

Vertrauliche Daten sicher und flexibel mit der Cell-Key-Methode veröffentlichen

Matthias Rumpf

Data Week Leipzig 2026

Das Problem

- **Verwaltungsdaten** mit großem gesellschaftlichem Potenzial bleiben aufgrund des Datenschutzes weitgehend unzugänglich
 - **Starre Aggregate** schützen zwar Einzelwerte, verhindern aber flexible und detaillierte Analysen
 - **Differenzierungsangriffe** machen selbst aggregierte Veröffentlichungen riskant – Zellsperren nur schwer umzusetzen
 - **Forschungsdatenzentren** sind die bisherige Lösung – aber teuer, bürokratisch und nur für wenige zugänglich
- ☐ **Ich stelle einen Ansatz vor**, der eine flexible, sichere Datenveröffentlichung ermöglicht – und den Verwaltungsaufwand deutlich senkt

Das Differenzierungsproblem – Cell-Key-Perturbation als Lösung

Mikrodaten (anonymisiert)

Nr	Alter	Beruf	Stadt	Einkommen
1	34	Lehrer	Berlin	41.000 €
2	36	Lehrerin	Berlin	37.000 €
3	35	Lehrer	Berlin	42.000 €
4	62	Managerin	Berlin	100.000 €

Veröffentlichte Gesamttabelle (Durchschnittseinkommen)

Beruf	Anzahl	Durchschnittseinkommen
Lehrkräfte	3	40.000
Führungskräfte	1	unterdrückt
Gesamt	4	55.000 €

Rekonstruktion durch Differenzierung:

Gesamtsumme: $4 \times 55.000 \text{ €} = 220.000$

Lehrkräfte: $3 \times 40.000 \text{ €} = 120.000 \text{ €}$

Führungskraft: $220.000 \text{ €} - 120.000 \text{ €} = 100.000 \text{ €}$

1 Zuweisung von Datensatzschlüsseln

Jeder Mikrodatensatz wird ein geheimer Wert zugewiesen. Dieser Schlüssel wird niemals veröffentlicht.

2 Zellschlüssel berechnen

Für jede Abfragezelle werden die Schlüssel der zugehörigen Datensätze addiert; daraus ergibt sich der Zellschlüssel.

3 Zuweisung der Perturbation (p-Tabelle)

Zellschlüssel wird deterministisch einem kleinen Rauschwert (zb -2 bis +2) zugeordnet. Gleiche Datensätze → immer gleiches Rauschen.

4 Veröffentlichung des perturbierten Werts

Veröffentlichter Wert = tatsächlicher Wert + Rauschen. Unterschiede zwischen zwei Abfragen ebenfalls verrauscht – Rekonstruktion schlägt fehl.

Beruf	Anzahl	Veröffentlichter Durchschnitt
Lehrkräfte	3	40.700 €
Führungskräfte	1	unterdrückt
Gesamt	4	54.200 €

Rekonstruktion fehlgeschlagen:

Gesamtsumme: $4 \times 54.200 \text{ €} = 216.800$

Lehrkräfte: $3 \times 40.700 \text{ €} = 122.100 \text{ €}$

Führungskraft: $216.800 \text{ €} - 122.100 \text{ €} = 94.700 \text{ €}$

Perturbationsverfahren

- **Zellschlüssel-Störung** bei der Häufigkeitsverteilung
- Unabhängige **Perturbation von Nullwerten**
- Erhält **struktureller Nullwerte**

Weitere Informationen unter

Stephanie Blanchard: “[The methodological challenges of protecting outputs from a Flexible Dissemination System](#)”

