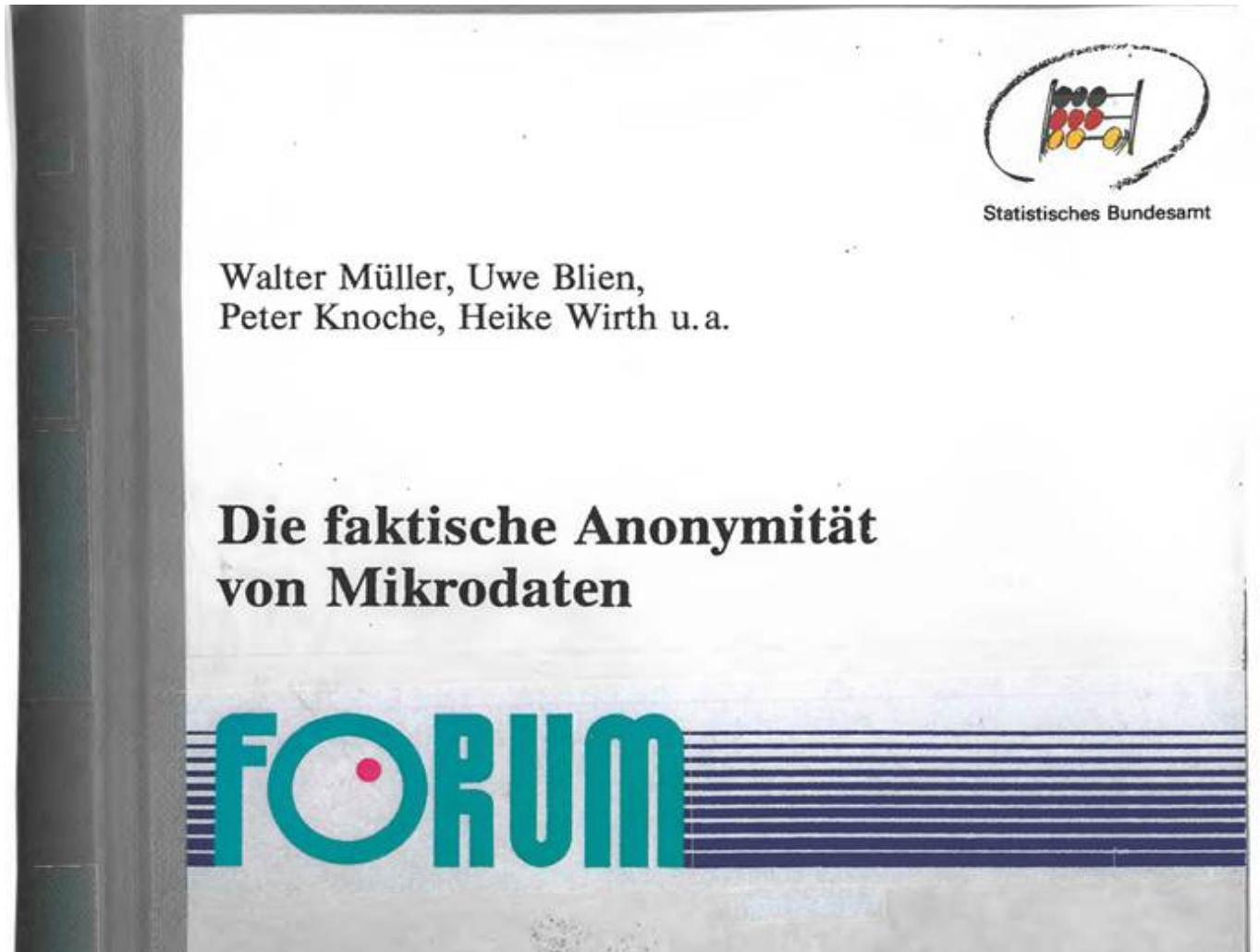

Maßnahmen zur Erzielung von faktischer Anonymität in ausgewählten Datenbeständen

Workshop ‚Scientific Use Files aus verknüpften Sozialdaten‘
SOEP im DIW Berlin, 17.-18.11.2015

Klaudia Erhardt



Allgemeine Datenschutzvorkehrungen

- Vertragliche Bindung des Datenempfängers
- Geheimhaltung der lokalen Umsetzung der Stichprobenpläne
- Systemfreie Daten-Anordnung im File

Empfehlungen für den Mikrozensus

Mikrozensus Grundfile	Mikrozensus Regionalfile
70% Substichprobe	85% Substichprobe
Kleinste Regionaleinheit mindestens 500.000 (400.000) Einwohner	Kleinste Regionaleinheit mindestens 100.000 Einwohner
„Brisante“ Merkmale:	
jede ausgewiesene Kategorie muss in der Grundgesamtheit mind. 50.000 Fälle umfassen	jede ausgewiesene Kategorie muss in der Grundgesamtheit mind. 50.000 und im Mikrozensus-Originalfile mind. 3 Fälle pro Regionaleinheit umfassen.
Alle anderen Merkmale: jede ausgewiesene Kategorie muss in der Grundgesamtheit mind. 5.000 Fälle umfassen	

„Brisante Merkmale“

Mikrozensus Grundfile	Mikrozensus Regionalfile
<ul style="list-style-type: none">• Staatsangehörigkeit	<ul style="list-style-type: none">• Staatsangehörigkeit• Beruf• Wirtschaftszweig• Alter

