

## 1. Allgemeine Hinweise zur Anonymisierung des Remote Access SUF (Berichtsjahre bis 2019)

Das Datenprodukt Remote Access SUF (Remote SUF) ordnet sich hinsichtlich des Analysepotentials und des Anonymisierungsgrades zwischen den bestehenden Produkten Off-Site-SUF und On-Site-Daten ein. Wie beim Off-Site-SUF handelt es sich auch beim Remote SUF um faktisch anonyme Daten. Solche Daten dürfen gemäß den Vorgaben des Bundesstatistikgesetzes grundsätzlich an wissenschaftliche Einrichtungen übermittelt werden (vgl. BStatG § 16 (6)). Faktisch anonym heißt, dass die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können.

## 2. Anonymisierungsmaßnahmen am Datenmaterial

Die Basis für den Remote SUF bilden die formal anonymen Daten des Mikrozensus (On-Site-Daten<sup>1</sup>). Um faktische Anonymität zu erreichen, werden am Remote SUF folgende Maßnahmen ergriffen:

### 2.1. Entfernung räumlicher Informationen

Die tiefste ausgewiesene räumliche Gliederung der Daten bilden die Anpassungsschichten. Die Anpassungsschichten sind in der Regel kreisscharf und bestehen aus einem oder mehreren meist benachbarten Kreisen oder kreisfreien Städten mit durchschnittlich 500.000 Einwohnern. Deutschlandweit gibt es aktuell 147 Anpassungsschichten.

Die notwendigen Zusammenfassungen unterscheiden sich damit in Abhängigkeit der Einwohnerzahl der Kreise und kreisfreien Städte. Das verdeutlicht ein Vergleich einiger Bundesländer: So gibt es in Nordrhein-Westfalen bspw. viele Kreise oder kreisfreie Städte, die eine eigene Anpassungsschicht bilden. Insgesamt werden aus den 53 Kreisen und kreisfreien Städten 34 Anpassungsschichten gebildet. In Rheinland-Pfalz hingegen verteilen sich die 36 Landkreise und kreisfreien Städte auf nur fünf Anpassungsschichten. Das Saarland bildet eine einzige Anpassungsschicht, das Land Bremen besteht aus zweien (Bremen und Bremerhaven). In den weiteren Stadtstaaten bilden grundsätzlich die Bezirke die Anpassungsschichten<sup>2</sup>. Der Remote SUF bietet damit deutlich mehr regionales Analysepotential als der Off-Site-SUF – der maximal bis zur Ebene der Bundesländer ausgewertet werden kann – und kann tw. auch als Ersatz für On-Site-Projekte mit regionalem Fokus genutzt werden.

Alle räumlichen Informationen, die unterhalb der Anpassungsschichten liegen oder in Kombination mit der Anpassungsschicht Rückschlüsse auf eine darunterliegende Regionalinformation geben, wurden aus den Daten entfernt. Hierzu zählen Regions- und Kreistypen, Gemeindegrößenklassen, Arbeitsmarktregionen, Arbeitsagenturbezirke, Gemeindetyp und die Stadt-Land-Gliederung nach Eurostat.

### 2.2. Vergrößerungen

Zusätzliche inhaltliche Eingriffe in das Datenmaterial bilden einen weiteren Schritt, um faktische Anonymität zu erreichen. Das Vorgehen ist hierbei von zwei Überlegungen geleitet. Zum einen sollen im Datensatz grundsätzlich keine Merkmalsausprägungen vorhanden sein, die bei einfacher

---

<sup>1</sup> Grundsätzliche Informationen zu den On-Site-Daten des Mikrozensus finden sich in den Metadatenreports zur Statistik (Teil I) und zum Produkt (d.h. zum jahresspezifischen Datenangebot des Mikrozensus, Teil II): <https://www.forschungsdatenzentrum.de/de/haushalte/mikrozensus>

<sup>2</sup> In Hamburg sind Harburg und Bergedorf zu einer Anpassungsschicht zusammengefasst.

Auszählung für die Stichprobe ein  $n < 3$  aufweisen. Zum anderen muss in Betracht gezogen werden, dass bestimmte Merkmale in der im Mikrozensus vorliegenden Form das Risiko einer Deanonymisierung stärker als andere erhöhen. Als Beispiele können hier die Angaben zur Nationalität oder auch Informationen zu den Berufen genannt werden.

