

Challenges of publishing with synthetic data

- but good for exploratory, teaching
- creating code

Next work with restricted access

Restricted access through VPN
strict rules around how data used.

Stages within this spectrum

- physical clean room, paper.
- tech workarounds e.g. VPN

restricted data like census

lightly redacted

- e.g. public data, facebook ad library.

Cutting costs of getting data

- private vs public.

Privacy concerns

But also competitiveness

e.g. Social media algorithms
and recommender systems

- bias

100,000 person panel -

Synthetic data
based on 1.3mil
devices

Chat prompts

X^{10M}
LLMs

Can sell in
format' data
from a commercial
perspective.

5 Safes (ONS)

- 1 Data
- 2 Projects
- 3 People
- 4 Settings
- 5 Outputs

- Consent (informed)

- Licensing - EUL.

- Perivision

Quali - Soc Science

Lessons learnt

(free text)

- cleanse + anonymise

Younger + SDDS

= severe + open.

OKPS - largely
Safeguarded