

Delivering more data, faster, with automated safe data publishing

Matthias Rumpf

IFC BdP Workshop on data security and privacy
Lisbon 14-15 September 2026

The Differentiation Problem – Cell Key Perturbation as a solution

Microdata (de-identified)

Rec .	Age	Occupation	City	Income
1	34	Teacher	Berlin	€41,000
2	36	Teacher	Berlin	€37,000
3	35	Teacher	Berlin	€42,000
4	62	Manager	Berlin	€100,000

Published aggregate table (average income)

Occupation	Count	Avg. Income
Teacher	3	€40,000
Manager	1	suppressed
Total	4	€55,000

Disclosure by differentiation:

Total income: $4 \times €55,000 = 220,000$

Teachers: $3 \times €40,000 = €120,000$

Manager income: $€220,000 - €120,000 = €100,000$

1 Assign record keys

Each microdata record receives a secret random number. These keys are never published.

2 Compute cell key

For any query cell, the record keys of contributing records are summed; the fractional part becomes the cell key.

3 Look up perturbation (p-table)

The cell key maps deterministically to a small noise value (typically -2 to $+2$). Same records $\rightarrow$ always same noise.

4 Publish perturbed value

Published figure = true value + noise. Differences between any two queries are also noisy — reconstruction fails.

Occupation	Count	Published Avg.
Teacher	3	€40,700
Manager	1	suppressed
Total	4	€54,200

Disclosure attempt fails:

Total income: $4 \times €54,200 = 216,800$

Teachers: $3 \times €40,700 = €122,100$

Manager income: $€216,800 - €122,100 = €94,700$

Some advantages

- No or less blocked cells in tables
- Results are consistent across different tables (provided the same p table parameters are used)
- Less effort for confidentiality checks
- No cross-table disclosure risks

Some disadvantages

- CKM tables no longer add up to summation tables
- Perturbation for small values can be very large
- Less intuitive than other anonymisation method

Additional Challenges

- Magnitude tables
 - Can be addressed when outliers are not too big
 - Winsorisation and other methods to provide additional layers of security
 - Time series
 - Differencing attacks on time series possible when same p-value is used
 - Can be addressed by combination of base and period-specific p-value
- ➔ Adds complexity and may limit data usability

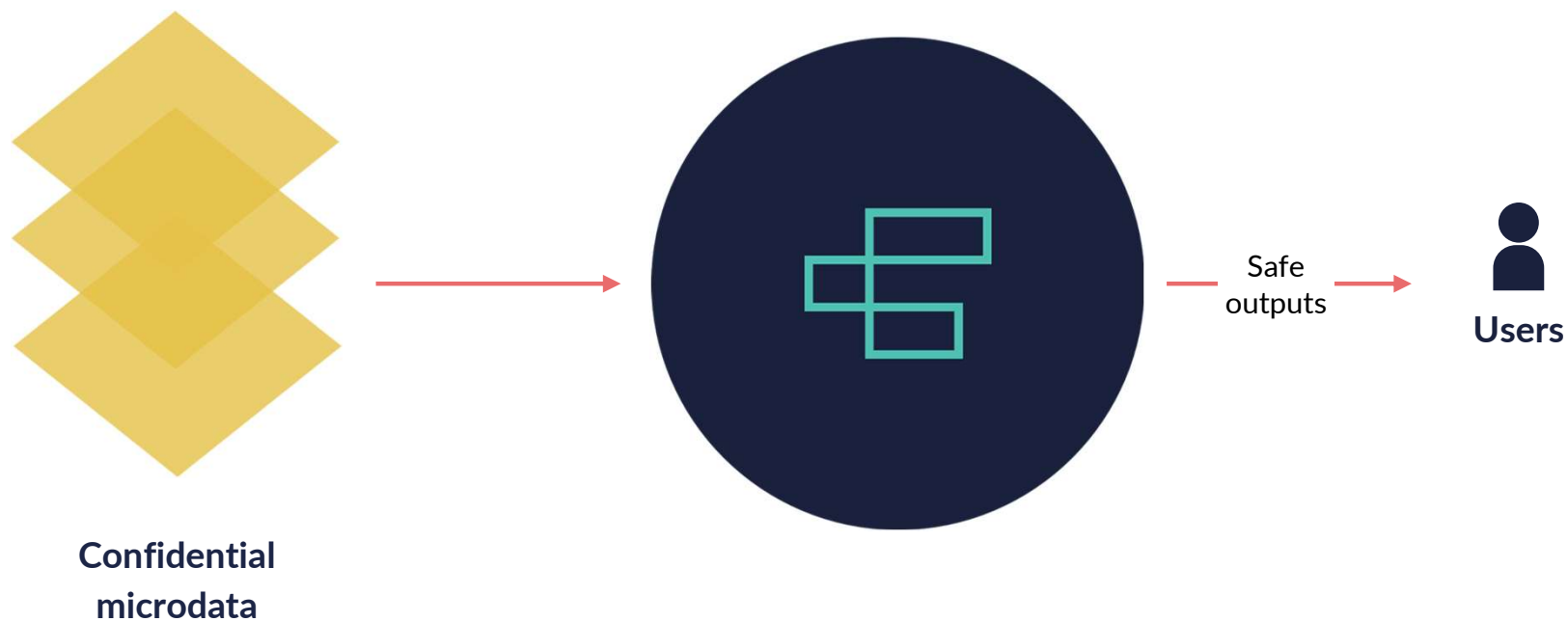
CKM today is widely used in the NSI world

