

Nationales Gemeindemonitoring

StatistTreffen – Use Case
Datenanonymisierung

Hochschulbibliothek/ZHAW
Services Forschungsdaten
2026-09-08 | Simon van Rekum

Projektüberblick in Kürze

- SNF-finanzierte Langzeitstudie
- Erhebung alle 6 Jahre, durchgeführt durch ZHAW SML (Fachstelle Public Networks and Service Delivery) in Zusammenarbeit mit UNIL
- Befragung Gemeindeschreiberinnen und Gemeindeschreiber (Verwaltungsseite) – seit 1988
- Befragung Exekutivmitglieder (politische Perspektive) – seit 2008
- Letzte Erhebung 2023
- Vorgesehen: Publikation anonymisierter Datensätze auf SWISSUBase

Projektüberblick in Kürze

Gemeindeschreiberbefragung

Zielgruppe: Gemeindeschreiber/-in (Verwaltung)

Ebene: Gemeinde – 1 Person pro Gemeinde

Inhalt: Verwaltung, Finanzen, Demographie, Fusionen

Vollerhebung (2000+ Befragte) Rücklauf: 85 %

Exekutivbefragung

Zielgruppe: alle gewählte Exekutivmitglieder (Politik)

Ebene: Personen – 3-9+

Inhalt: Motivation, politische Position, Belastung, Familie

Vollerhebung (6000+ Befragte) Rücklauf: 50%

Besonderheiten der Daten

(Halb)öffentliche Personen

- Grunddaten zu Personen und Ämter i.d.R. bekannt (→ Anonymisierung betrifft v.a. Verknüpfung mit Antworten)
- Linkage mit offenen Daten (Wahlresultate und Gemeindewebsites zeigen oft Namen, Partei, Ressort von Exekutivmitgliedern)
- Politisch & finanziell sensibel: Einkommen, Einschätzungen und Abweichungen von Parteilinie

Hohe Dimensionalität

- In beiden Erhebungen viele Variablen (~350, ~245) — Hohes Kombinationsrisiko
- Cross-Survey-Linkage (Erhebungen betreffen dieselben Gemeinden im selben Jahr → gesteigertes Risiko)

Variablen klassifizieren

Gemeindeschreiberbefragung: 65 Fragen, 350 Variablen **Exekutivbefragung:** 61 Fragen, 245 Einzelvariablen
→ Zuordnung der Variablen (bzw. Variablen Blöcke – je nach Fragenstruktur)

Identifiers

Direct Identifiers

Eindeutig, für sich allein re-identifizierend)

- Gemeindenamen, E-Mail, Telefon

Quasi-Identifiers

In Kombination re-identifizierend → keyVars

- Kanton, Gemeindegrösse
- Geburtsjahr, Geschlecht, Zivilstand
- Beruf, Partei, Amtsdauer, Ressort

Attributdaten

Geringeres Re-Identifikationspotenzial
(Achtung: externe Quellen, z.B. lokale Reportagen, Interviews)

- Einschätzungen, Meinungen
- Verwaltungsroutine

Sensible Variablen (→ *sensibleVar*)

- Parteinähe (z.B. Einfluss von Parteiprogramm auf Entscheidungen)
- Psychische Belastungen

Variablen priorisieren

Recap – Warum Daten publizieren? → Nachvollziehbarkeit & Nachnutzung

- Wissenschaftlicher „Wert“ der Variablen gilt es zu berücksichtigen (verschiedene Nutzungsszenarien, versch. Disziplinen)
- Vorgängige Priorisierung der Variablen → tief-priorisierte Variablen früh generalisieren/entfernen
- Jedoch: Wer-Risiko Abwägung

	Hohes Re-ID-Risiko	Niedriges Re-ID-Risiko
Hoher Wert	Schonend und erst nach anderen Variablen anonymisieren. Bsp.: Parteizusammensetzung, Exekutive nach Geschlecht (→ führt in Kombination zu eindeutiger Zuordnung)	Belassen/minimal anpassen. Bsp.: Autonomie-Einschätzung, Fragen zu Privatleben (→ generische 5-stufige Skala)
Niedriger Wert	(früh) entfernen / stark generalisieren. Bsp.: Mitgliedschaft im kantonalen/nationalen Parlament (Doppelmandat), Freitexte «andere» bei Ressort, Berufsgruppen, etc. (→ seltene, unike Angaben; für quantitative Auswertung irrelevant)	Kandidat für frühe Entfernen im Risk Assessment. Bsp.: Sitzungsrhythmus/-dauer (→ Randaspekt)

Quasi-Identifiers & k-anonymity

Fiktiver Beispielauszug: Exekutivmitglieder von GR ~ 250 Antworten

Gemeinde	Alter	Geschlecht	Partei	Ressort	Amtseintritt
Thusis	41	Weiblich	SP	Finanzen	2022
Thusis	62	Männlich	Mitte	Sozial	2017
Safiental	39	Männlich	FDP	Bau	2020
Vals	52	Weiblich	FDP	Sozial	2023
Vaz	49	Weiblich	SVP	Bildung	2020

Gemeinde	Alter	Geschlecht	Partei	Ressort	Amtseintritt	k
GR	50-59	Männlich	FDP	Finanzen/Wirtschaft	2018-2020	5
...	...	...	...	...	...	...
GR	60-69	Weiblich	SVP	Bildung/Soziales	2021-2023	2

✓ $k \geq 3$

✗ $k < 3$

1. Iteration:

Variablen generalisieren, z.B. mit Geschlecht und Partei priorisiert

- Gemeinde → Kanton
- Alter → Bänder (30-39, 40-49...)
- Ressorts → thematisch zusammenlegen (z.B. Bildung/Soziales)
- Amtseintritt → Bänder (...2018-2020, 2021-2023)

2. Iteration:

Stärker generalisieren, weitere Variablen generalisieren (z.B. Parteien zusammenlegen) oder lokale Unterdrückung (KA) → bis Zielwert erreicht (z.B. $k \geq 3$)

Prüfung von k-anonymity in Cross-Survey-Linkage!

