



DARE UK

How is sensitive data made
accessible for research?

Webinar, Tuesday 5 July 2022, 12-1pm

DARE UK (Data and Analytics Research Environments UK) aims to design and deliver a national data research infrastructure that is **joined-up, demonstrates trustworthiness** and supports **research at scale for public good.**



Recap: What is a trusted research environment (TRE)?

A trusted research environment (TRE) is a **highly secure digital environment** that provides access to sensitive data for analysis by approved researchers.



Agenda

12:05-12:20: Public good research under the Digital Economy Act
Lily O'Flynn, UK Statistics Authority

The Secure Research Service
Bill South, Office for National Statistics

12:20-12:30: The SAIL Databank and the ZOE COVID Symptom Study
Chris Orton, Swansea University

12:30-12:40: The Administrative Data | Agricultural Research Collection project
Nick Webster, Welsh Government

12:40-12:55: Q&A and discussion

12:55-13:00: Closing remarks

Public good research under the Digital Economy Act

Lily O'Flynn, UK Statistics Authority

07 July 2022





Digital Economy Act 2017

2017 CHAPTER 30

An Act to make provision about electronic communications infrastructure and services; to provide for restricting access to online pornography; to make provision about protection of intellectual property in connection with electronic communications; to make provision about data-sharing; to make provision in connection with section 68 of the Telecommunications Act 1984; to make provision about functions of OFCOM in relation to the BBC; to provide for determination by the BBC of age-related TV licence fee concessions; to make provision about the regulation of direct marketing; to make other provision about OFCOM and its functions; to make provision about internet filters; to make provision about preventing or restricting the use of communication devices in connection with drug dealing offences; to confer power to create an offence of breaching limits on ticket sales; to make provision about the payment of charges to the Information Commissioner; to make provision about payment systems and securities settlement systems; to make provision about qualifications in information technology; and for connected purposes. [27th April 2017]

BE IT ENACTED by the Queen's most Excellent Majesty, by and with the advice and consent of the Lords Spiritual and Temporal, and Commons, in this present Parliament assembled, and by the authority of the same, as follows: –

DEA Research Powers

- 1) Allows **public authorities** to share their data for research purposes, if they wish to.
- 2) Provides a **set of principles** and rules that must be met in order for the data to be shared **safely and securely**.
- 3) Names the **UK Statistics Authority** as the body **responsible for implementing** and operationalising these rules and principles, in order for secure data sharing to go ahead.



