Smoothing the way for secure data access using synthetic data

Emily Oliver, Head of Research and Capacity Building, ADR UK

Richard Welpton, Head of Data Service Infrastructure, ESRC

Would you buy this car?

The way data is accessed right now is like buying a used car, **BUT:**

- Not allowed to see it first, no pictures
- Not allowed a test drive
- No reviews anywhere about it
- Delivered in 2-6 months
- When you get it, it might be missing parts

You also need to:

- Justify why you need the car and specify the journeys you will take

What are the advantages of synthetic forms of secure data?

Researchers:
- Allows researchers to assess data before applying to use it
- **Explore** whether the real dataset **could** answer the research question
- Can be used to develop and test code

Secure Data Environments and Data Owners:
- Higher quality applications to process
- Reduced demand on computational resources
- Reduced staff time undertaking statistical disclosure control

Future users:
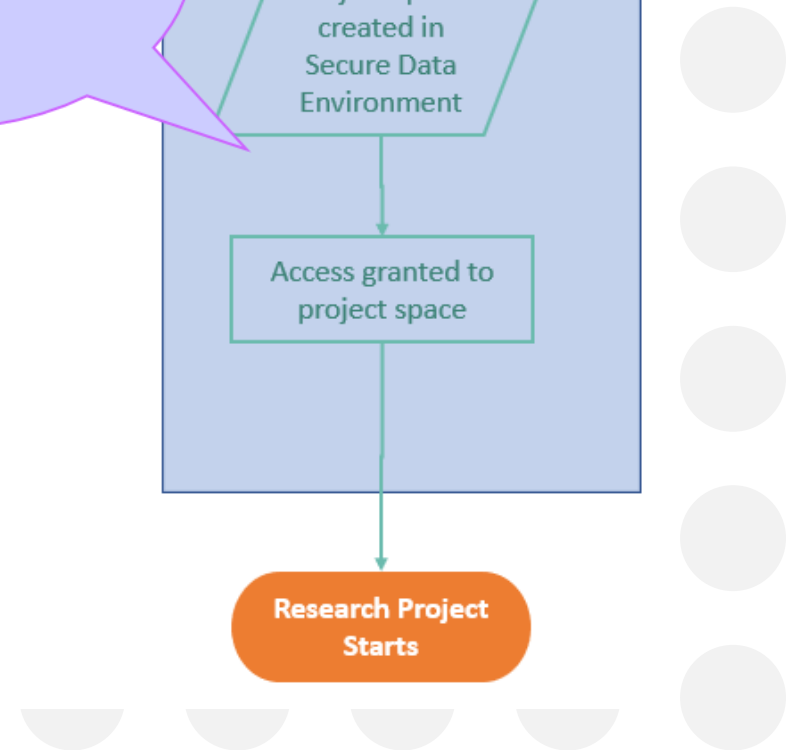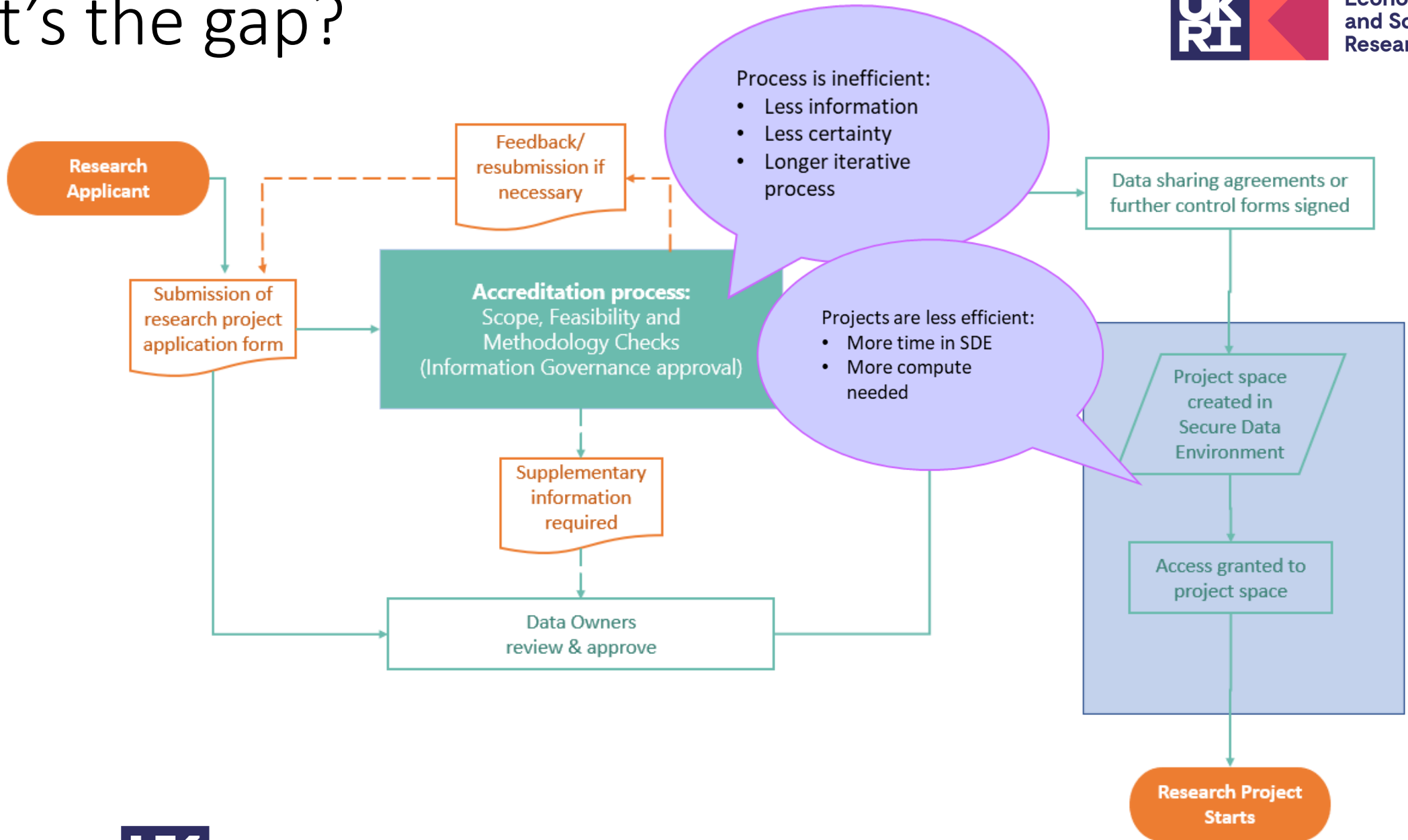- Can be used for training - methodological and specific to the dataset
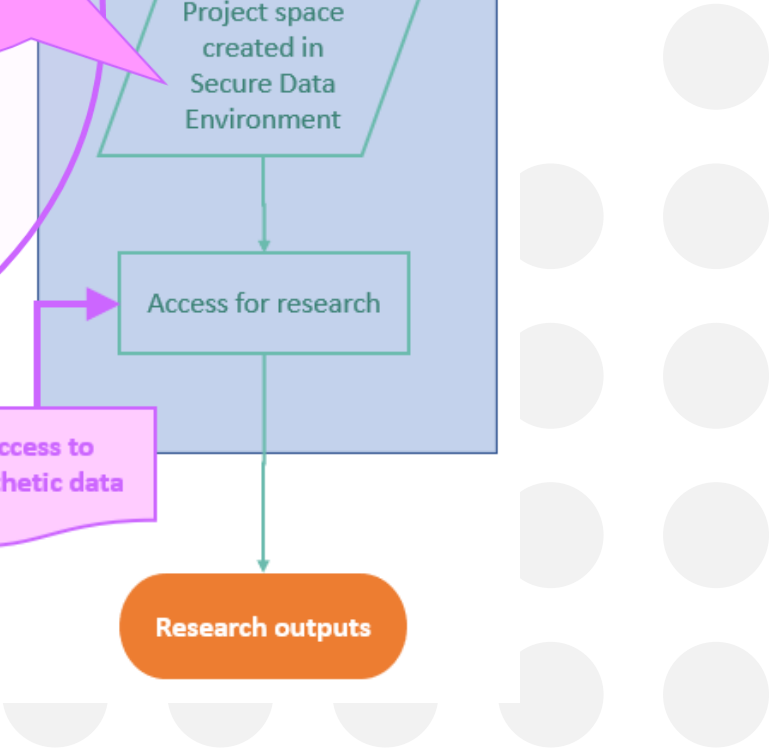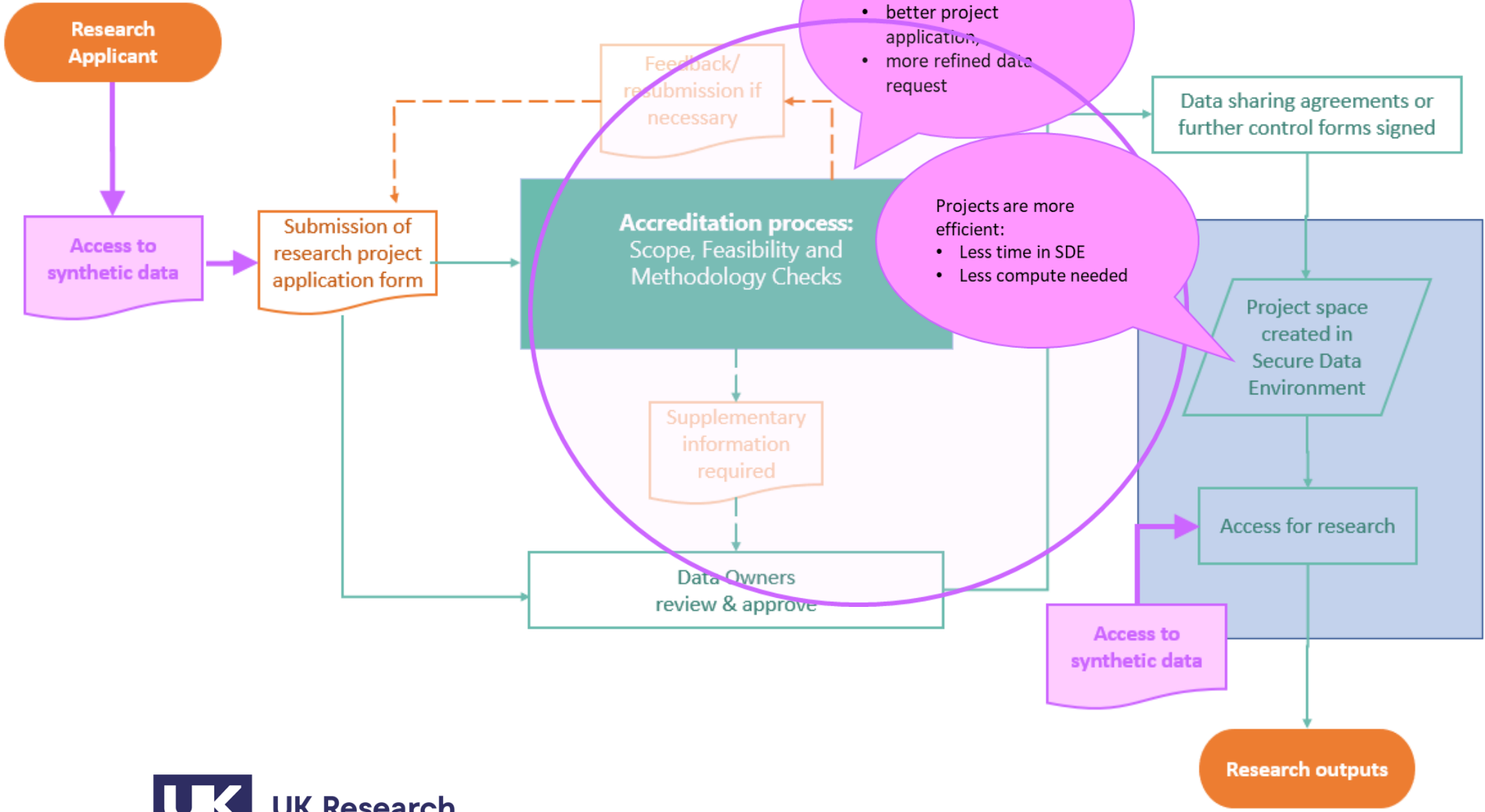
….and the problems?