Von k-anonymity zu l-diversity

Fiktiver Beispielauszug: Antwort zu Frage 8: *“Wie beurteilen Sie die finanzielle Entschädigung des Exekutivamts?”*

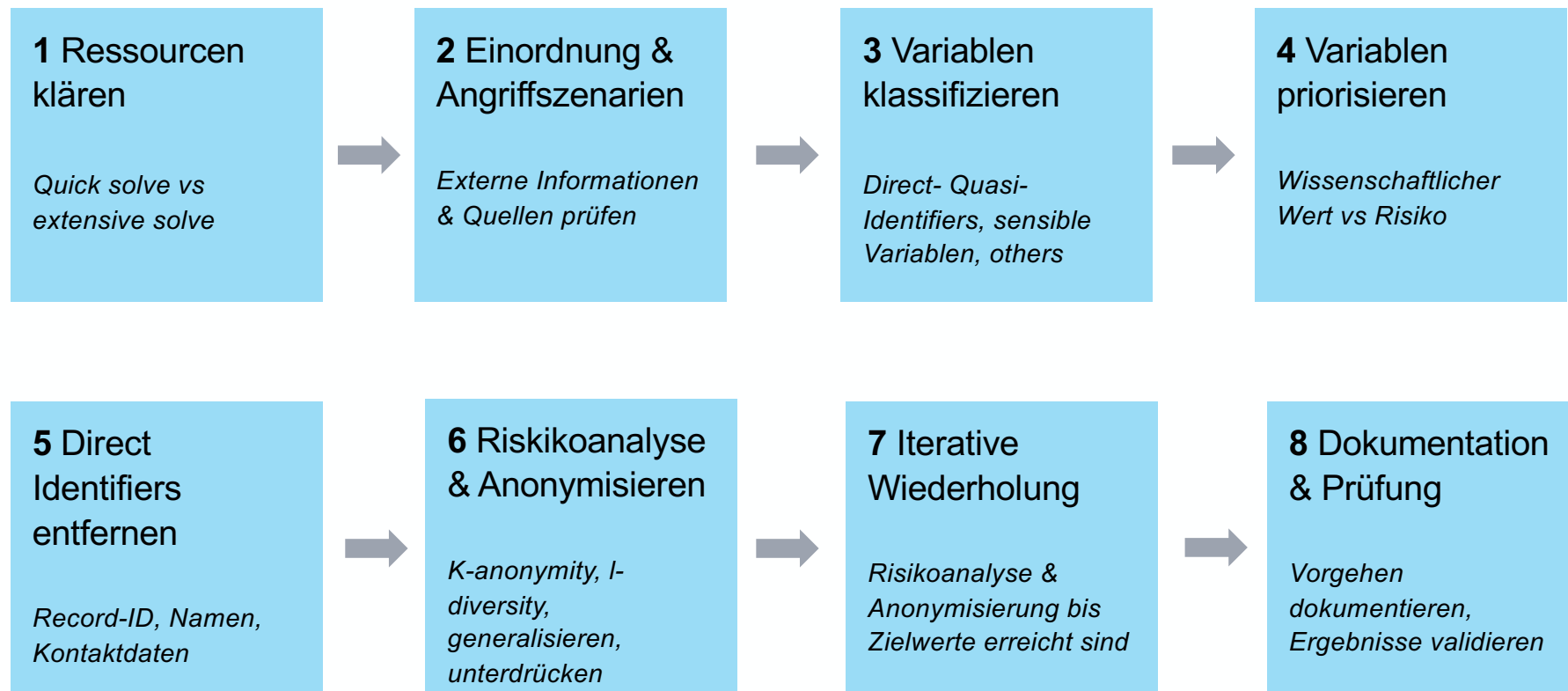
→ Quasi-Identifikatoren: Kanton + Partei + Amtsantritt-Kohorte

Kanton	Partei	Amtseintritt	Antwort F8
GR	SVP	2018 – 2020	zu tief
GR	SVP	2018 – 2020	zu tief
GR	SVP	2018 – 2020	zu tief
GR	SVP	2018 – 2020	zu tief

k = 4 (anonym) Person nicht eindeutig identifizierbar. Jedoch: **l = 1** (nicht divers!)

- k-anonymity schützt vor identity disclosure, l-diversity vor attribute disclosure
- Voraussetzung: zur statistischen Risikoanalyse (z.B. via sdcMicro) müssen sensitive identifiers als solche gekennzeichnet werden (hier Antwort zu Frage 8)
- Äquivalenzgruppen mit identischen sensitiven Attributen identifizieren (hier: Kanton x Partei x Amtseintritt)
- Post Randomization oder Quasi-Identifiers vergrößern (bis $l \geq 2$)

Workflow



ZHAW Services Forschungsdaten

researchdata@zhaw.ch

Unterstützung bei der
Datenanonymisierung durch ZHAW
Services Forschungsdaten

- Informieren
- Support
- Tool (?)
- Review