What is the



**UK Statistics
Authority ?**

**Our role in
research
accreditation...**

**Independent
body that
operates at
arm's length
from
government**

**UKSA Strategy:
Joined up data**

**Statutory
responsibility
set out in the
legislation**

**Accreditation of
researchers,
processors and
research
projects**

Research Code of Practice and Accreditation Criteria

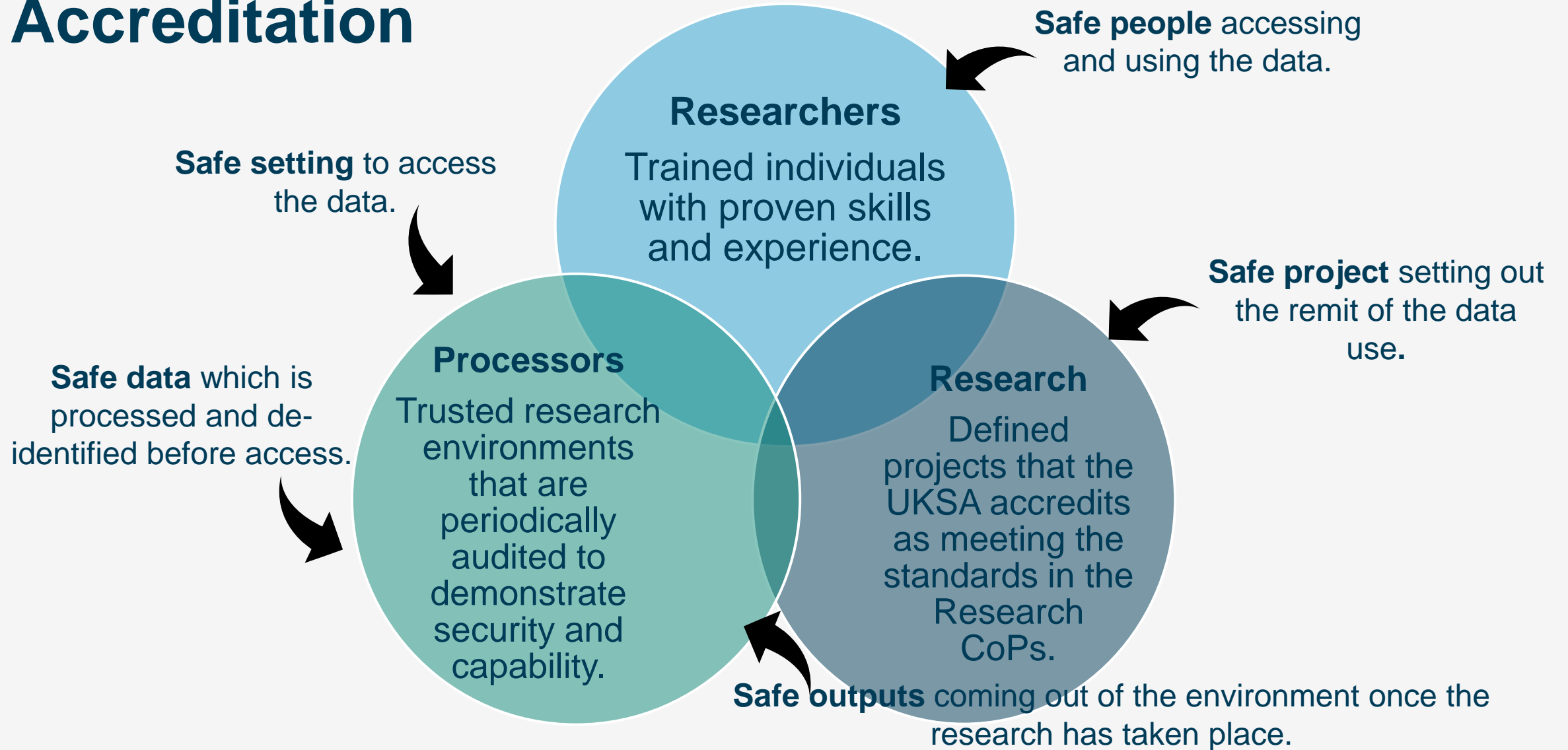
- Confidentiality
- Transparency
- Ethics and the law
- Public interest
- Proportionality
- Accreditation
- Retention and onward disclosure

33.1 The Act makes it a condition of the disclosure of data that the research for which the data is disclosed is in the public interest. For the purposes of accrediting research projects the Authority interprets public interest in the same way as 'public good', as set out in the Statistics and Registration Service Act 2007. To secure accreditation, the primary purpose of a research project must therefore be to serve the public interest in one or more of the following ways, to:

- provide an evidence base for public policy decision-making;
- provide an evidence base for public service delivery;
- provide an evidence base for decisions which are likely to significantly benefit the economy, society or quality of life of people in the UK, UK nationals or people born in the UK now living abroad;
- replicate, validate, challenge or review existing research and proposed research publications, including official statistics;
- significantly extend understanding of social or economic trends or events by improving knowledge or challenging widely accepted analyses; and/or
- improve the quality, coverage or presentation of existing research, including official or National Statistics. [\[footnote 11\]](#)

<https://www.gov.uk/government/publications/digital-economy-act-2017-part-5-codes-of-practice/research-code-of-practice-and-accreditation-criteria>

Accreditation



Processor Accreditation: Trusting Research Environments



UK Statistics Authority



[UK Statistics Authority](#) > [Digital Economy Act: Research and Statistics Powers](#)

> [Digital Economy Act Accredited Processing Environments](#)

> [List of Digital Economy Act Accredited Processing Environments](#)

List of Digital Economy Act Accredited Processing Environments

Digital Economy Act Accredited Processing Environments are accredited for the purposes of processing data, whether that processing be concerned principally with the linkage or de-identification of data, or the storage and provision of secure access to de-identified data.

