

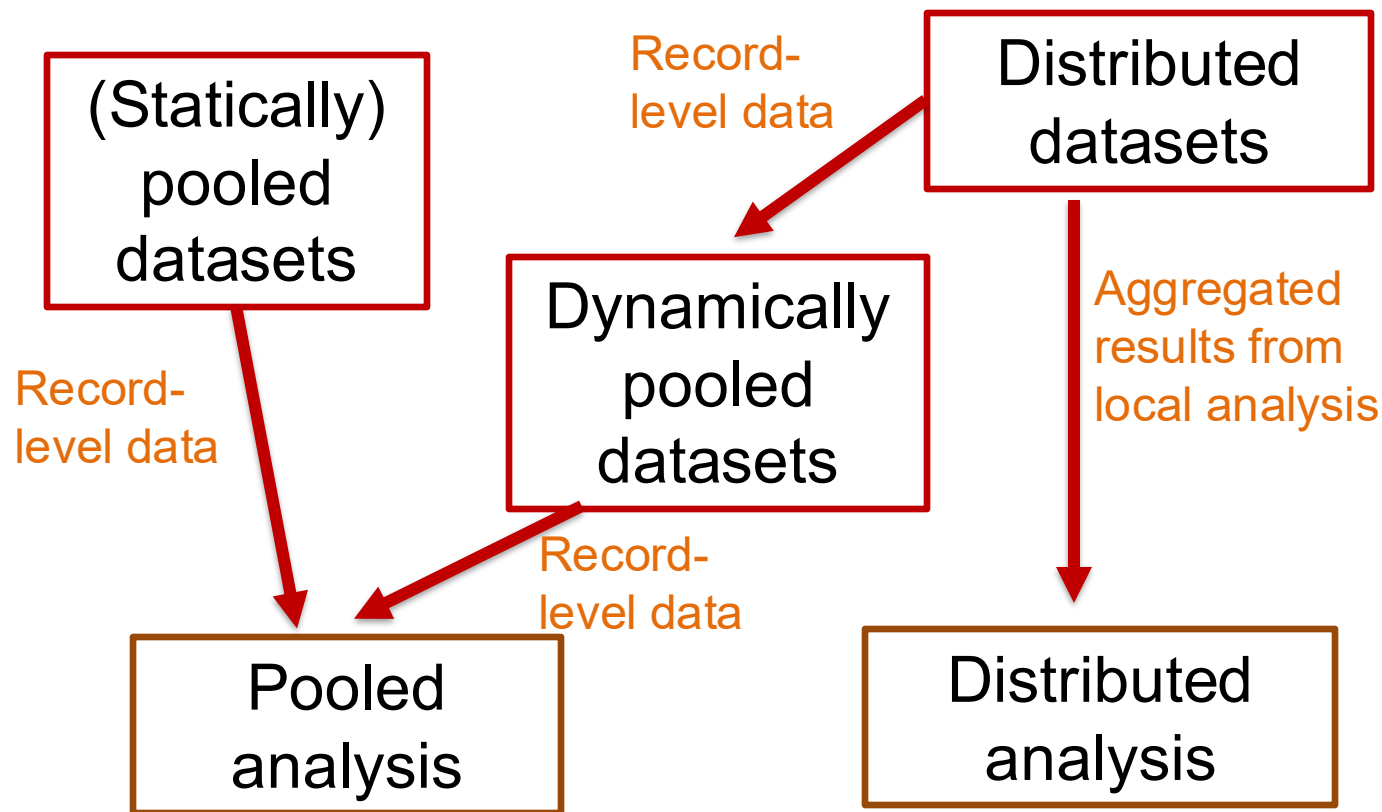
# Operations and output checking in a federated world