Bei den Variablen, die Merkmalsausprägungen mit lediglich einem oder zwei Fällen aufweisen, handelt es sich im Wesentlichen um Staatsangehörigkeiten, Herkunftsländer, Berufs-, Hauptfachrichtungs- oder Wirtschaftszweiginformationen sowie Jahresangaben (siehe auch die Spalte „Maßnahme im Remote SUF“ im Schlüsselverzeichnis<sup>3</sup>). In vielen Fällen sind dies auch Merkmale, die ein höheres Deanonymisierungsrisiko mit sich bringen. Diese Merkmale werden daher durch Zusammenfassungen vergrößert, wobei sich das Vorgehen im Wesentlichen am Vorgehen beim bisherigen Off-Site-SUF orientiert. D.h. bei den aufgeführten Variablen wird die gesamte Variable auf hochgerechnet 5.000 Fälle vergrößert, im Falle der Angaben zu Nationalitäten auf 50.000 Fälle hochgerechnet auf die Grundgesamtheit<sup>4</sup>. Bei den Berufen wird auf einen [künstlichen Viersteller zurückgegriffen](#), der auch im Off-Site-SUF verwendet wird und das Anforderungsniveau (5.Stelle der KldB 2010) soweit wie möglich erhält.

In einzelnen Fällen, die nur in den Randbereichen zu Einzelwerten oder sehr geringen Fallzahlen führen, wird ein Bottom- oder Top-Coding angewendet.

### **3. Unterschiede zwischen Remote SUF und On-Site-Material aufgrund der Anonymisierungsmaßnahmen**

Im Schlüsselverzeichnis des Remote SUF ist über die Spalte „Maßnahme im Remote SUF“ grundsätzlich erkennbar, wie mit jeder der aufgeführten On-Site-Variablen verfahren wird. So lässt sich erkennen, ob für ein Projekt der Remote SUF infrage kommt, oder ob ein Rückgriff auf das On-Site-Material notwendig ist. Relevante Unterschiede sind hier überblicksartig zusammengefasst.

#### **3.1. Umfang des Datenmaterials**

Die regionale Tiefe ist im Vergleich zum On-Site-Material eingeschränkt. Dies betrifft wie oben beschrieben die Variablen mit Regionalinformationen unterhalb der Anpassungsschichten. Einige Variablen sind darüber hinaus nicht im Remote SUF aber im On-Site-Material enthalten:

- 5-Steller der Berufsgattungen
- die Hauptfachrichtungen (7-Steller und Stellen 5-7)
- 2-Steller Meisterausbildung
- Hinterbliebenenrente

Infolge von notwendigen Vergrößerungen sind einige Originalvariablen aus den Daten entfernt worden, da sie sonst Vergrößerungen aufheben würden. Diese Variablen können grundsätzlich aus den

---

<sup>3</sup> Hier herrscht auch über die Erhebungsjahre weitgehende Kontinuität bei den kontinuierlich abgefragten Variablen.

<sup>4</sup> Wie genau im jeweiligen Berichtsjahr bei den Variablen, die vergrößert werden müssen, zusammengefasst wird, lässt sich aus dem Datenhandbuch zum Off-Site-SUF erkennen. Die Datenhandbücher finden sich auf den Seiten der Forschungsdatenzentren (hier das Beispiel für den MZ 2019):

<https://www.forschungsdatenzentrum.de/de/10-21242-12211-2019-00-00-3-1-0>

Im Schlüsselverzeichnis zum Remote SUF sind die Variablen oder die betroffenen Bereiche blau hervorgehoben.

vergrößerten Daten selbst generiert werden. Es ist dabei zu beachten, dass sich durch die Vergrößerungen in Einzelfällen auch auf den ersten Stellen Abweichungen zu den Originaldaten ergeben.

- 1- bis 3-Steller bei den Berufen
- 1- bis 3-Steller bei den Berufsgattungen

Im On-Site-Material sind teilweise Merkmale zu Haupteinkommensbeziehern oder den Bezugspersonen in eigenständigen Variablen vorhanden, die aber aus dem Datenmaterial selbstständig generiert werden können. Diese werden im Remote SUF nicht bereitgestellt.

### **3.2. Vergrößerungen**

Quantitativ betrachtet ist nur ein kleiner Teil von wenigen Prozent der im Remote SUF vorhandenen Variablen von Vergrößerungen nach dem oben beschriebenen Vorgehen betroffen. Die Vergrößerungen konzentrieren sich auf folgende inhaltliche Bereiche:

- Berufe und Berufsgattungen
- Wirtschaftszweige
- Hauptfachrichtungen
- Arbeitsstunden
- Jahresangaben (z.B. Zuzugsjahre, Geburtsjahr)
- Staatsangehörigkeiten und Geburtsländer