteten Ergebnisse zu reproduzieren.

Typischerweise gelten folgende Bedingungen:

* + Provide the URL, doi, or other permanent path for accessing the data in a public, open-access repository:

Confirm that there is sufficient information for an independent researcher to reproduce all of the reported results, including codebook if relevant.

**Kann ich finanzielle Mittel für das Datenmanagement beantragen?**

 Die meisten Förderer erkennen zunehmend den Nutzen offener Daten und fordern diese auch ein. Entsprechend können häufig Mittel für das Datenmanagement beantragt werden (bspw. zusätzliche Hilfskraftmittel). Bei der DFG ist dies beispielsweise möglich. Die Beantragung solcher Mittel sollte in jedem Fall gut begründet werden.

# zusatzinformationen

## Textbausteine einer einwilligungserklärung INKL. INORMATIONEN ZUR VERWENDUNG DER ANONYMISIERTEN DATEN

Optimalerweise sollten bereits die bei der Datenerhebung eingesetzten Einverständniserklärungen einen Hinweis auf die offene Zurverfügungstellung der anonymisierten Daten enthalten. Das Fehlen solcher Hinweise (beispielsweise, weil es sich um Daten älterer Projekte handelt oder entsprechende Hinweise vergessen wurden) sollte aber nicht als Begründung verwendet werden, um Daten nicht offen zur Verfügung zu stellen. Anonymisierte und daher prinzipiell nicht auf einzelne Personen rückführbare Daten können auch ohne entsprechende Hinweise offen zur Verfügung gestellt werden.

Mögliche Textbausteine

*Freiwilligkeit:*

Die Teilnahme an der Studie ist freiwillig. Sie können jederzeit und ohne Angabe von Gründen Ihre Einwilligung zur Teilnahme an dieser Studie widerrufen, ohne dass Ihnen daraus Nachteile entstehen. Auch wenn Sie die Studie vorzeitig abbrechen, haben Sie Anspruch auf eine entsprechende anteilige Vergütung oder die entsprechende Anzahl Versuchspersonenstunden für den bis dahin erbrachten Zeitaufwand. Sie können Ihre Einwilligung zur Speicherung der Daten bis zum Ende der Datenerhebung widerrufen, ohne dass Ihnen daraus Nachteile entstehen.

*Datenschutz:*

Da keine personenbezogenen Daten erhoben werden, ist nach Abschluss der Datenerhebung prinzipiell keine Zuordnung mehr zwischen den Daten im Datensatz und Ihrer Person möglich ‑ der Datensatz ist anonym. Entsprechend ist nach Abschluss dieser Datenerhebung auch keine gezielte Löschung Ihres persönlichen Datensatzes möglich, da wir diesen nicht zuordnen können.

*Verwendung der anonymisierten Daten:*

Die Ergebnisse und Daten dieser Studie werden als wissenschaftliche Publikation veröffentlicht. Dies geschieht in anonymisierter Form, d. h. ohne dass die Daten einer spezifischen Person zugeordnet werden können. Die vollständig anonymisierten Daten dieser Studie werden als offene Daten im Internet in einem Datenarchiv namens \_\_\_\_\_\_\_\_\_\_\_ zugänglich gemacht. Damit folgt diese Studie den Empfehlungen der Deutschen Forschungsgemeinschaft (DFG) und der Deutschen Gesellschaft für Psychologie (DGPs) zur Qualitätssicherung in der Forschung.

Hiermit versichere ich, dass ich die oben beschriebenen Teilnehmerinformationen verstanden habe und mit den genannten Teilnahmebedingungen einverstanden bin.

