UK Research and Innovation

HDRUK Health Data Research UK

ADR UK Data-driven change

DARE UK

SACRO

SACRO: Semi-Automated Checking of Research Outputs

Professor Jim Smith,

University of the West of England

DARE UK

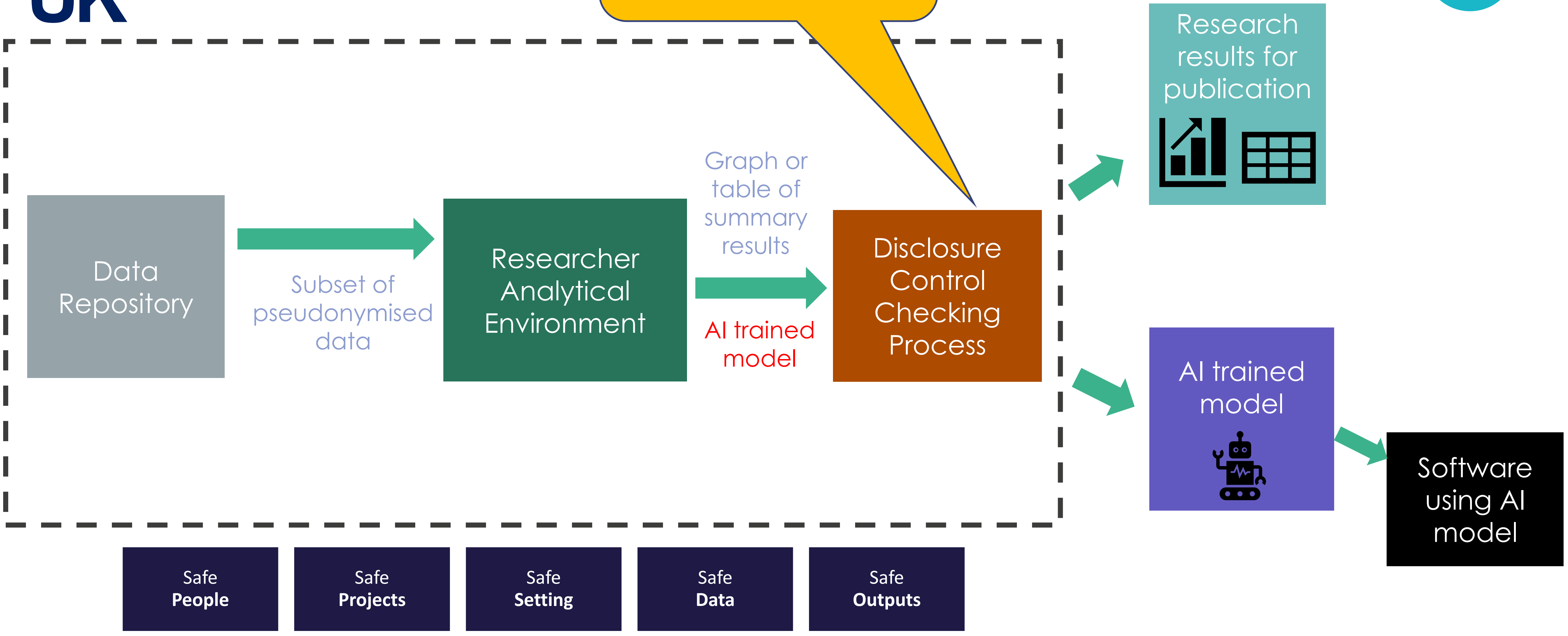
HDRUK Health Data Research UK

ADR UK

UK Research and Innovation

SACRO

(Semi) Automating this bottleneck!



