



# Offener Zugang zu Forschungsdaten vs. Datenschutz:

## Institutionelle und technische Lösungsmodelle in den Sozial- und Wirtschaftswissenschaften

*Stefan Liebig*  
*Universität Bielefeld*

# Das Datenspektrum in den Sozial- und Wirtschaftswissenschaften

## Quantitativ:

- Surveygenerierte Mikrodaten: Bevölkerungsumfragen
- Experimentell generierte Mikrodaten
- Prozessgenerierte Mikrodaten: Mikrodaten, die im Bereich der Sozialversicherung anfallen (= Renten-, Kranken- und Arbeitslosenversicherung), amtliche Statistik (Zensus)
- Verknüpfte survey-/prozessgenerierte Mikrodaten: Ergänzung durch zusätzliche Informationen (u.a. auch Raumdaten)

## „Qualitativ“:

- Offene Interviews (als transkribierte Texte, Audiodateien) und Textcorpora
- Audio-/Videodaten

# Sozial- und Wirtschaftswissenschaftliche Forschungsdaten: Besonderheiten und Herausforderungen

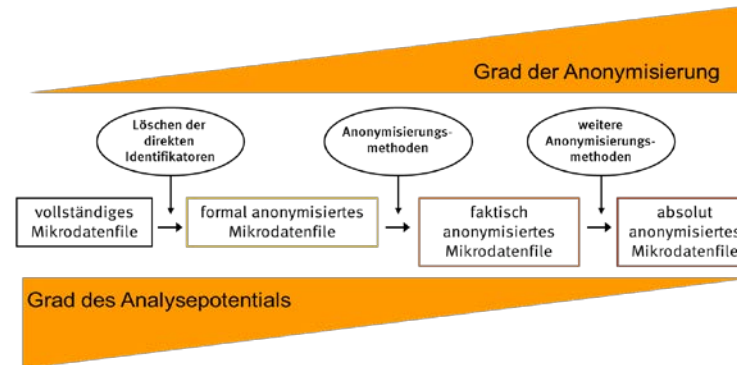
- 1) Mikrodaten mit Informationen über natürliche oder juristische Personen: Datenschutz betrifft unterschiedliche Rechtskreise
  - nP: Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person -> BDSG
  - jP: z.B. Verschwiegenheitspflicht -> BGB etc.

**→ Wahrung der Rechte der Auskunftgebenden**
  
- 2) Umfasst auch Daten, die nicht zum Zwecke der Forschung generiert wurden und die besonderen rechtl. Anforderungen ihrer Nutzung unterliegen (z.B. SGB und Nutzbarkeit außerhalb Deutschlands)

**→ Ermöglichung rechtskonformer Nutzung durch alle Mitglieder der scientific community**

# Sozial- und Wirtschaftswissenschaftliche Forschungsdaten: Besonderheiten und Herausforderungen

- 3) Durch Löschen von Informationen zum Zwecke der Anonymisierung wird Informationsgehalt und damit Wertigkeit für die Forschung reduziert (gilt insbesondere für qualitative Daten)



→ **Bereitstellung des maximalen Analysepotenzials**

- 4) Zunehmende *methodische Heterogenität* (Erhebungsverfahren, Datenarten) und *Komplexität* (Längsschnitt, Mehrebenen, Datenverknüpfung; Mixed-Methods)  
→ **Qualitätssicherung bei Bereitstellung und Nutzung**

# Klassische Wege der Bereitstellung

Bereitstellung sozialwissenschaftlicher Forschungsdaten bereits seit 1960 durch das Zentralarchiv für empirische Sozialforschung in Köln.

Aufgaben: Katalogisierung, Archivierung und Bereitstellung „anonymisierter“ Daten für die wissenschaftliche Nutzung

Nutzungsformen:

- „public use file“ („OpenAccess“) → Verringerung des Re-Identifikationsrisiko durch Bereitstellung eines deutlich reduzierten Datenumfangs/oder durch datenverfremdende Maßnahmen
- „scientific use file“ → Reduzierter Datenumfang und besondere Sicherungsmaßnahmen auf Seiten der Datennutzer (i.d.R. vertragliche Vereinbarungen & technische Vorkehrungen auf Nutzerseite)

# Grenzen der klassischen Archivlösung

- Zugang zu prozessgenerierten Daten nicht möglich, da Daten aus rechtlichen Gründen nicht „migrierbar“
  - Aufbau spezifischer Zugangsmöglichkeiten jenseits von Archiven, d.h. „Produzentennah“ (amtl. Statistik, SGB)
- Aufwand der datenschutzgemäßen Aufbereitung komplexerer Forschungsdaten für klassische Archivierung zu umfangreich
  - Bereitstellungsmöglichkeiten ohne weitergehende datenverändernde Maßnahmen / Reduktion der Information
- Steigendes Risiko der Re-Identifikation bei komplexeren Designs macht „anonymisierte“ Forschungsdaten wissenschaftlich „wertlos“
  - Bereitstellung in Umgebungen, in denen weitgehende Kontrolle der Nutzerinnen und Nutzer gewährleistet ist.