## Verarbeitung und Bereitstellung von anonymen und personenbezogenen DATEN

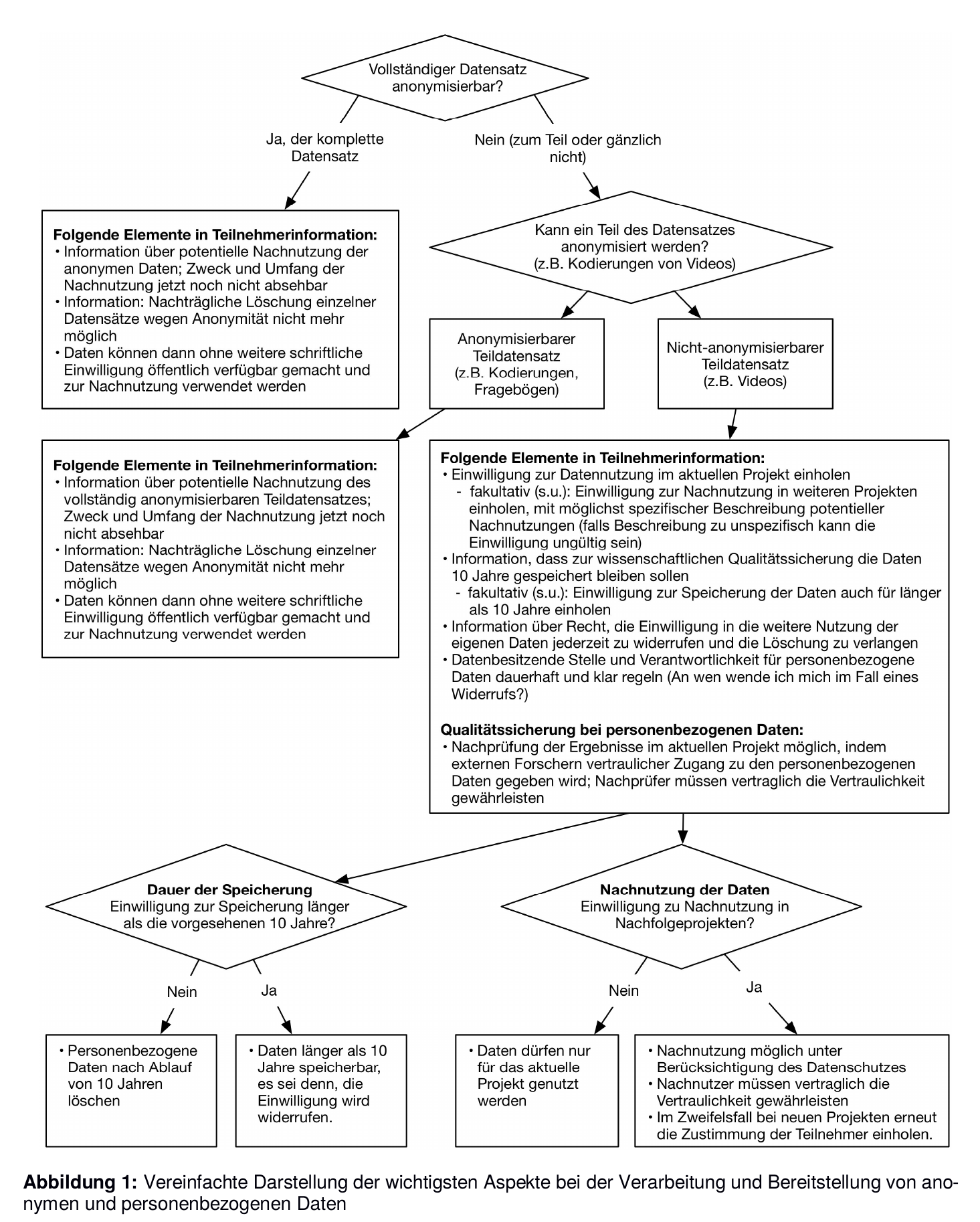


Abbildung 1: Wichtige Aspekte bei der Verarbeitung und Bereitstellung von anonymen und personenbezogenen Daten (Quelle: http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien\_forschungsdaten.pdf)

## Rohdaten und Primärdaten

Als **Rohdaten** oder Ursprungsaufzeichnungen werden all jene Daten bezeichnet, die in der Form ihrer Erstaufzeichnung bestehen. Dazu gehören beispielsweise die Kreuze auf einem Fragebogen, oder die Videoaufzeichnung eines Interviews.

Daten gelten als **Primärdaten**, nachdem sie auf der Basis der Rohdaten in ein digitales Format übertragen wurden. Durch diesen Transformationsprozess erhalten die Daten eine stärker strukturierte Form, die allerdings auch mit einem Informationsverlust einhergeht. Es handelt sich hierbei immer noch um vollkommen unbearbeitete quantitative oder qualitative Daten. Beispiele für Primärdaten wäre z.B. Code „1“ für eine Ja-Antwort.

## Rechtliche Rahmenbedingungen

### Urheberrechtsgesetz (UrhG)

Wissenschaftliche Primärdaten unterliegen grundsätzlich nicht dem Schutz des Urheberrechtsgesetzes.[[2]](#footnote-2) Ausnahmen bestünden, sofern der Datensatz eine persönliche intellektuelle Schöpfung mit einem Mindestniveau an Originalität aufwiese, was durch die Verwendung wissenschaftlicher Konventionen normalerweise nicht der Fall ist. Sollten Urheberrechte bestehen, so kann mit so genannten **Creative Commons Lizenzen** festgelegt werden, wem welche Nutzungsrechte eingeräumt werden. Bei der Verwendung von psychologischen Test- und Fragebogeninstrumenten allerdings liegen die Nutzungsrechte bei den Verlagen, weswegen dort in jedem Fall eine Konsultation über die Nutzung erfolgen muss.

