



AUSSDA

AUSTRIAN  
SOCIAL SCIENCE  
DATA ARCHIVE

# Forschungsdaten archivieren und nachnutzbar machen – zwischen Data Sharing und Datenschutz

OEFEB-Tagung - Linz, September 2019

Mag. Dr. Otto Bodi-Fernandez  
Universität Graz

Have data? Need data? | [www.ausdda.at](http://www.ausdda.at)

# Warum Data Sharing in den Sozialwissenschaften? – Vorteile für...



...die wissenschaftliche Community



...die Öffentlichkeit



...Datenproduzent\*innen

# Vorteile für die Community

- **Transparenz und Nachvollziehbarkeit von Studienergebnissen (replication data)**
- **Nachnutzbarkeit bei neuen Fragestellungen (Sekundäranalyse)**
- **Nutzung realer Datenbasis für Lehrzwecke**
- **Schaffung neuer Datengrundlagen durch Kombination von Datensätzen**

# Vorteile für die Öffentlichkeit

- Ableitung von politischen Handlungsempfehlungen durch vorhandene Datengrundlagen Daten
- Mehr Effizienz bei investierten Fördermitteln (durch Mehrfachnutzung vorhandener)
- Einhaltung des datenschutzrechtliches Gebot der Datenminimierung (weniger Erhebungen nötig)

# Vorteile für Datenproduzent\*innen

- Zukünftige Datennutzung durch Produzent\*innen selbst wird einfacher
- Leichte Zitierbarkeit der Daten durch persistente Identifier (zB doi)
- Höhere Sichtbarkeit der Daten und zugehörigen Forschungsleistungen → mehr Impact
- Schaffung neuer Kooperationsmöglichkeiten
- Einhaltung von Förderrichtlinien und guter wissenschaftlicher Praxis

# Vorgaben des FWF

„Für **Forschungsdaten**, die den wissenschaftlichen Publikationen des Projekts zugrunde liegen, ist der **offene Zugang verpflichtend**. Das sind alle Daten, die zur Reproduktion und Überprüfbarkeit der Ergebnisse der Publikationen erforderlich sind, einschließlich der zugehörigen Metadaten. [...] **Sollte aus rechtlichen, ethischen oder anderen Gründen ein offener Zugang zu diesen Daten nicht oder nur teilweise möglich sein, ist das im Datenmanagementplan (DMP) zu begründen**“.

# Vorgaben des FWF

„Alle Forschungsdaten und ihre Metadaten müssen **auffindbar, zugänglich, interoperabel und wiederverwendbar** sein (siehe [FAIR-Prinzipien](#)). Weiters müssen die folgenden Kriterien erfüllt werden:

- Es können **institutionelle, disziplinspezifische oder disziplinübergreifende Repositorien** zur Archivierung herangezogen werden. [...]
- Die Daten müssen so abgelegt werden, dass sie **uneingeschränkt wiederverwendet** werden können [...]
- Hinterlegte Datensätze müssen über einen **persistenten Identifikator** zitierbar sein.“

# Vorgaben des FWF

„Alle Forschungsdaten und ihre Metadaten müssen auffindbar, zugänglich, interoperabel und wiederverwendbar sein (siehe [FAIR-Prinzipien](#)). Weiters müssen die folgenden Kriterien erfüllt werden:

zB AUSSDA

- Es können institutionelle, disziplinspezifische oder disziplinübergreifende **Repositorien** zur Archivierung herangezogen werden. [...]
- Die Daten müssen so abgelegt werden, dass sie uneingeschränkt wiederverwendet werden können [...]
- Hinterlegte Datensätze müssen über einen persistenten Identifikator zitierbar sein.“



# Vorgaben des FWF

„Alle Forschungsdaten und ihre Metadaten müssen auffindbar, zugänglich, interoperabel und wiederverwendbar sein (siehe [FAIR-Prinzipien](#)). Weiters müssen die folgenden Kriterien erfüllt werden:

- Es können institutionelle, disziplinspezifische oder disziplinübergreifende Repositorien zur Archivierung herangezogen werden. [...]
- Die Daten müssen so abgelegt werden, dass sie **uneingeschränkt wiederverwendet** werden können [...]
- Hinterlegte Datensätze müssen über einen persistenten Identifikator zitierbar sein.“

zB AUSSDA

zB Creative Commons Lizenz

# Vorgaben des FWF

„Alle Forschungsdaten und ihre Metadaten müssen auffindbar, zugänglich, interoperabel und wiederverwendbar sein (siehe [FAIR-Prinzipien](#)). Weiters müssen die folgenden Kriterien erfüllt werden:

- Es können institutionelle, disziplinspezifische oder disziplinübergreifende Repositorien zur Archivierung herangezogen werden. [...]
- Die Daten müssen so abgelegt werden, dass sie uneingeschränkt wiederverwendet werden können [...]
- Hinterlegte Datensätze müssen über einen **persistenten Identifikator** zitierbar sein.“

zB AUSSDA

zB Creative Commons Lizenz

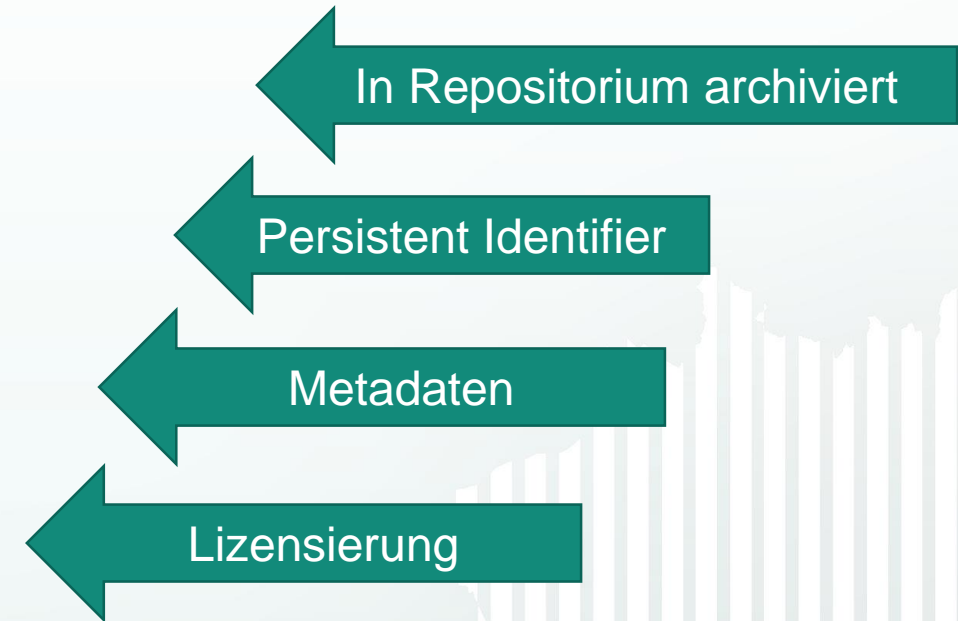
zB DOI-Nummer

# Was ist AUSSDA?

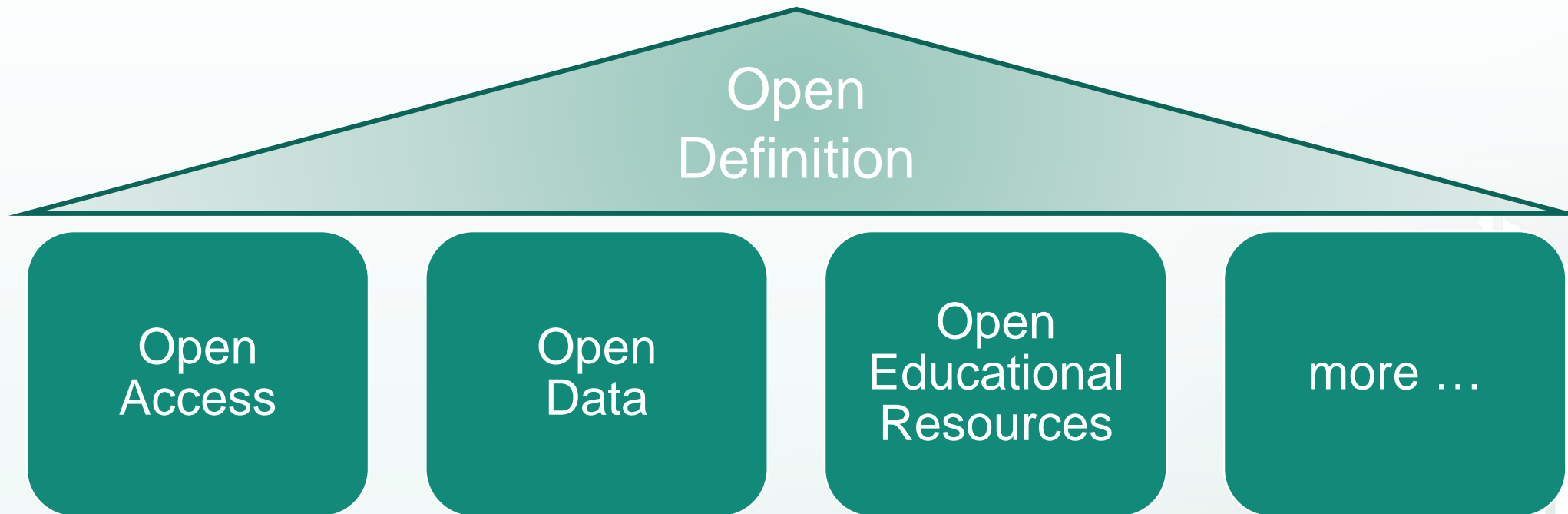
- **AUSSA – The Austrian Social Science Data Archive:**  
 Zentrale Infrastruktureinrichtung („Core Facility“) für sozialwissenschaftliche Forschungsdaten in Österreich
- **Aufgabe:** Archivierung und zur Verfügungen Stellung sozialwissenschaftlicher Datenbestände für Wissenschaft und Gesellschaft
- **Träger:** Konsortium aus den Universitäten Wien (Universitätsbibliothek), Linz (Institut für Soziologie) und Graz (Center for Social Research)
- **Finanzierung:** Bundesministerium für Bildung, Wissenschaft und Forschung
- **Internationaler Kontext:** AUSSDA ist der europäische Partner im Consortium of European Social Science Data Archives (CESSDA ERIC)