# Grenzen der klassischen Archivlösung

- Steigende Datenschutzrisiken durch den bloßen Aufwuchs von Primärdatenerhebungen oder das unreflektierte Umgehen mit personenbezogenen Daten  
→ niederschwelligere Archivierungsangebote (Datensicherung)
- Notwendigkeit spezifischer fachlicher und methodischer Kompetenzen für die Aufbereitung und Nutzung bei komplexen Forschungsdesigns und heterogenen Datenarten  
→ Notwendigkeit spezifischer (Nutzer-/Produzenten)Services

## Phase I: Aufbau von Forschungsdatenzentren

- Aufbau von Forschungsdatenzentren bei nicht-wissenschaftlichen Datenproduzenten (KVI-Prozess): Statistische Ämter, Bundesagentur für Arbeit, Deutsche Rentenversicherung (= neuartige Datenzugänge)
- Aufbau von Forschungsdatenzentren bei wissenschaftlichen Datenproduzenten: Sozio-ökonomisches Panel, Bildungspanel, pairfam (= Zugänge für sensiblere Datenbestände), Stifterverband, RWI
- Aufbau thematischer Forschungsdatenzentren: Bildung (IQB, DIPF), Betriebe und Organisationen (BO), Arbeitsmarkt (IZA)
- Akkreditierung durch RatSWD



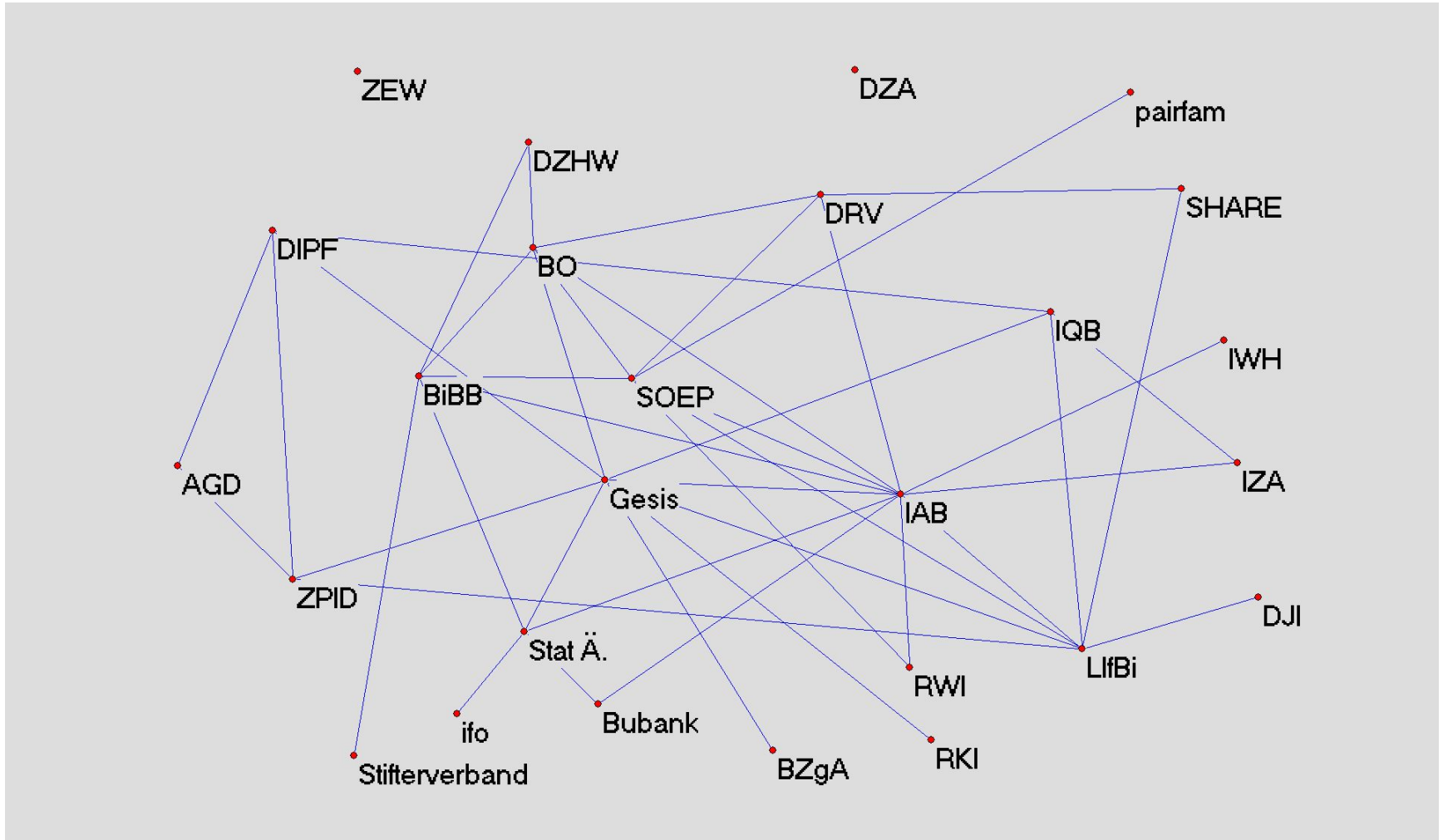
# Phase I: Aufbau von Forschungsdatenzentren