- Several NSIs in the EU and beyond started to use CKM to publish census and other data (Australia, Germany, UK, Sweden)
- Standardisation and documentation through Eurostat and UNECE

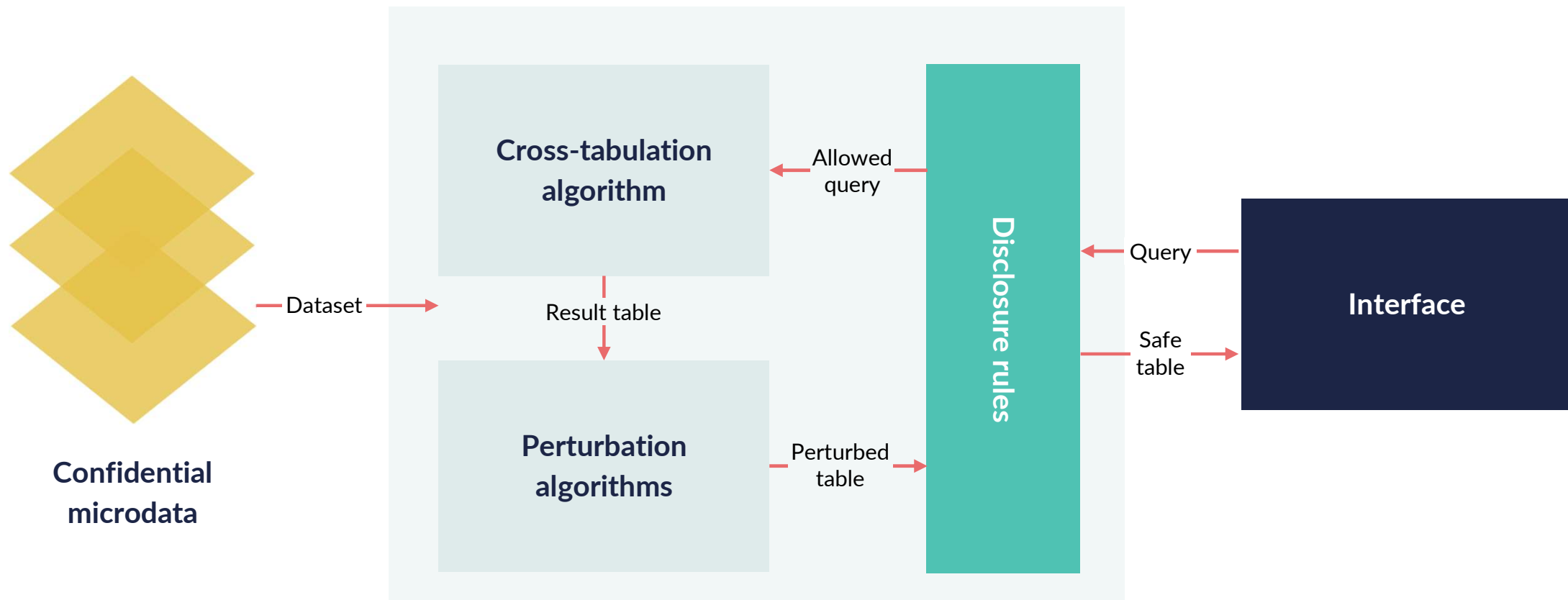
Innovations in the production process

- Most NSIs use CKM in traditional production of output table, where output become more granular and better secured against differencing attacks