# AUSSDA entspricht **FAIR DATA** Prinzipien

- **F** indable
- **A** ccessible
- **I** nteroperable
- **R** eusable



# Open Definition



*Open Definition:*

*"Open data and content can be **freely used, modified, and shared** by **anyone** for **any purpose**" (Open Knowledge Foundation).*

# Daten in den Sozialwissenschaften

## ● Personenbezogene und sensible Daten

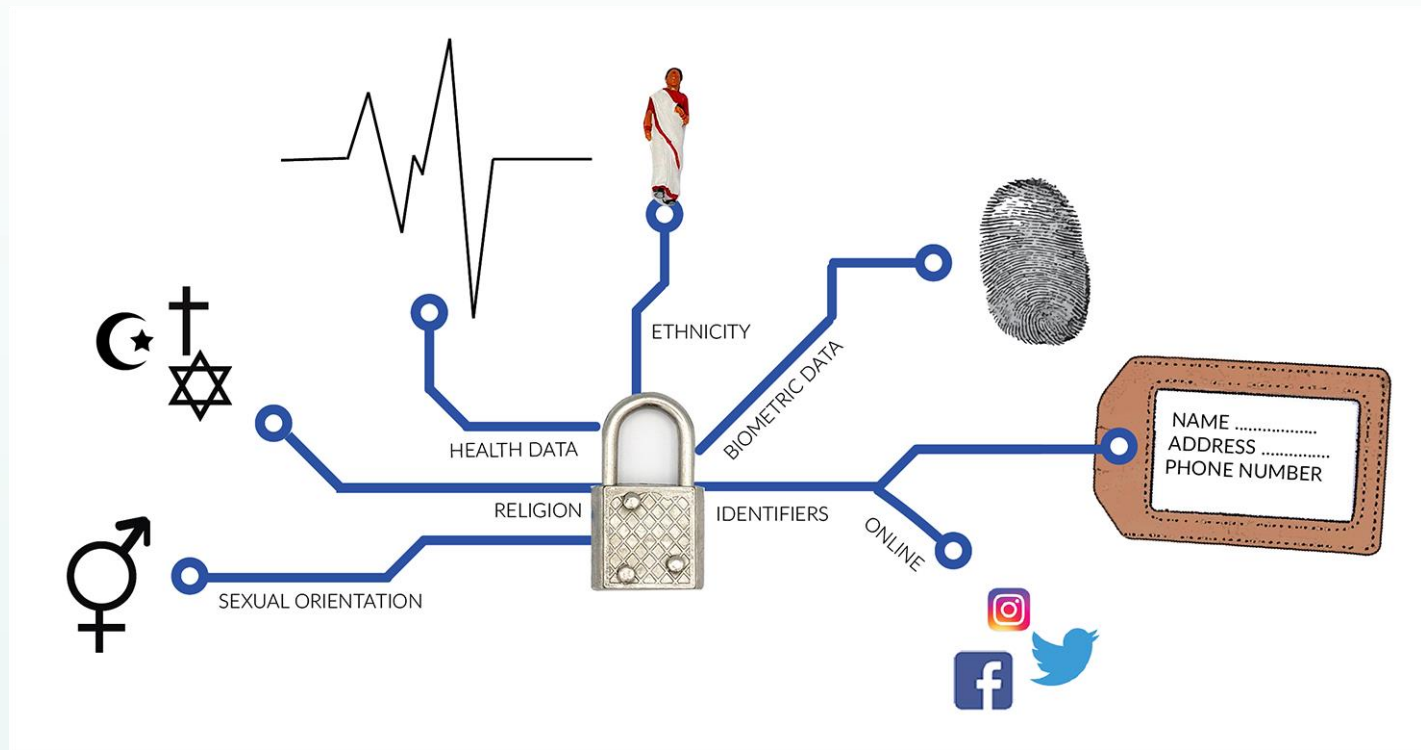


Bild: CC BY-SA CESSDA Training Working Group (2017 - 2018)

# Was sind personenbezogene Daten?

- Lt. DSGVO: alle Informationen, die sich auf eine **identifizierte** oder **identifizierbare** natürliche Person („betroffene Person“) beziehen.