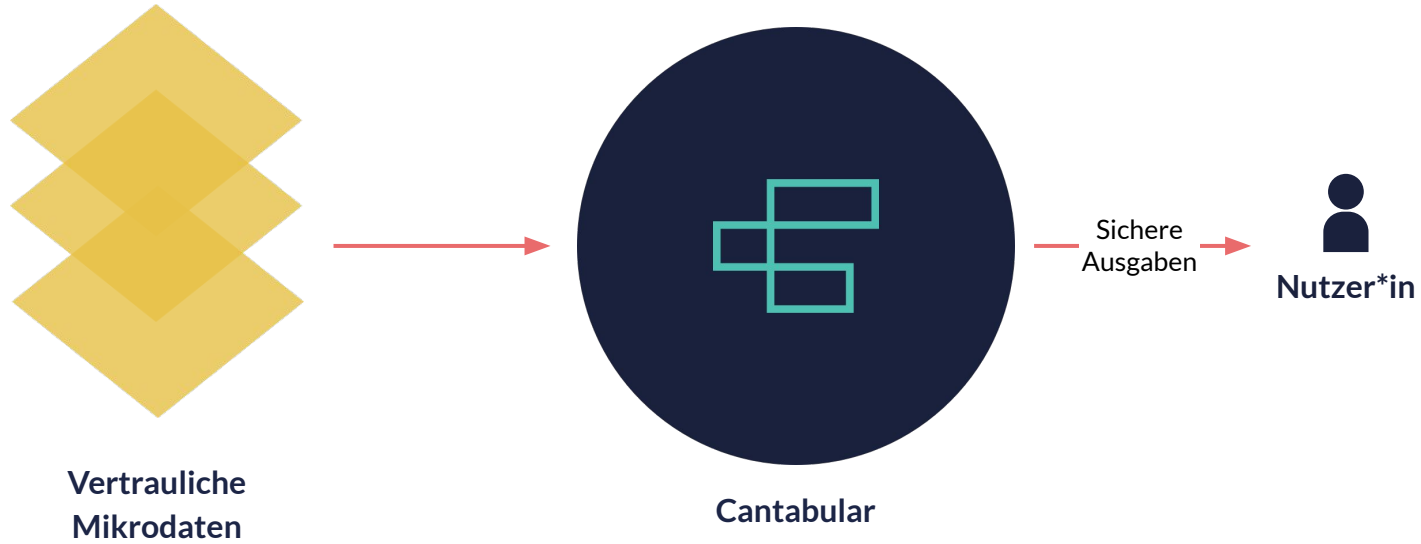
Weitere Kontrollmöglichkeiten I

- **Maximale Variablenanzahl:** Blockiert Abfragen, die zu sehr geringen Fallzahlen führen.
- **Maximale Anzahl an Zellen:** Blockiert Abfragen, die zu viele Beobachtungen enthalten.
- **Sensible Variablen:** Abfragen bei sensiblen oder sehr detaillierten Variablen einschränken.

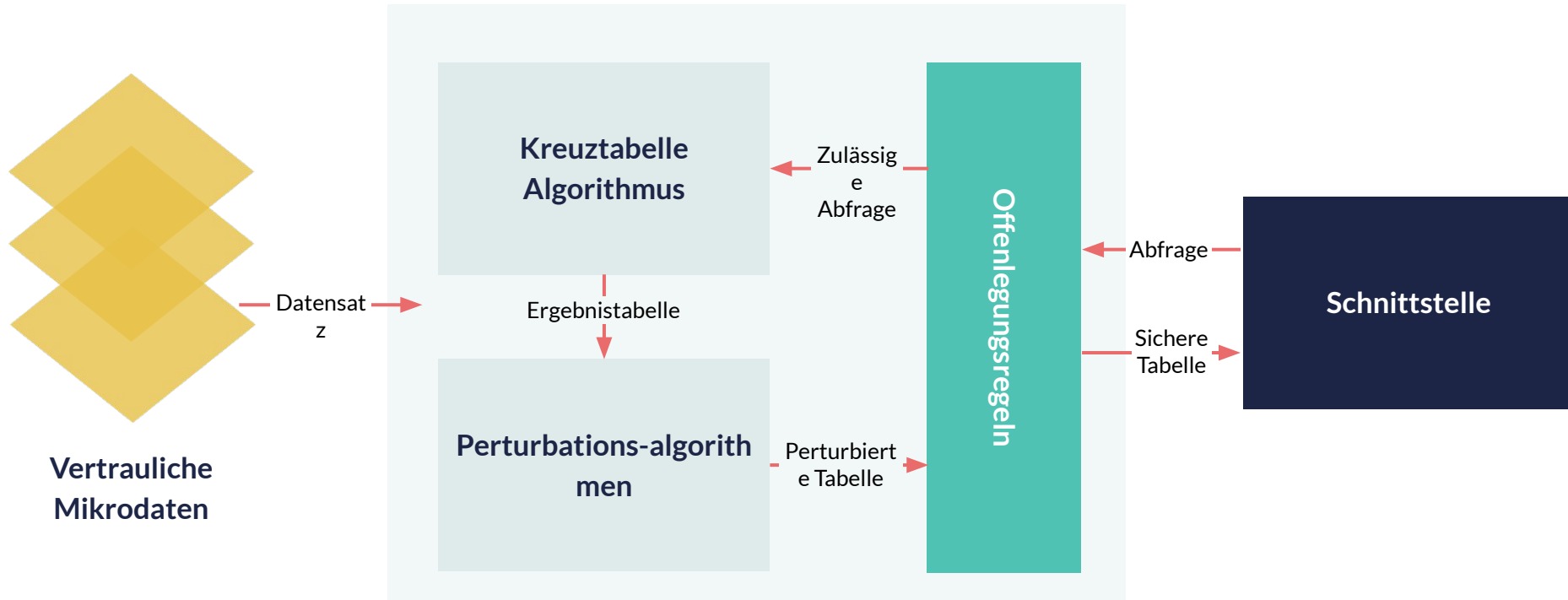
Weitere Kontrollmöglichkeiten II

- **Attribute:** Die Offenlegung einzelner oder gruppierter Attribute in einer Tabelle kann erkannt und unterdrückt werden.
- **Einzelwerte:** Tabellen, die zu viele Werte von „1“ enthalten, können ebenfalls blockiert werden.
- **Sparsität und Dominanz:** Blockiert Tabellen, die proportional zu viele Nullen enthalten oder zu stark von einer Kategorie bestimmt sind.