## Implications for systems design

Felix Ritchie, Elizabeth Green

occ\_intro  
August2024

# Definitions

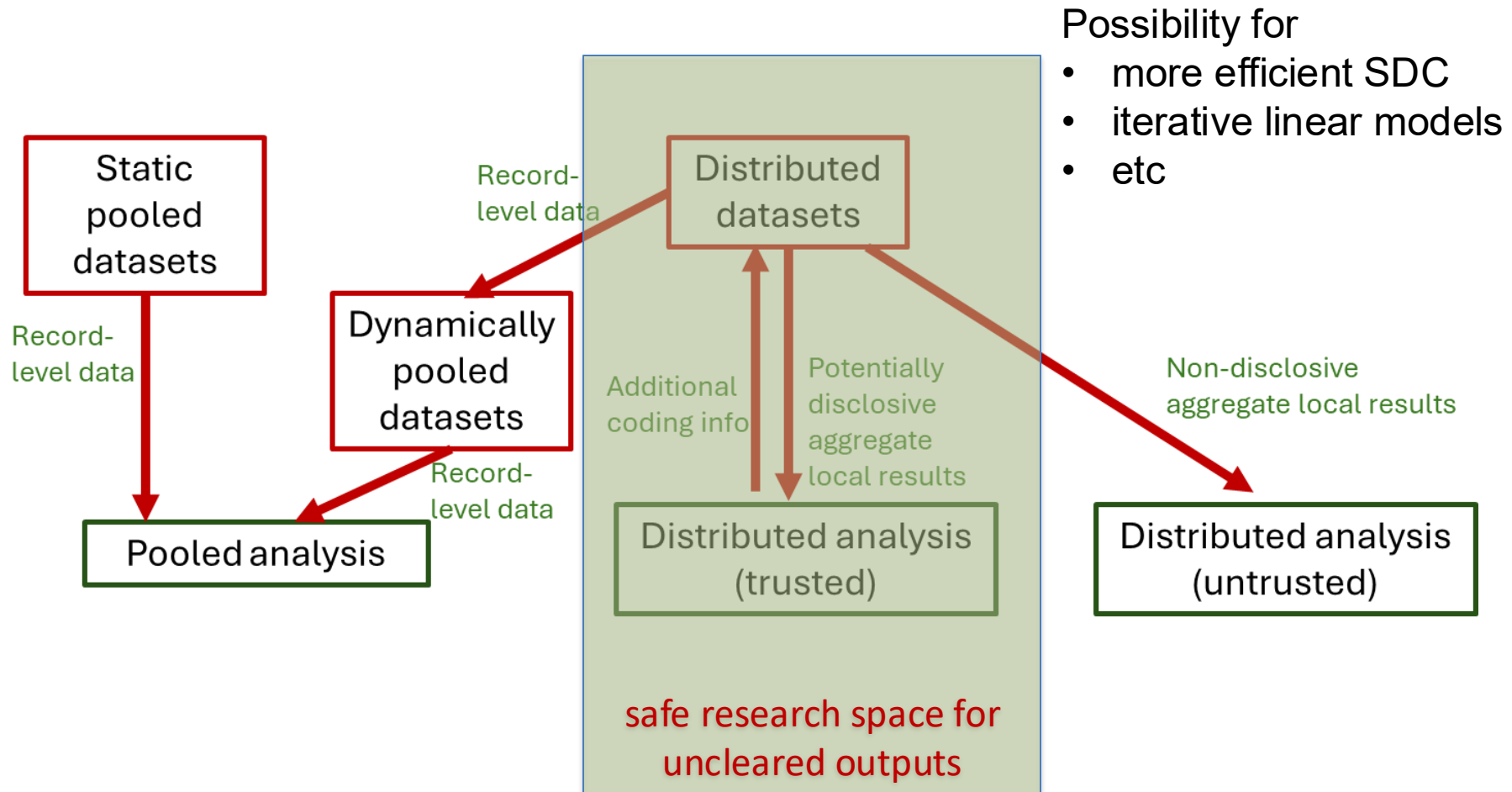


**'meta analysis'**  
**'federated analysis'** if you run the same query on all

# Why different models?

- Data management
  - Pros and cons for pooled and distributed data
    - Technical
    - Governance
- Analysis
  - Pros and cons for pooled and distributed data
    - Technical
    - Governance
  - But potentially severe analytical problems
    - Range of analysis more limited
    - SDC becomes inefficient or not possible
  - Ameliorated by 3<sup>rd</sup> option: 'trusted aggregator'

# The 'trusted aggregator' model



# Aim of today

- Understand the governance problems associated with these different models
- Focus on output checking to explore these because
  - it's a clearly defined problem
  - It's a familiar process
  - For Trusted Aggregators it's a two-stage process
  - SDC rules are common; parameters differ, and exceptions
  - Without agreement on rules or delegation, this is going to be really messy

\* We are *not* interested in which bits of SDC or analysis are feasible \*

# So: over to you

- Get allocated into 3 groups:
  - pooled analysis
  - distributed analysis with no trusted aggregator
  - distributed analysis with a trusted aggregator
- Work through the exercise questions
- Report back

# Activity 1: general principles

- **In general**, should we try to get common agreed rules or delegated authority?
  - What are the pros and cons of each?
    - Focus on governance here (practicalities later)
  - Are there alternatives?
  - What is your experience from your own governance teams?
  - What works, to get agreement on either?

Delegation or agreed rules?

	Del.	Rules
Pros		
Barriers		
Practical experience		
What works?		

# Activity 2: practicalities (ignoring exceptions)

- What does SDC require from a federated system?
  - What is necessary to allow automatic tools such as SACRO to work?
    - For an aggregator, should there be any checks on an intermediate output?
  - Do we need to agree on SDC rules?\*
  - If so, how? The most popular, the most privacy-friendly, the most research-friendly?
  - What has been your experience of
    - agreeing SDC rules?
    - operating output checking with multiple rules?

## Practicalities

What is necessary for federated checking to work?

- 
- 

What makes federated checking more effective?

- 
- 

Experience:

\* SACRO will work with multiple rules; doesn't mean it should do!



# Activity 3: exceptions

- Do we need agreement on what counts as an exception?
- Common training in handling them?
- Who reviews/makes decisions?
- For an aggregator, do they only occur at the final release stage?
  - What happens if an exception is only relevant to some of the data owners?

Again, distinguish between *necessary* and *efficient* requirements

## Exceptions

What is necessary to allow for exceptions?

- 
- 

What makes exception review more efficient?

- 
-

# Thank you for your input!

- We will collate results into evidence for a report on
    - how *best* to manage the governance for output checking
    - What needs to be done *at a minimum*, and
    - What lessons other parts of the federated analysis process can learn from this
- ⇒ Presentation for feedback at the SDC-Reboot sessions

# Acknowledgements

- This work is funded by DARE UK - Phase 2 (funded by UKRI Digital Research Infrastructure Programme; and delivered by HDR UK, and ADR UK)