- Etablierung von Konzepten zur Anonymisierung und zu Datenzugangswegen: Je restriktiver der Datenzugangsweg, desto schwächer die an den Daten vorgenommenen Anonymisierungsmaßnahmen.
- Neuartige Datenzugänge:
  - „On-site“ Nutzung: Datennutzung durch Gastaufenthalte vor Ort (Einrichtung spezieller „Gastwissenschaftlerarbeitsplätze“ in den Einrichtungen)
  - Datenfernverarbeitung mit manueller/automatisierter Output-Kontrolle
  - Remote-Access: FDZ im FDZ, Zugang über persönliches Login in Kombination mit zertifizierter biometrischer Authentifizierung („keystroke biometrics“).
  - SUF, PUF, Campus files

## Phase II: Vernetzung von Forschungsdatenzentren

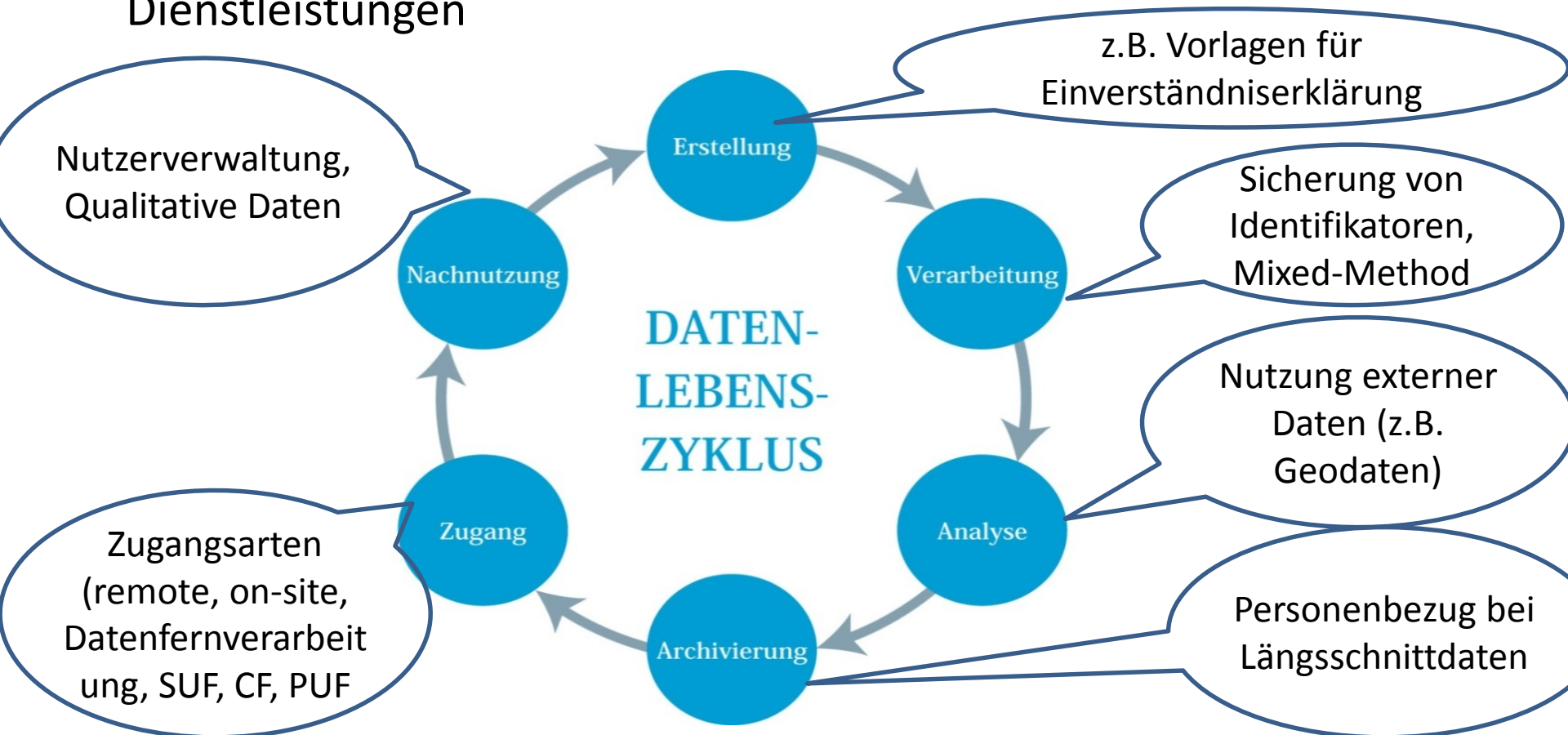
- Etablierung abgestimmter Konzepte zum Datenschutz (Anonymisierung) und zur Datensicherheit.
- Etablierung von gemeinsamen Standards (Interoperabilität, Datenzugang etc.)
- Neuartige Datenzugänge:
  - „FDZ im FDZ“: Datenzugang („On-site“ Nutzung) über andere FDZs (z.B. Zugangsmöglichkeiten zu Daten der Bundesagentur für Arbeit/IAB über lokale Niederlassungen der FDZs der Statistischen Ämter)
  - Internationaler Datenzugang: Außenstellen in den USA und Großbritannien (FDZ der BA im IAB)
- Vernetzung: Ständiger Ausschuss Forschungsdateninfrastruktur

# Vernetzte Forschungsdateninfrastruktur in den Sozial- und Wirtschaftswissenschaften



# Datenschutz durch Forschungsdatenzentren

Datenschutz entlang der einzelnen Phasen des „data life cycles“ durch methodische, technische und organisatorische Dienstleistungen