So funktioniert es



So funktioniert es (unter der Haube)



Die wichtigsten Funktionen

- Erstellung von beliebig vielen Tabellen in Echtzeit aus Mikrodaten
- Automatisierte Output-Kontrolle unter Verwendung von Zellschlüssel-Perturbation in Kombination mit anderen Methoden wie der Zellunterdrückung
- Metadaten, die bei der Berechnung einer Tabelle zusammengestellt werden

Demo

Flexibler Tabellen-Generator zum britischen Zensus 2021

 Office for
National Statistics

English (EN) | [Cymraeg \(CY\)](#)

[Release calendar](#) | [Methodology](#) | [Media](#) | [About](#) | [Blog](#)

Home | Business, industry and trade | Economy | Employment and labour market | People, population and community | Taking part in a survey?

Search for a keyword(s) or time series ID



[Home](#) > [Census](#)

Create a custom dataset

We group Census 2021 data together based on who or what the information is about, for example, people or households. We make population types from these groups or subsets of them. For example, people who are usually resident in England or Wales make up the population type usual residents. [Read about the measurements we used for Census 2021 data.](#)

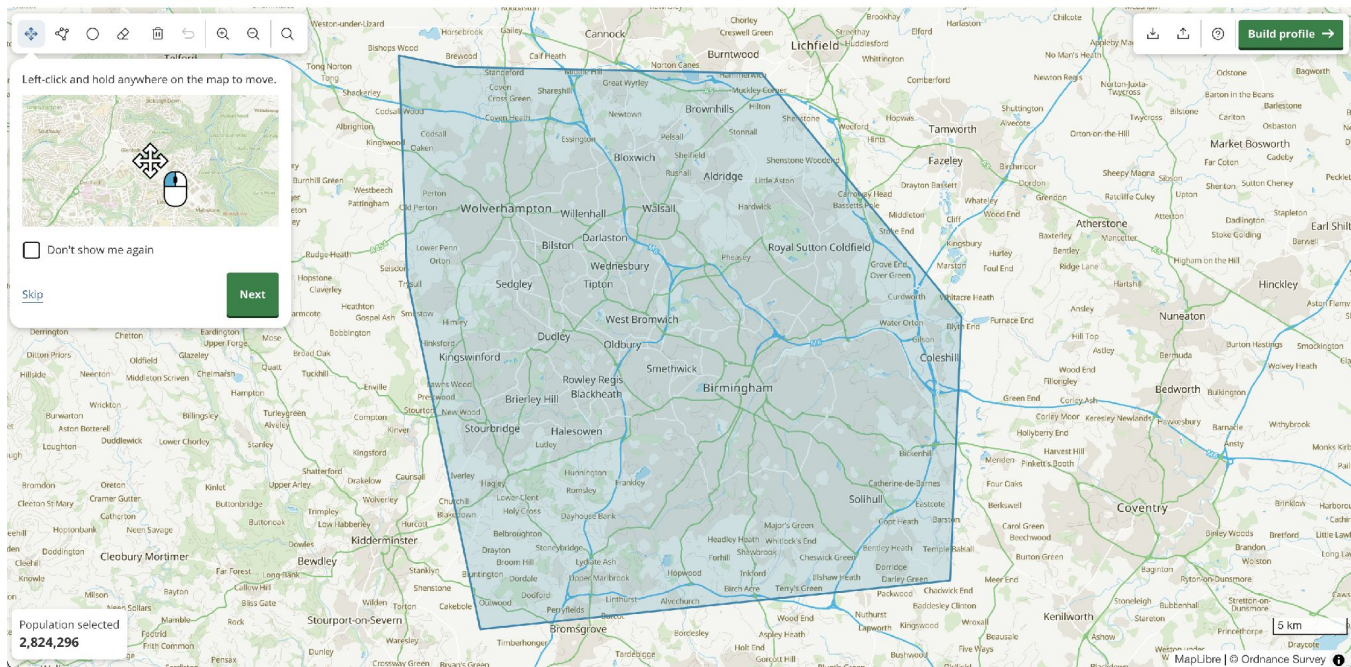
Select population type

- ☐ All households
Either one usual resident living alone or a group of people who share cooking and living facilities, where that group includes at least one usual resident.
- ☐ All Household Reference Persons
A person who serves as a reference point, mainly based on economic activity, to characterise a whole household.