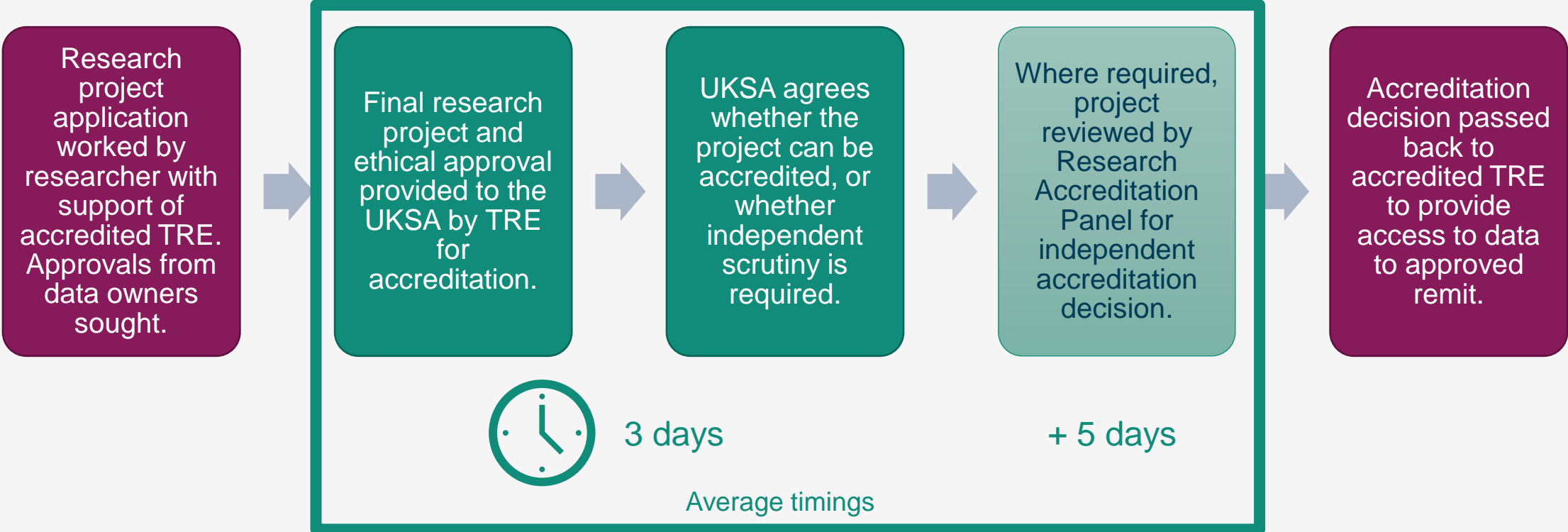
If your organisation is interested in becoming an Accredited Processing Environment under the Digital Economy Act, please email Research.Accreditation@statistics.gov.uk.

- UKSA responsible for assessing the security and capability of processing environments to award accreditation.
- Transparent framework setting out requirements for accreditation.
- Audits of accredited TREs to ensure maintained compliance as a minimum.
- Emphasis on improving maturity and service provision for researchers and analysts.



UK Statistics Authority

Project Accreditation: Approving Access to Data



Support from the UK Statistics Authority

Ensure you demonstrate how the research project serves the public interest as set out in the Research Code of Practice and Accreditation Criteria. Public interest should be articulated in a way that meets the required standard, as set out in this good practice example.

13 Public Good:
13.1 Please describe how your research project will provide a public good. Complete all the sections that apply.

Ensure you consider the wider impact of your research outputs and explain how these outputs will inform policy and decision-making processes.

Provide as much detail as possible about how your analysis will provide a public good, including references to specific policies or decisions that you anticipate the research will provide an evidence base for.