# Datenschutz durch Forschungsdatenzentren

Drei Varianten eines nachhaltigen und datenschutzkonformen Umgangs mit Forschungsdaten:

**Variante 1:** nachhaltige Sicherung von Forschungsdaten ohne nutzerfreundliche Dokumentation und ohne Nachnutzungsmöglichkeiten

**Variante 2:** nachhaltige Sicherung von Forschungsdaten mit nutzerfreundlicher Dokumentation und mit restriktiven Nutzungsmöglichkeiten

**Variante 3:** nachhaltige Sicherung und Bereitstellung von Forschungsdaten mit nutzerfreundlicher Dokumentation für die Datennutzung außerhalb des Aufbewahrungsortes

# Offener Zugang zu Forschungsdaten vs. Datenschutz?

Standarddatensätze weiterhin „offener Zugang“ – aber mit Abstrichen beim Analysepotenzial

Komplexere Datensätze (Längsschnitt, Mehrebenen, Verknüpfte Daten) sowie besondere Anforderungen des Datenschutzes (Widerrufsrecht) erfordern FDZs und Zugangsregelungen.

Technische und organisatorische Lösungen werden aber Zugangsbarrieren abbauen.

# Referenzen

<http://www.ratswd.de>

[http://www.ratswd.de/dl/RatSWD\\_Output3\\_Forschungsdatenmanagement.pdf](http://www.ratswd.de/dl/RatSWD_Output3_Forschungsdatenmanagement.pdf)

Liebig, S. et al. (2014): Datenschutzrechtliche Anforderungen bei der Generierung und Archivierung qualitativer Interviewdaten. RatSWD Working Papers No 238. Berlin: RatSWD

RatSWD (2016): Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften. RatSWD Output 3, Berlin: RatSWD.