

Challenges of publishing with
synthetic data

- but good for exploratory,
teaching
- creating code.

next work with restricted secure
access

restricted access through VPN
strict rules around how data
used.

stages within this ^{broader spectrum}
- physical clean room, ^{over} laptop.

- tech workarounds e.g. VPN
restricted data like census

lightly redacted

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

~~Costs~~ 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 1.5mil
devices

Chat prompts
from
LLMs

Can sell raw
format data
from a commercial
perspective.

5 safes (ONS)

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

- Consent (informed)

- Licensing - EUL

- Penetration

Quali - soc science

Lessons learnt

(free text)

- clearance + anonymisation

Younger + SDDS

= secure + open.

UKDS - largely
safeguarded