 - In Germany, Destatis together with state-level and local authorities use CKM in a distributed system
- ➔ Can we take this a step further?

How it works



How it works (under the bonnet)



Flexible table builder from the UK census 2021

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.

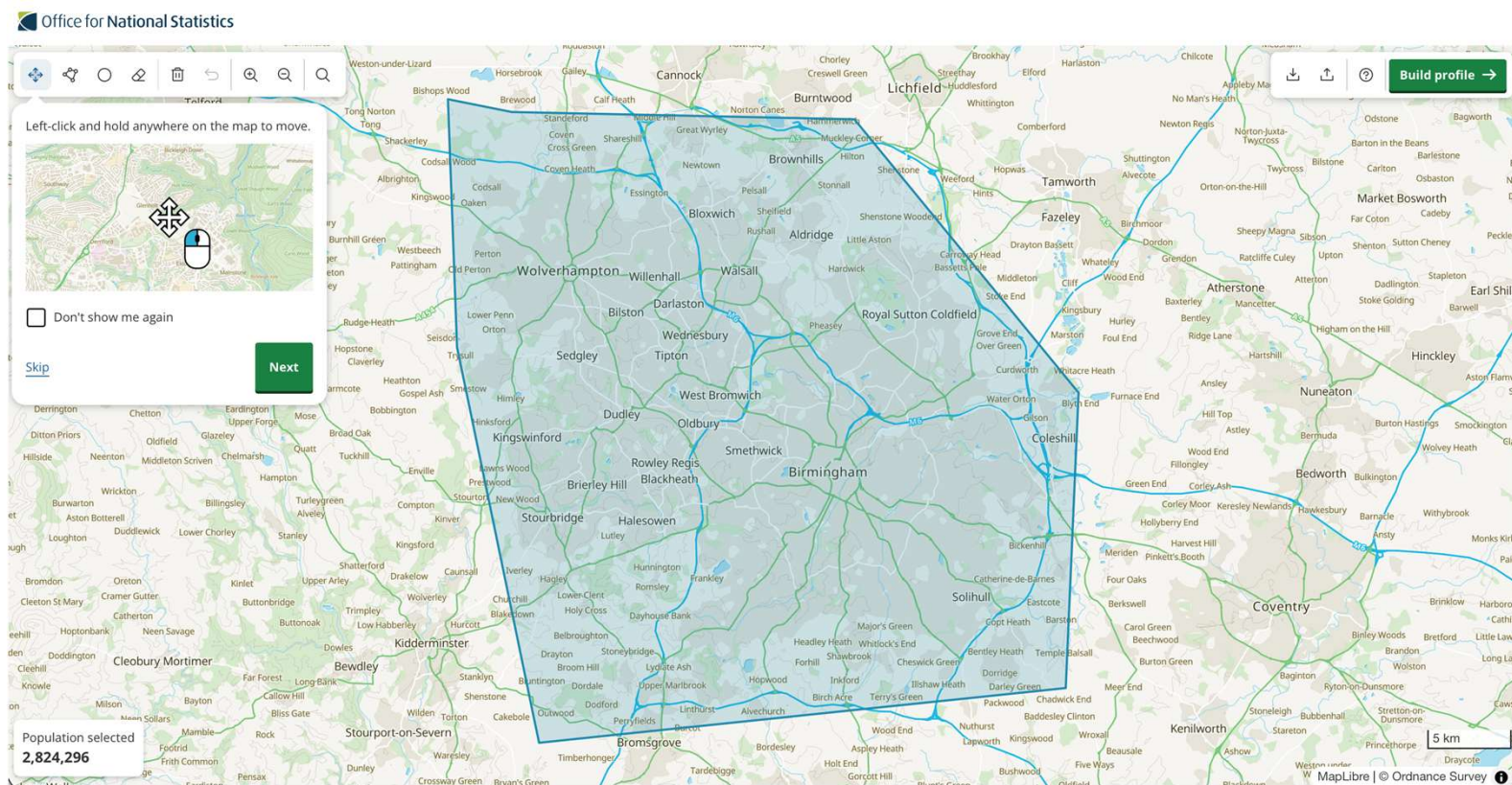


All Household Reference Persons

A person who serves as a reference point, mainly based on economic activity, to characterise a whole household.