### Bundesdatenschutzgesetz (BDSG)

Das BDSG legt fest, dass personenbezogene Daten (also Daten, die eindeutig einer bestimmten Person zugeordnet sind) nur für die Zwecke wissenschaftlicher Forschung verarbeitet oder genutzt werden. Sie sind weiterhin nach Erreichen des primären Forschungszweckes zu anonymisieren.

Eine erfolgreiche Anonymisierung liegt vor, wenn eine direkte oder indirekte Identifizierung der einzelnen Merkmalsträger unmöglich ist. Anonymisierte Daten unterliegen keinen datenschutzrechtlichen Bestimmungen[[3]](#footnote-3).

### Scientific Use Files (SUF)

Bei Scientific Use Files (SUF) handelt es sich um faktisch anonymisierte Daten, die nur auf Antrag Forscherinnen und Forschern wissenschaftlicher Einrichtungen zur Verfügung gestellt werden. Es handelt sich hierbei also um eine abgestufte Einschränkung des Nutzerkreises, welche bei datenschutzrechtlichen Bedenken von Vorteil sein kann. Für die Nutzung der Daten müssen Nachnutzende eine entsprechende Geheimhaltungs- und Nichtweitergabeerklärung unterschreiben und die Rohdaten müssen meist nach einer bestimmten Frist wieder gelöscht werden. Obwohl SUF in manchen Fällen sinnvoll sein können, sollen Daten vorrangig uneingeschränkt geteilt werden, sodass alle daran teilhaben können. Das Datenteilen auf Anfrage klappt zudem nicht gut. Trotz einer unterschriebenen Erklärung von Autoren, die Daten auf Anfrage zu teilen, führen viele solcher Anfragen ins Nichts, da sich die Autoren schlichtweg nicht zurückmelden. Daher sollte eine Datenveröffentlichung im Sinne der Open Data immer die erste Option sein.

## Datenmanagementpläne erstellen

Datenmanagementpläne werden vor Beginn des Forschungsvorhabens erstellt und dienen der Beschreibung des Lebenszyklus‘ von Daten. Heutzutage werden sie von immer mehr Förderungsinstitutionen als Vorgabe gefordert. Inhaltlich werden die Rahmenbedingungen und Strategien zum Umgang mit den Forschungsdaten festgelegt. Dazu gehören die Erstellung, die Verarbeitung, der Umfang, die Sicherung und Veröffentlichung der anfallenden Forschungsdaten. Für den Plan gibt es keine generelle starre Struktur, eine sinnvolle Gliederung könnte allerdings folgende Punkte enthalten:

1. Projektbeschreibung
2. Art der Daten
3. Umfang und Qualität der projekteigenen Daten
4. Organisation der Daten
5. Speicherung, Backupstrategie und Sicherheit
6. Rechtliche Aspekte
7. Archivierung, Austausch und Publikation
8. Verantwortliche und Pflichten
9. Kosten und Ressourcen

Weitere Informationen finden sich in der Checkliste des WissGrid-Projekts (<http://www.forschungsdaten.org/images/b/b0/Leitfaden_Data-Management-WissGrid.pdf>), sowie auf den Hinweisen der WWU zur Organisation von Forschungsdaten: <https://www.uni-muenster.de/Forschungsdaten/organisieren/datenmanagementplan/> .

Auch das oben bereits erwähnte Tool [Data Wiz](https://datawiz.zpid.de/) soll bei der Erstellung von Datenmanagementplänen eine Hilfestellung geben.

## Creative Commons Lizenzen und ihre Module

Wie bereits oben erwähnt gibt der Urheber mit einer Creative Commons Lizenz keinesfalls seine Rechte vollständig auf, sondern lässt die Nutzung seiner Daten unter bestimmten Bedingungen zu. Wie diese Bedingungen aussehen, kann eigenständig mit der Aktivierung bzw. Deaktivierung folgender vier Module gestaltet werden (für Details siehe: <https://creativecommons.org/licenses/?lang=de>) .