Als **identifizierbar** wird eine natürliche Person angesehen, die **direkt oder indirekt**, insbesondere mittels Zuordnung zu einer Kennung wie einem

- Namen,
- zu einer Kennnummer,
- zu Standortdaten,
- zu einer Online-Kennung

oder zu einem oder mehreren besonderen Merkmalen **identifiziert werden kann**, die Ausdruck der

- physischen,
- physiologischen,
- genetischen,
- psychischen,
- wirtschaftlichen,
- kulturellen oder
- sozialen Identität

dieser natürlichen Person sind.

# Was sind personenbezogene Daten?

- Das DSGVO ist nur auf lebende natürliche Personen anwendbar.
- Anonyme oder anonymisierte Daten gelten nicht als personenbezogene Daten.
- Anonymisierung: in einer Weise, dass die Personen nicht mehr identifiziert werden kann.
- Was nicht als personenbezogene Daten zählt, fällt nicht unter DSGVO → Wenngleich es andere Gründe geben kann, diese Informationen dennoch zu schützen – zB ethische.



# Besondere Kategorien personenbezogener Daten (sensible Daten)

- 🌐 Informationen, die ein Risiko für die Grund- und Freiheitsrechte bergen.
  - 🌐 *ethnische Herkunft*
  - 🌐 *politische Meinungen*
  - 🌐 *religiöse oder weltanschauliche Überzeugungen*
  - 🌐 *Gewerkschaftszugehörigkeit*
  - 🌐 *genetischen Daten*
  - 🌐 *biometrischen Daten*
  - 🌐 *Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person*

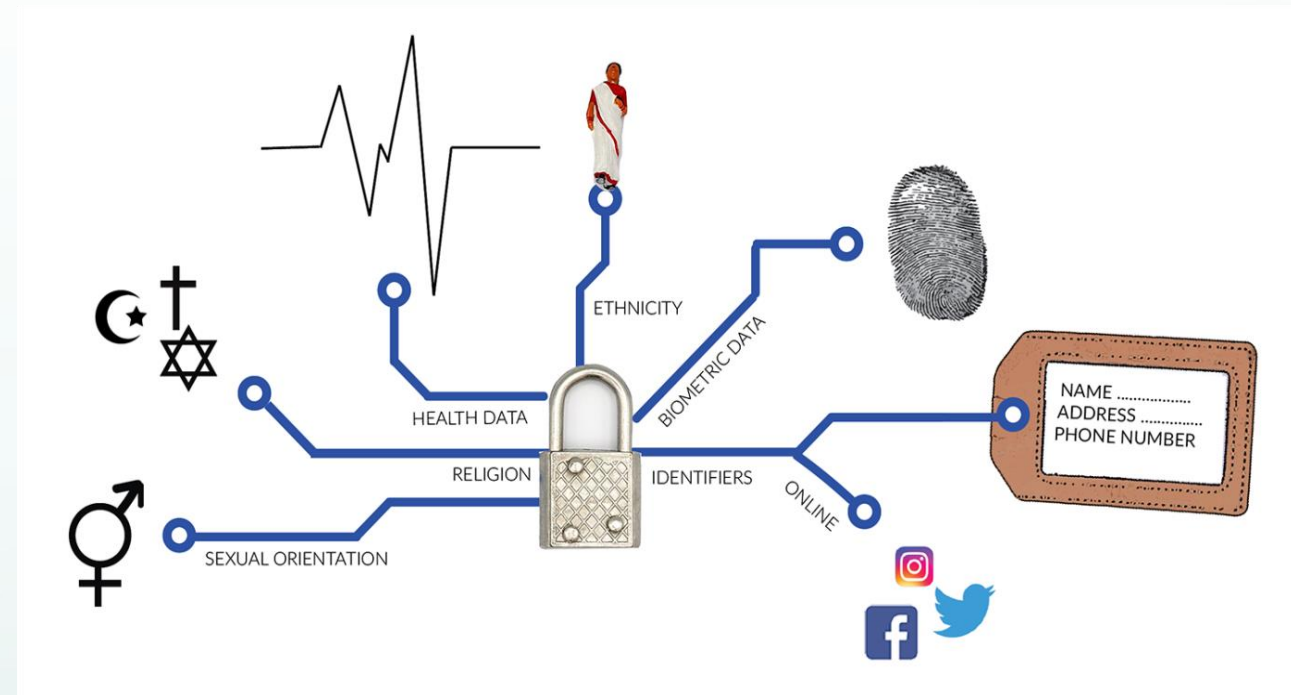







Bild: CC BY-SA CESSDA Training Working Group (2017 - 2018)