Datenbestand	(Sub)-Stichprobe
Mikrozensus Grundfile	0,70% der Wohnbevölkerung in Deutschland
Mikrozensus Regionalfile	0,85% der Wohnbevölkerung in Deutschland
SIAB Regionalfile	2% d. Originaldaten (Vollerhebung der Zielpopulation)
DRV (zahlreiche SUFs)	Zw. 1% u. 35% der Originaldaten (Vollerh. d. Zielpop.)
NEPS	3% der Zielpopulation
DZA Freiwilligensurvey	0,03% der Zielpopulat. (Wohnbevölk. DE ab 14 Jahren)
DZA Alterssurvey	0,022% der Zielpopulation (Wohnbevölkerung 40 u. mehr Jahre alt)
SOEP	0,045% der Privathaushalte in Deutschland

Datenbestand	Regionalangaben
Mikrozensus Grundfile	500.000 (400.000)er Regel
Mikrozensus Regionalfile	100.000er Regel
SIAB Regionalfile	100.000er Regel
DRV (zahlreiche SUFs)	Bundesland, bei Einhaltung der Gruppengröße für brisante/nicht brisante Merkmale max. Tiefe: Kreise
NEPS	Bundesland
DZA Freiwilligensurvey	Bundesland plus Indikator Kreistyp
DZA Alterssurvey	Bundesland plus Indikatoren Kreistyp, systemfreie Kreiskennziffer
SOEP	Bundesland; mit erweit. Datenschutzkonzept Regional- indikatoren (GGK, BIK-Regionen, 97 Raumordn.reg.)

Datenbestand	Staatsangehörigkeit u.a. Länderangaben
Mikrozensus Grundfile	50.000er Regel
Mikrozensus Regionalfile	50.000er Regel & mind. 3 Fälle pro Regionaleinheit
SIAB Regionalfile	Staatangehörigkeit dichotom (deutsch/nicht deutsch)
DRV (zahlreiche SUFs)	50.000er Regel
NEPS	17 Kategorien (Länder), 21 Kategorien (Sprachen)
DZA Freiwilligensurvey	Nur deutsch/nicht deutsch erhoben
DZA Alterssurvey	Mind. 3 Fälle pro Kategorie, sonst Zus.fassung
SOEP	Keine Vergrößerung

Datenbestand	Welche Merkmale gelten darüberhinaus als brisant u. werden besonderen Maßnahmen unterzogen?
Mikrozensus Grundfile	Keine
Mikrozensus Regionalfile	Beruf / Wirtschaftszweig / Alter (50.000er Regel & mind. 3 pro Regionaleinheit)
SIAB Regionalfile	Beruf / Wirtschaftszweig / Alter (50.000er Regel & mind. 3 pro Regionaleinheit)
DRV (zahlreiche SUFs)	Alter: datensatzspez. Altersbegrenzung an den Rändern
NEPS	Zahl d. Angestellten bei Selbständigen: Top Coded ab 20
DZA Freiwilligensurvey	u.a. seltene Geburtsjahrgänge, Ausübungsjahre d. Engagements, Zuzugsjahr nach Deutschland
DZA Alterssurvey	u.a. seltene Jahresangaben, Top-Coding HH-Größe, Kinder- und Geschwisterzahl
SOEP	Keine Vergrößerungen

Datenbestand	Maßnahmen für „alle übrigen Merkmale“
Mikrozensus Grundfile	5.000er Regel (univariat)
Mikrozensus Regionalfile	5.000er Regel (univariat)
SIAB Regionalfile	5.000er Regel (univariat)
DRV (zahlreiche SUFs)	5.000er Regel (univariat)
NEPS	keine
DZA Freiwilligensurvey	keine
DZA Alterssurvey	keine
SOEP	keine

Auf welche Risikokonstellation sind die Maßnahmen von Müller et al. gemünzt?

- Die gesuchte Person gehört einer kleinen, abgrenzbaren Subpopulation an (sachliche Tiefengliederung)
- Regionale Tiefengliederung im File vorhanden
- Ein Forscher weiß, dass ein ganz bestimmter Angehöriger der Subpopulation im File enthalten ist
- Die Merkmale dieser Person sind in der Weise im File abgebildet, wie es der Forscher vermutet

Diese Bedingungen müssen alle gleichzeitig zutreffen...

Auf welche Risikokonstellation sind die Maßnahmen von Müller et al. gemünzt?

„Das gleichzeitige Zusammentreffen aller Bedingungen kann bei Stichprobenerhebungen als außergewöhnlich seltenes Ereignis betrachtet werden.“ (Müller et al. 1001: 388)

...und außerdem „muss vorausgesetzt werden, daß ein Datenangreifer ein subjektives Interesse daran hat, das die denkbaren Kosten der Konsequenzen des Angriffs (...) übersteigt.“ (Müller et al. 1991: 387, FN 2)

Auf welche Risikokonstellation sind die Maßnahmen von Müller et al. gemünzt?

Ein Deanonymisierungsversuch unter dieser Risikokonstellation könnte durch wissenschaftliches Personal nur aufgrund einer außerberuflichen Motivlage und grob vertragswidrigen Handelns stattfinden.

Vielen Dank für Ihre
Aufmerksamkeit

Klaudia Erhardt kerhardt@diw.de