|  |  |  |
| --- | --- | --- |
| **Modul** | **Icon** | **Erläuterung** |
| 1. by, Attribution (Namensnennung) | Bildergebnis für creative commons icons | Bei Verwendung des Werkes durch Dritte muss der Name des Urhebers und ggf. der Name des Werkes genannt werden. |
| 1. nc, Non Commercial (nicht kommerziell) | Bildergebnis für creative commons icons | Das Werk darf nicht für kommerzielle Zwecke verwendet werden. |
| 1. nd, No Derivative Work (keine Bearbeitung) | http://mirrors.creativecommons.org/presskit/icons/nd.large.png | Das Werk darf nicht bearbeitet oder in anderer Weise verändert werden. |
| 1. sa, Share Alike (Weitergabe unter gleichen Bedingungen) | Bildergebnis für creative commons icons sa | Wird das Werk verändert oder als Grundlage für ein anderes Werk verwendet, so darf das neu entstandene Werk nur unter Verwendung von identischen oder vergleichbaren Lizenzen weitergegeben werden. |

## Dateiformate für die Langzeitarchivierung

Oft ist das Ende eines Projektes der kritischste Moment im Lebenszyklus eines Datensatzes, da hier oft die Finanzierung endet. Gerade aus diesem Grund ist es sehr wichtig, Maßnahmen zur Datenaufbereitung und Veröffentlichung vorher geklärt zu haben. Ein wichtiger Schritt hierbei ist die Auswahl eines geeigneten Formates für die Langzeitarchivierung. Folgende Merkmale zeichnen geeignete Dateiformate aus:

* **Offenheit:** Das Dateiformat ist frei zugänglich und open source
* **Hoher Verbreitungsgrad**: Das Format ist weit verbreitet und bleibt voraussichtlich lange bestehen
* **Geringe Komplexität**: Der Inhalt sollte unmittelbar lesbar sein
* **Keine integrierten Schutzmechanismen**
* **Möglichst umfassende Selbstdokumentation**: Die nötigen Metadaten sind möglichst bereits integriert
* **Robustheit:** Geringe Anfälligkeit gegenüber einzelnen Bit-Fehlern
* **Keine Abhängigkeiten**: Von bestimmter Hardware, Betriebssystemen etc.
* **Verlustfreiheit:** Es gehen keine relevanten Informationen verloren

Für quantitative Datensätze empfiehlt sich eine Plaintextdatei zusammen mit einem command file eines Statistikprogramms, das auch Metadaten enthalten kann. [[4]](#footnote-4)

## Repositorien

Wie bereits oben erwähnt stellen Repositorien die ideale Plattform dar, um Daten dauerhaft speichern und öffentlich zugänglich machen zu können. Eine grobe Unterscheidung lässt sich zwischen institutionellen und disziplinären Repositorien treffen.

Institutionelle Repositorien werden, wie der Name bereits sagt, von Institutionen betrieben (Universitäten, Forschungsorganisationen etc.). Die WWU bietet einen eigenen Publikationsserver unter dem Namen „miami“ (Münstersche Informations- und Archivsystem multimedialer Inhalte) an. Forschungsdaten können hier nach Absprache veröffentlicht werden und werden für die Zitierbarkeit mit einem persistent identifier versehen. Mehr Informationen unter: [www.uni-muenster.de/Publizieren/dienstleistungen/repository/index.html](http://www.uni-muenster.de/Publizieren/dienstleistungen/repository/index.html)

Im Gegensatz dazu sind disziplinäre Repositorien institutionsübergreifend und bündeln die Forschung eines bestimmten Faches. Für das Fach Psychologie gibt es beispielsweise das Repositorium PsychData des ZPID (Leibniz-Zentrum für Psychologische Information und Dokumentation). Mehr Informationen unter [www.psychdata.de](http://www.psychdata.de).

Das in der Psychologie international bekannteste Repositorium, welches zugleich diverse weitere Open Science-Funktionalitäten anbietet, ist das Open Science Framework ([www.osf.io](http://www.osf.io))

Bei der Auswahl eines geeigneten Repositoriums lohnt es sich, auf folgende Qualitätsmerkmale zu achten:

* **Unabhängigkeit** & wissenschaftliche **Professionalität** der bereitstellenden Institution
* **Persistenz der Daten:** Sind die Daten über einen längeren Zeitraum gesichert?
* **Zugänglichkeit der Daten:** Öffentlicher und kostenloser Abruf
* **Identifizierbarkeit der Daten:** Vergabe eines persistenten Identifikators
* **Klärung der Rechte** an den Daten
* Die Möglichkeit, **Daten sowohl öffentlich, als auch nicht-öffentlich** abzuspeichern

1. http://opendatahandbook.org/guide/de/what-is-open-data/ [↑](#footnote-ref-1)
2. Spindler & Hillegeist, 2011 [↑](#footnote-ref-2)
3. https://www.psychdata.de/downloads/PsychData-Handbuch\_2013.pdf [↑](#footnote-ref-3)
4. https://www.psychdata.de/downloads/PsychData-Handbuch\_2013.pdf [↑](#footnote-ref-4)