# Forschungsethische Grundsätze in den Sozial- und Wirtschaftswissenschaften (Rat für Sozial- und Wirtschaftsdaten 2017)

- Wissenschaftliche Güte und Integrität der Forschenden
  - Allgemeine wissenschaftliche Standards
  - Integrität der Forschenden
  - Dokumentation, Archivierung und Sekundärnutzung
- Risikoabwägung und Schadensvermeidung
  - Vertraulichkeit und Anonymisierung
- Informiertes Einverständnis
  - Information der Teilnehmenden
  - Freiwilligkeit der Teilnahme

# Datenschutz

- 
 Datenschutzrechtliche Vorgaben sind zu beachten, wenn **personenbezogene** Daten **natürlicher** Personen im Spiel sind.
  
- 
 Rechtlicher Rahmen
  - 
 Datenschutzgrundverordnung (DSGVO)
  - 
 Datenschutzgesetz (DSG)
  - 
 Forschungsorganisationsgesetz (FOG)

# Datenschutz

„Verbot mit Erlaubnisvorbehalt“: Verarbeitung personenbezogener Daten ist grundsätzlich verboten und nur unter bestimmten Bedingungen erlaubt:

- wenn dies gesetzlich oder durch andere Rechtsvorschriften erlaubt oder vorgeschrieben ist
- oder die jeweils betroffene Person, um deren Daten es geht, hierin eingewilligt hat.

Rechtlich unproblematisch ist daher die Forschung mit

- **anonymisierten** Daten (bei diesen fehlt der Personenbezug), und mit
- personenbezogenen Daten, wenn eine **Einwilligung der Betroffenen** vorliegt

# Verarbeitung personenbezogener Daten

- Es gelten folgende sechs Prinzipien (DSGVO, Artikel 5/1)
  - a) Rechtmäßigkeit
  - b) Zweckbindung (*Ausnahmen für Wissenschaft gem. Artikel 89*)
  - c) Datenminimierung
  - d) Richtigkeit
  - e) Speicherbegrenzung (*Ausnahmen für Wissenschaft gem. Artikel 89*)

# Data Sharing und Datenschutz

- **Forschungsfördereinrichtungen und Journals**  
→ Forderung: Daten möglichst offen zugänglich machen (Open Data)
- **Forschungsethik und Datenschutzrecht**  
→ Forderung: Daten möglichst schützen

*Wie können diese Anforderungen vereinbart werden?*

# DATA SHARING

Einholen  
von  
Einverständnis

Anonymi-  
sierung

Zugriffs-  
beschränkung

## DATENSCHUTZ

# Informierte Einwilligung

- Informiertes Einverständnis ist erforderlich...
  - aufgrund forschungsethischer Grundsätze
    - → Teilnehmer\*innen treffen freiwillige Entscheidung über Teilnahme oder Nicht-Teilnahme auf Basis, dass sie zuvor ausreichend über die Studie informiert wurden.
  - aufgrund datenschutzrechtlicher Bestimmungen
    - Informierte Einwilligung ist eine (von mehreren möglichen) Rechtsgrundlagen zur Verarbeitung personenbezogener Daten
      - Betroffene Personen willigen in die Verarbeitung ihrer persönlichen Daten ein.



# Informierte Einwilligung

- Wie hat informierte Einwilligung zu erfolgen?
  - freiwillig
  - spezifisch (für bestimmten Zweck)
  - in informierter Weise
  - durch eine unmissverständlich abgegebene Willensbekundung
    - in Form einer Erklärung oder
    - einer sonstigen eindeutigen bestätigenden Handlung
  
- *Betroffene Person, gibt damit zu verstehen, dass Sie in die Verarbeitung ihrer personenbezogenen Daten einverstanden ist.*

# Informierte Einwilligung

## ● Bestandteile

- Informationsteil mit allgemeinen Angaben zum Projekt
- Datenschutzrechtlicher Informationsteil
- Text der Einwilligung selbst mit Unterschrift der betroffenen Person

# Informierte Einwilligung

## ● Bestandteile

- Informationsteil mit allgemeinen Angaben zum Projekt
  - Angaben zum Projekt: (zB Titel, Ziele, Fragestellungen)
  - Wer erhebt die Daten? (Institution, Projektleiter, Ansprechpartner)
  - Welche Daten werden erhoben? (zB Kompetenzen, Einstellungen, Verhalten)
    - Bei Verarbeitung besonderer Kategorien ist ausdrücklich darauf hinzuweisen!
  - Wie werden die Daten erhoben? (Interview, Fragbogen)
  - Wie werden die Daten verarbeitet? (Digitalisierung, Transkription, Kodierung, Anonymisierung)
  - Zu welchem Zweck werden die Daten verwendet? (Auswertung, Veröffentlichung, Archivierung und Weitergabe zur Sekundärnutzung, Archivierung und Weitergabe zur Sekundärnutzung)