Data Repository

Subset of pseudonymised data

Researcher Analytical Environment

Graph or table of summary results

AI trained model

Disclosure Control Checking Process

Research results for publication

AI trained model

Software using AI model

Safe **People**

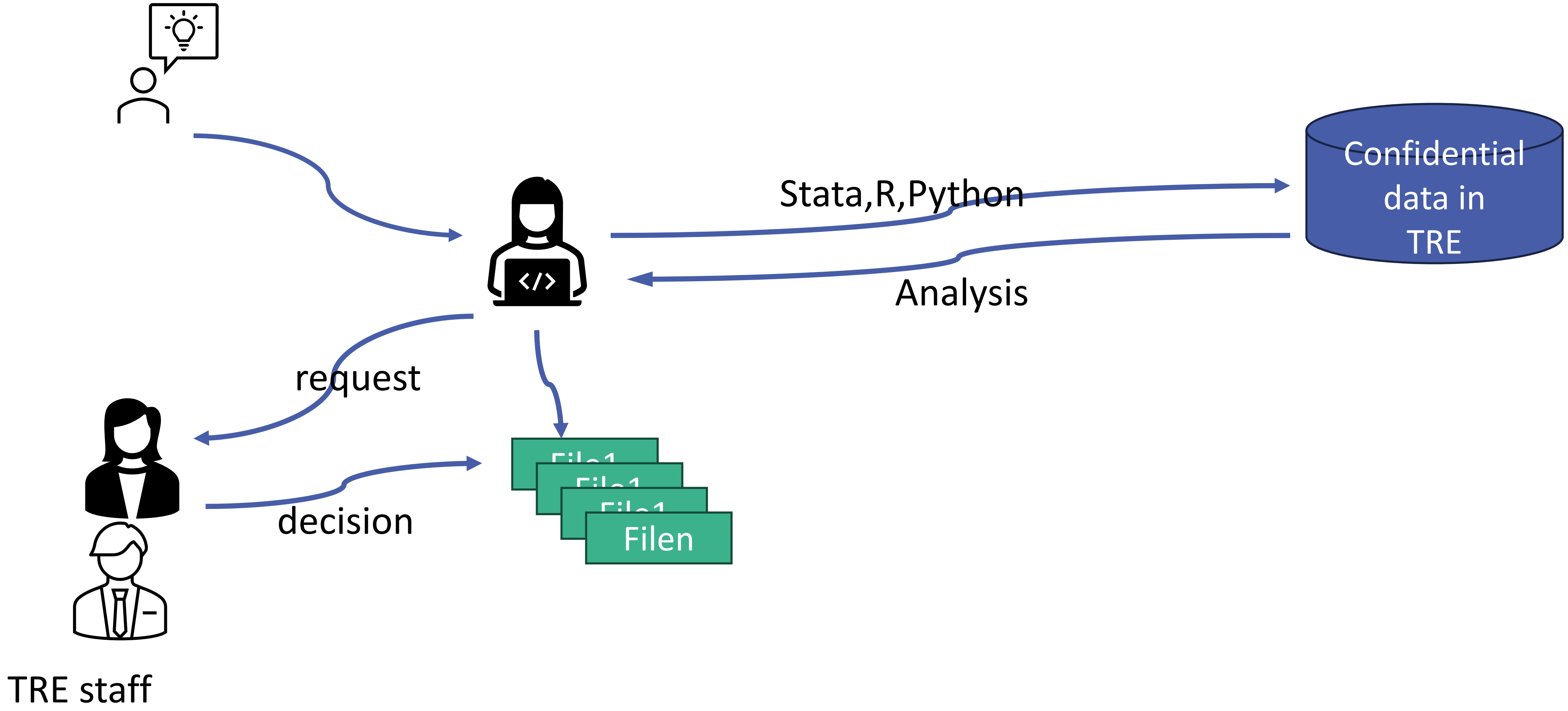
Safe **Projects**

Safe **Setting**

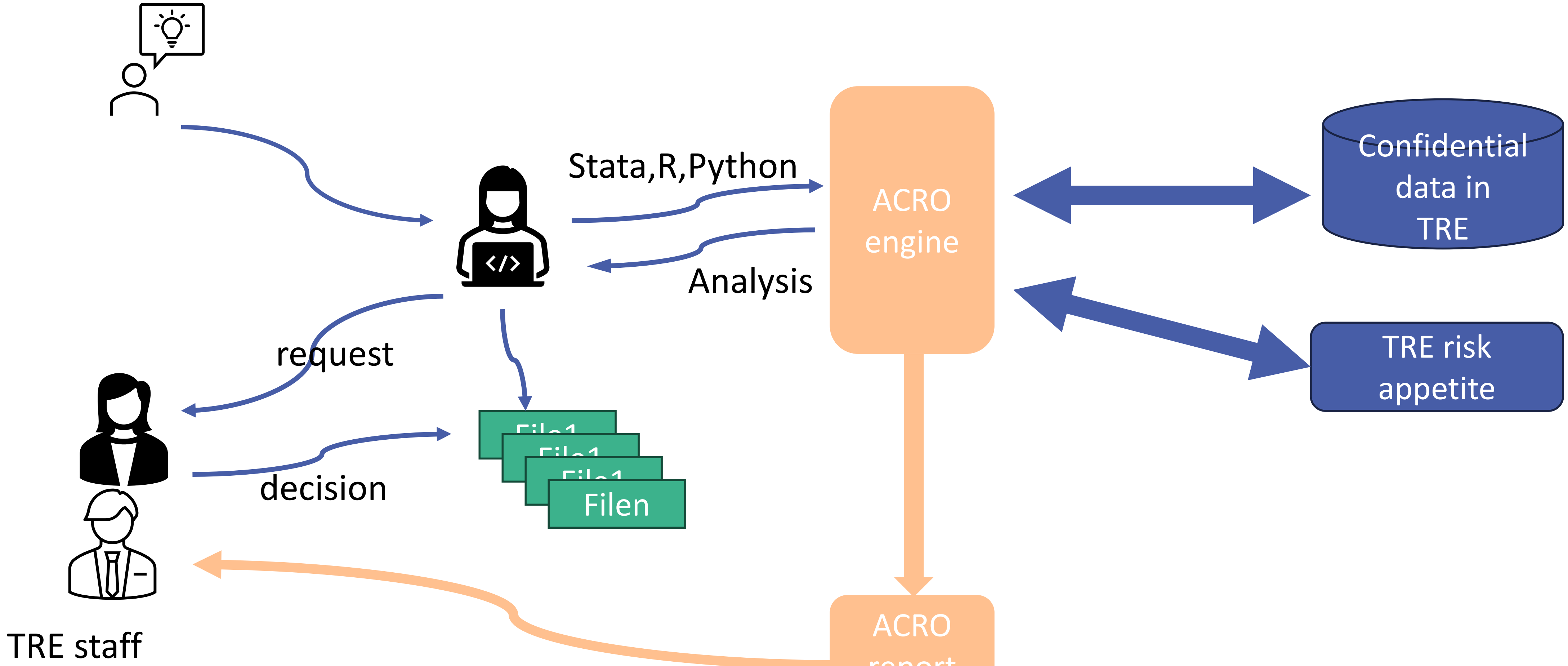
Safe **Data**

Safe **Outputs**

The current situation



SACRO in a nutshell



Similar for Machine Learning Models

Except that we :

- Run a range of “inference” attacks”
- Aim to support more ‘user journeys’
- Don’t have a set of ‘tried and trusted’ guidelines to work with

Progress to Date: technical

- First sprint ‘creation phase’ completed:
  - Initial requirements gathering
  - Technical workpackages: code and test scripts
  - ‘Conceptual Framework’: draft taxonomy
  - TRE partner feedback meetings 6th and 7th June
- Sprint 2 ‘refinement/consolidation’ underway
  - phase 2 requirements driven by TRE co-designers

Progress to Date: non-technical

- A public and a stakeholder meeting held with more planned
  - consensus statement in preparation
- International Steering Group has met twice
- Around 15 meetings with external parties
  - reverse science cafe’s
  - in person events: ESRC, UK LLC, ...
  - scoping meetings: Pictures, other Driver projects, ICO

Key (Unexpected) Findings

Finding	Adaptation
Wide range of skills and experience amongst output checkers	Add links to output to descriptions of • analysis ‘family’ • type of risk to look for • potential mitigations
IT staff at TREs quite realistic about risk from python vs e.g., R	Focus on driving governance risk assessment at a few key TREs during project- other TREs more likely to follow
Users already submitting trained ML models for output checking	Explore possibility of reverse engineer ‘SACRO’ outputs from existing code
SDC people and ML researchers aren’t so different after all	Lots of work currently underway on finding the right language to describe ML risks

‘Three stars and a wish’ ...

1. Linkages between people on SACRO and other driver projects

   but are we all fishing in the same pool of opinions?

2. PIE has benefitted from existing group at Bennett Institute

   who are familiar with the concept of a TRE already

3. Really helpful Steering Group

Really looking forward to meeting people from other projects (and beyond)

- Especially if they are willing to try out SACRO

Thank you for listening,

Questions?