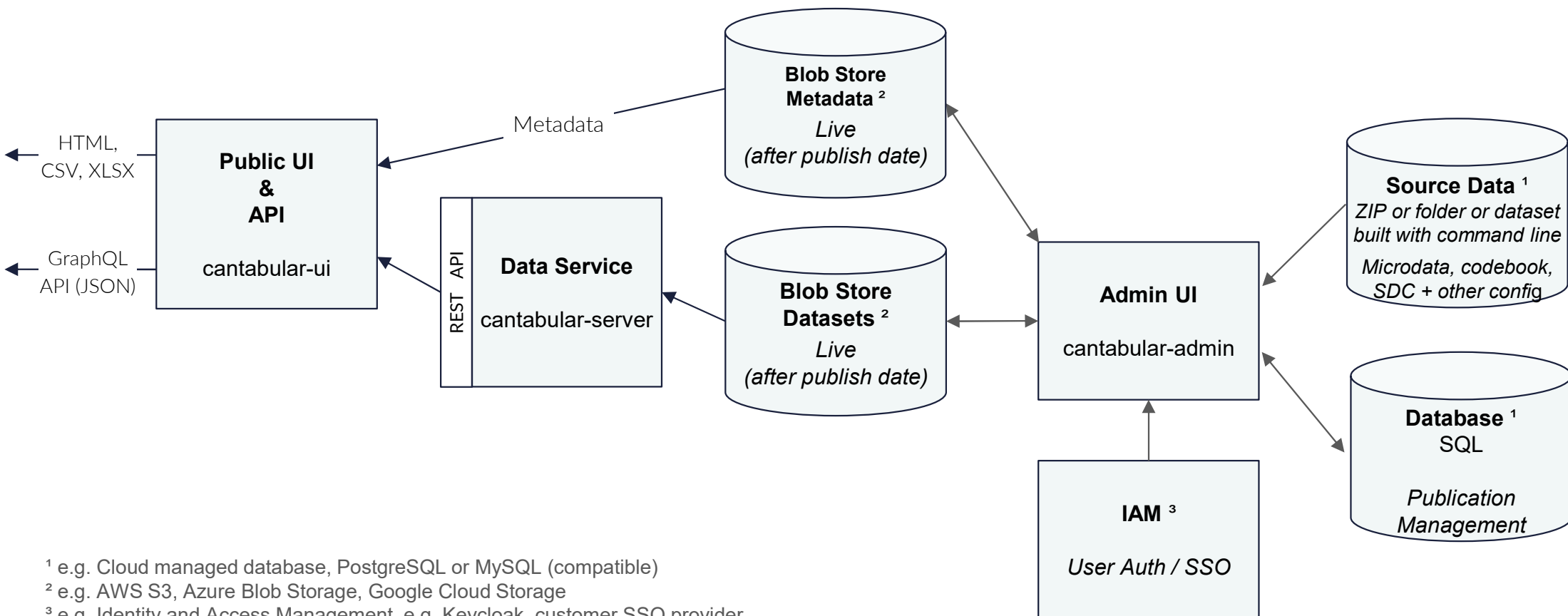
<https://www.ons.gov.uk/datasets/create>

Custom area selection from the UK census



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

Cantabular Architecture



CKM publication with flexible table builder

- More flexibility for users at lower costs; data access close to real-time
- CKM perturbed data could replace SUFs or sample data for RDCs
- Reduced exposure of sensitive data both for internal and external users while preserving usability
- Biggest potential with credit registries or other data focusing on consumers or NFCs

Thank you

matthias.rumpf@gmx.de