# Informierte Einwilligung

## ● Bestandteile

### ● Datenschutzrechtlicher Informationsteil

- Angaben zu den Rechten der Betroffenen:
  - Freiwilligkeit
  - Widerrufsrecht
  - Folgenlosigkeit bei Verweigerung oder Widerruf

# Informierte Einwilligung

## Bestandteile

- Text der Einwilligung selbst mit Unterschrift der betroffenen Person
  - Einwilligung kann für unterschiedliche Zwecke auch einzeln eingeholt werden (granular consent)

Ich mit einverstanden

*...mit der Teilnahme an der Studie und der Verwendung meiner personenbezogenen Daten.*    O ja / O nein

*...mit der weiteren Nutzung meiner Daten – über das Projekt hinaus – für die Bildungsallgemein.*    O ja / O nein

Datum, Unterschrift: \_\_\_\_\_

Quelle: Meyermann & Porzelt (2019)

Mit den voranstehenden (im Informationsteil) gemachten Ausführungen zur Verwendung meiner personenbezogenen Daten bin ich einverstanden.

Datum, Unterschrift: \_\_\_\_\_

# Informierte Einwilligung

## ☉ Einwilligung bei Surveys

- ☉ Wird in der Regel nicht schriftlich eingeholt, da keine persönlichen Identifikationen (Name, Adresse, etc.) erhoben werden.
- ☉ Teilnehmer\*innen erhalten Aviso-Schreiben mit Informationsblatt – Teilnahme an Befragung gilt als Einwilligung.
- ☉ Das Informationsblatt sollte auch die geplante Archivierung und Nachnutzung der Daten beinhalten!

# Informierte Einwilligung

## ☉ Zentrale Anforderungen

- ☉ Laienverständlichkeit
- ☉ Freiwilligkeit
- ☉ Zweckbindung

# Informierte Einwilligung

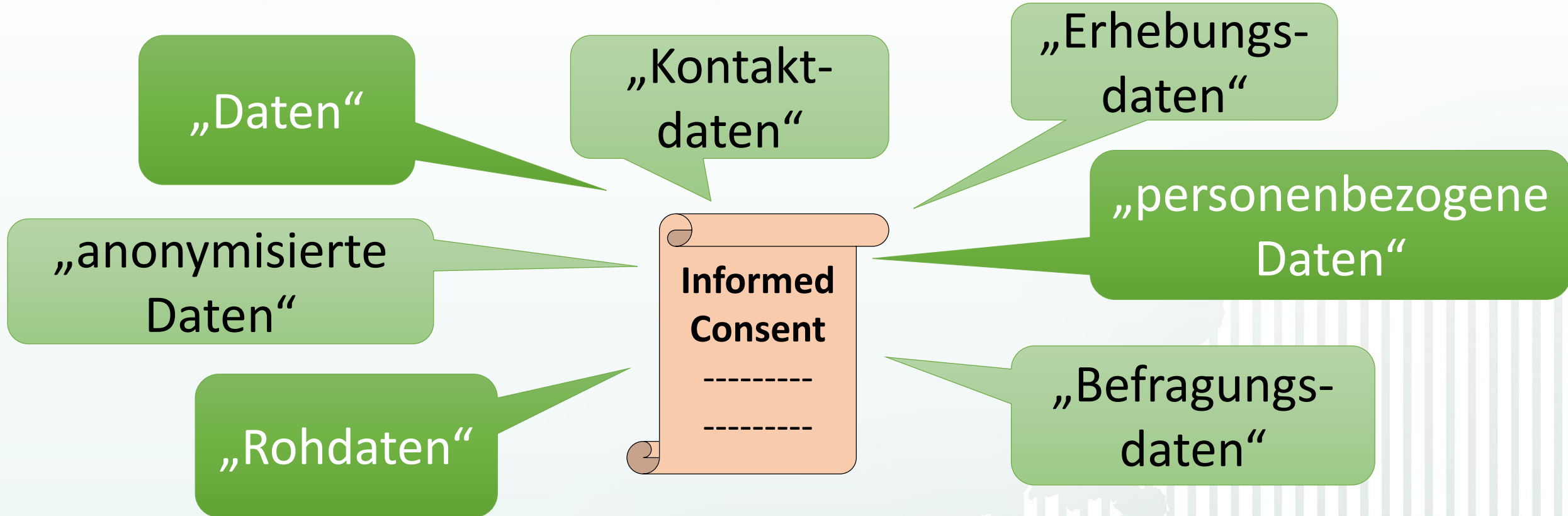
## ● Zentrale Anforderungen

### ● Laienverständlichkeit

- an den Empfängerhorizont angepasste Sprache verwenden
- Vermeiden von Fachtermini (Rohdaten, Transkription, Videographie, Anonymisierung, etc.)
- Auf trennscharfe, eindeutige, und konsistente Begriffe achten - insbesondere beim Datenbegriff



# Häufig verwendete Begriffe



- Datenschutzrechtliche Vorgaben beziehen sich nur auf „personenbezogene“ Daten
- Werden Aussagen zu anonymen Daten gemacht, können diese dennoch rechtlich binden sein.