- Routine creation
- Expectation
- Level of fidelity
- Public understanding and acceptance
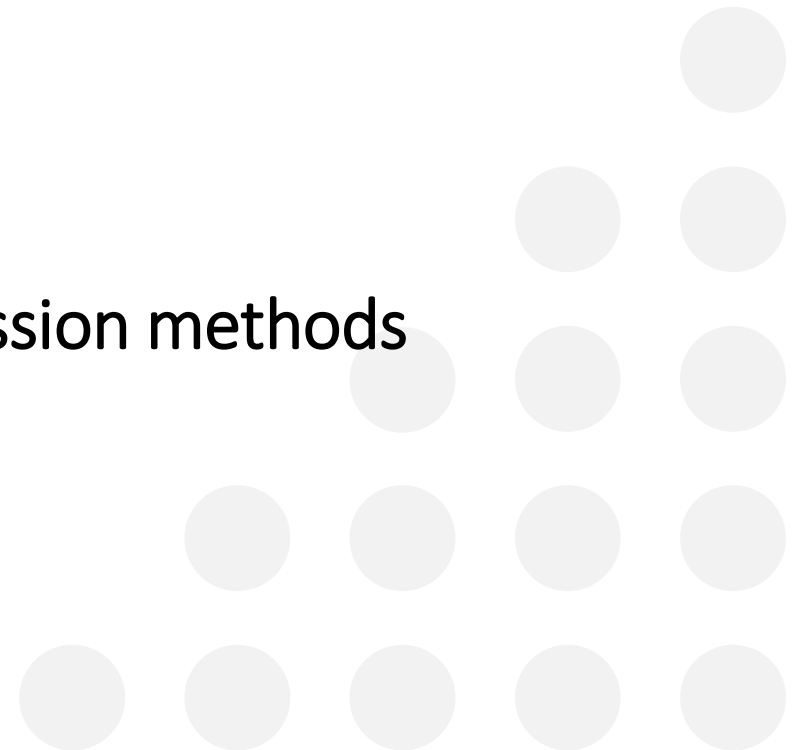- Any others?

What's the gap?

A smoother process

A framework for the future: questions and discussion

- Low-fidelity vs high-fidelity: does it matter?

- Routine production of synthetic data: what should we be aware of?

- Is synthetic data the key to privacy-preserving?

- Will synthetic datasets make traditional suppression methods redundant?

Thank you

Find us at **adruk.org** or
on X **@ADR_UK @**