<https://www.ons.gov.uk/datasets/create>

Benutzerdefinierte Gebietsauswahl

Office for National Statistics



<https://www.ons.gov.uk/visualisations/customprofiles/draw/>

Benutzeroberfläche – Tabellen

- Filtern aus Tabellen wird unterstützt, eine Sortierung kann implementiert werden
- Tabellen können auf Basis von Kombinationen aus Variablen eines Datensatzes erstellt werden.
- Pivot-Tabellen werden unterstützt
- Download in XLSX, CSV und CSV (flach) wird unterstützt, PDF möglich

Benutzeroberfläche – Karten

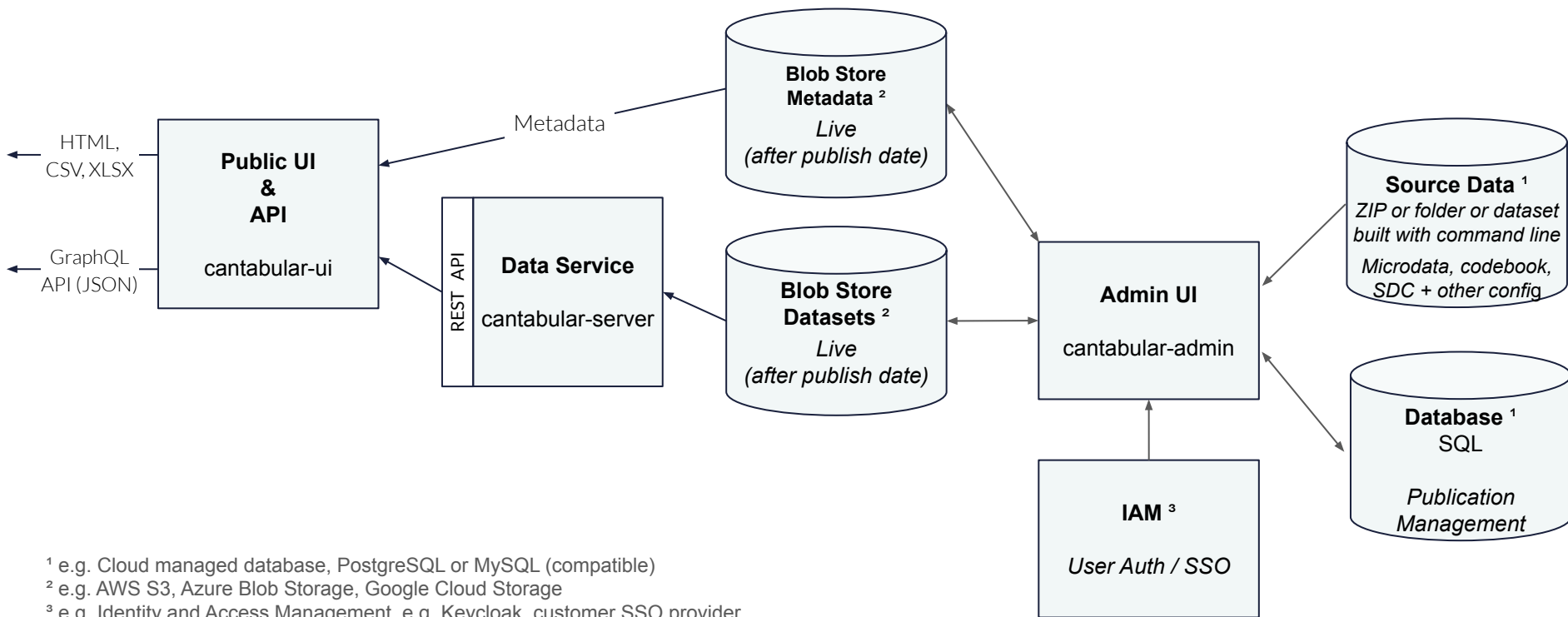
- Kartenbasierte Benutzeroberfläche mit Unterstützung für GeoJSON und Shapefiles
- Kartenbasierte Auswahl geografischer Gebiete, Anzeige von Informationen basierend auf dem ausgewählten Gebiet
- Die aktuelle Benutzeroberfläche wurde mit Mapbox erstellt, andere Anbieter sind möglich, einschließlich ESRI

Benutzeroberfläche – Grafiken

- Alle grundlegenden Chart-Typen werden unterstützt und können in die Benutzeroberfläche integriert werden
- Chart-Typen unterstützen Tooltips sowie das Hinzufügen und Entfernen von Kategorien
- Daten können direkt aus Grafiken heruntergeladen werden

IT-Architektur

Cantabular Architecture



Vielen Dank

matthias.rumpf@gmx.de