# Informierte Einwilligung

Beispiele Formulierungen:

*„Die erhobenen Daten werden ausschließlich im Forschungsprojekt verwendet und nicht an Dritte weitergegeben. Nach Abschluss des Projektes werden die erhobenen Daten gelöscht.“*

# Informierte Einwilligung

Beispiele Formulierungen:

*„Die anonymisierten Daten werden ausschließlich im Forschungsprojekt verwendet und nicht an Dritte weitergegeben.“*

# Informierte Einwilligung

Beispiele Formulierungen:

*„Die personenbezogenen Daten (Kontaktdaten) werden ausschließlich im Forschungsprojekt verwendet und nicht an Dritte weitergegeben. Nach Abschluss des Projektes werden die personenbezogenen Daten gelöscht.“*

# Informierte Einwilligung

## ● Zentrale Anforderungen

### ● Freiwilligkeit

- Betroffene müssen Einverständnis freiwillig geben können.
- Hierauf muss in Einwilligungserklärung explizit hingewiesen werden.
- Zustimmungen sind nicht als freiwillig anzusehen wenn...
  - Die Nichtteilnahme mit Nachteilen verbunden ist bzw. die Teilnahme durch die Vermeidung von Nachteilen motiviert ist
  - Sie auf Zustimmung zu unverhältnismäßigen Gegenleistungen (Incentives) beruhen.
- *Freiwilligkeit* ist besonders deutlich zu machen, wenn Abhängigkeitsverhältnisse bestehen:  
 zB: Schüler\*innen – Lehrer\*innen – Direktor\*innen

# Informierte Einwilligung

## ● *Beispiel A (Nachteil bei Nicht-Zustimmung):*

*„Wir würden uns freuen, wenn auch Sie Ihrem Kind die Chance geben würden, an unserer Studie teilzunehmen. Damit tragen Sie dazu bei, dass die Kinder mit Lese- und Rechtschreibschwächen in der Klasse Ihres Kindes besser betreut werden.“*

## ● *Beispiel B (unverhältnismäßige Gegenleistung):*

*„Für Ihre Teilnahme am einstündigen Interview bedanken wir uns bei Ihnen mit einer Aufwandsentschädigung in Höhe von 500 €.“*

# Informierte Einwilligung

- Zweckbindung
- Einverständniserklärungen müssen Aussagen zur Verwendung der Daten enthalten.
- Grundsätzlich ist die Verarbeitung der Daten an die genannten Zwecke gebunden.
  - Zwecke sollten so eng wie möglich, aber so breit wie nötig angegeben werden.
  - In Wissenschaft oft nicht möglich, da nicht alle Zwecke vorhersehbar sind.
  - Für wissenschaftliche Zwecke sieht die DSGVO (und DSG) die Möglichkeit des **Broad Consent** vor → breit gefasste Einwilligung für einen Forschungsbereich

# Informierte Einwilligung

## ● Zweckbindung

*Enge Zweckbindung:* Angabe eines Personenkreises und eines konkreten Zwecks:

*„Daten werden nur im Rahmen des Projektes zur Untersuchung der Einflüsse von Lehrerhandeln auf... genutzt und nur durch die genannten Projektmitarbeiter/innen bearbeitet.“*

*Weite Zweckbindung:* Angabe eines allgemeinen Verwendungszwecks und Nennung eines (unbestimmten) Personenkreises:

*„Die Daten werden ausschließlich zu Zwecken der Bildungsforschung durch ausgewiesene Wissenschaftler/innen verwendet.“*



# Anonymisierung

- Beste Lösung: Keine Informationen zur Identifizierung erheben
  
- Zweitbeste Lösung: Informationen so bearbeiten, dass Identifikation ausgeschlossen werden kann → „**Anonymisierung**“
  
- Direct Identifiers:
  - zB Name, Adresse, Telefonnummer
  
- Indirect Identifiers:
  - Informationen die in Kombination mit zusätzlichen Informationen eine Identifizierung erlauben. zB Kombination aus Beruf, Einkommen, Alter und Geschlecht

# Anonymisierung

- Formale Anonymisierung
- Faktische Anonymisierung
- Absolute Anonymisierung





VS.

- Pseudonymisierung → Daten werden so bearbeitet, dass ohne Hinzunahme zusätzlicher Informationen keine Identifizierung möglich ist.
- Pseudonymisierte Daten sind als personenbezogene Daten anzusehen.

# Anonymisierungsmethoden (quantitative Daten)

- **Variablen entfernen**
  - *zB. Namen, Adressen, sonstige Identifikatoren*
- **Reduzieren der Genauigkeit (Kategorien zusammenfassen)**
  - *zB deatillierte geographische Angaben, detaillierte Berufsangaben*
- **Bedeutungsinhalte bei Textvariablen verallgemeinern**
  - *zB offene Antworten kategorisieren*
- **Beschränken der oberen und unteren Bereiche kontinuierlicher Variablen**
  - *zB: Alter, Haushaltsgröße, Kinderzahl*

# Zugangsklassen für Forschungsdaten

- 
**Open Access Files (OA):**  
*sind als Open Data unter Creative Commons Attribution International 4.0 License (CC BY)*
- 
**Scientific Use Files (SUF):**  
*Zugang für registrierte User - stehen nur für wissenschaftliche Zwecke zur Verfügung – zB aus Datenschutzgründen*
- 
**Embargo:**  
*Metadaten sind im Repository zugänglich. Datensatz selbst wird erst nach Ablauf einer gesetzten Frist freigegeben.*
- 
**Auch Nutzungsverträge mit speziell vereinbarten Konditionen sind möglich**  
*→ zB. Zugriff nur auf Anfrage und nach Zustimmung (für sensible Daten)*

# Open Access bei AUSSDA

- Public Domain (CC0):
  - Metadaten
- CC BY-Lizenz:
  - Methodenberichte
  - Fragebögen
  - Codebücher
  - Weitere Begleitdokumente
  - Gegebenfalls (absolut anonymisierte) Forschungsdaten

# Zusammenfassung

- Daten erheben nötig?
  - Wenn schon vorhanden, dann vermeiden (Sekundäranalyse)
- Personenbezogene Daten nötig?
  - wenn nicht, dann vermeiden!
  - wenn ja, Einverständnis einholen, ggf. pseudonymisieren, getrennt aufbewahren
- Personenbezogene Daten nicht mehr nötig?
  - Löschen bzw.
  - Anonymisieren
- Wie sollen Daten archiviert werden?
  - Open Access:                   möglich bei absolut anonymen Daten, ansonsten...
  - Scientific Use:               für anonymisierte Daten mit Restrisiko der Identifizierbarkeit
  - Restricted Access:         für sensible Daten (Einwilligung einholen)

# Fehlerquellen

- Untätigkeit aus Unsicherheit bzw. überzogene selbstauferlegte Einschränkungen
  - zB Löschen von anonymen Erhebungsdaten am Projektende
- Unzureichende Planung
  - zB mangelhafte Einwilligungserklärung
- Ex-post-Anonymisierung
  - Vermeidbarer Aufbereitungsaufwand und Fehlerrisiko
- Pseudonyme aus personenbezogenen Angaben zusammensetzen
  - zB Buchstaben aus Vornamen der Eltern
  - Risiko der Identifizierbarkeit



AUSSDA  
AUSTRIAN  
SOCIAL SCIENCE  
DATA ARCHIVE

# Danke!

Für Fragen besuchen Sie unsere Webseite oder kontaktieren Sie uns:

<https://aussda.at>

[info@aussda.at](mailto:info@aussda.at)

+43 01 4277 15323






Twitter: @TheAUSSDA

YouTube: [https://youtu.be/Y\\_IKvkCUHuU](https://youtu.be/Y_IKvkCUHuU)

Have data? Need data? | [www.aussda.at](http://www.aussda.at)



# Quellen:

- 
 CESSDA Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. Retrieved from <https://www.cessda.eu/DMGuide>
- 
 FWF - der Wissenschaftsfond. Open Access für Forschungsdaten. <https://www.fwf.ac.at/de/forschungsfoerderung/open-access-policy/open-access-fuer-forschungsdaten/>
- 
 Meyermann, A. & Porzelt, M. (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. forschungsdaten bildung informiert 6, Version 2.
- 
 Open Knowledge Foundation. The Open Definition. <https://opendefinition.org/>
- 
 RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2017): Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften. RatSWD Output 9 (5). Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.1>
- 
 Watteler, O., & Ebel, T. (2019). Datenschutz im Forschungsdatenmanagement. In Jensen U., Netscher S., & Weller K. (Eds.), Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten (pp. 57-80). Opladen; Berlin; Toronto: Verlag Barbara Budrich. Retrieved from <http://www.jstor.org/stable/j.ctvbkk1p8.7>