Only complete the public good

Public Good	Describe how this research project will provide this public good
Provide an evidence base for public policy decision-making	Small business owners are an increasingly important part of the workforce, and often attract preferential tax treatment and are targeted by a range of other policies. These are often justified by the desire to promote economic growth and boost productivity, but can lead to costly tax avoidance and a misallocation of people and capital. This research project will provide vital evidence on how policy affects a range of business outcomes, and the potential impact of various policy reforms.
Provide an evidence base for public service delivery	We will be able to quantify the costs and benefits of various existing policies and potential reforms. These have important implications for the government finances and the wider economy. Favourable tax treatment of small business can be costly – for example, Entrepreneurs' Relief (a reduced rate of capital gains tax applied to business assets) is estimated to cost the government in excess of £2.4 billion a year.
Provide an evidence base for decisions which are likely to significantly benefit the UK economy, society or quality of life of people in the UK	
To replicate, validate or challenge Official Statistics	
To replicate, validate or challenge existing research	
To significantly extend understanding of social or economic trends or events by improving knowledge or challenging widely accepted analyses	

Self-Assessment Form Version 3.0

Project Title
Please provide some details about your project timeline (e.g. start date, end date, any key dates for dissemination activities)

Project Timeline
Please provide a detailed description of your project's research aims/questions including project partners and sponsors

Project Purpose
Please provide some details about your project (e.g. how data are collected, used, processed and shared, as well as the research environment used)

Research Overview
Please provide detail of what data is being used, and what kind of data it is (i.e. Survey, Admin, Social Media, Web scraped etc)

Data Sources

UK Statistics Authority

When completing the purpose and overview boxes please provide a level of detail that would enable someone not familiar with the project to understand the aims of your project and how it is being undertaken

- [Research Project Application Guidance](#)
- [Ethics Self-Assessment Tool](#)
- TRE training on application standards
- TRE project accreditation tool
- Developing online video guidance

Contact and Support:

Accessing Data via the DEA: <https://uksa.statisticsauthority.gov.uk/digitaleconomyact-research-statistics/better-useofdata-for-research-information-for-researchers/>

Ethics Self-Assessment Tool: <https://uksa.statisticsauthority.gov.uk/the-authority-board/committees/national-statisticians-advisory-committees-and-panels/national-statisticians-data-ethics-advisory-committee/ethics-self-assessment-tool/>

DEA Accredited Projects: <https://uksa.statisticsauthority.gov.uk/digitaleconomyact-research-statistics/better-useofdata-for-research-information-for-researchers/list-of-accredited-researchers-and-research-projects-under-the-research-strand-of-the-digital-economy-act/>

DEA Accredited Processors: <https://uksa.statisticsauthority.gov.uk/digitaleconomyact-research-statistics/better-access-to-data-for-research-information-for-processors/list-of-digital-economy-act-accredited-processing-environments/>

Email: Research.Accreditation@Statistics.gov.uk

Secure Research Service

Bill South

Secure Research Service

ONS's Secure Research Service gives accredited researchers secure access to de-identified, unpublished data in order to work on research projects for the public good.

The SRS operates under the Five Safes framework to protect data confidentiality. The framework is a set of principles adopted by a range of secure labs, including the ONS.



Circa
5,000
accredited
researchers



Over
600
live
research projects



Approx
125
accessible
datasets

SRS Metadata Catalogue

Easy to search, by all dimensions, including theme, geographic coverage, spatial level, periodicity, data format, data typology (admin, survey etc.)

Full details of variables contained in the dataset, with descriptions, available before research applications are submitted

Viewable in browser, with downloadable extracts functionality

<https://ons.metadata.works/domain/index.html>



The screenshot shows the homepage of the ONS Secure Research Service Metadata Catalogue. At the top left is the ONS logo. To the right are navigation links: ABOUT, DATASETS, SAFE ACCESS, and PARTNERS. The main heading reads "Search the ONS Secure Research Service Metadata Catalogue". Below this is a search bar with the placeholder text "Search all datasets" and a green "SEARCH" button. Underneath the search bar are four category buttons: "BROWSE ALL", "BUSINESS", "EDUCATION", and "HEALTH". On the right side of the page, there is a decorative graphic consisting of a grid of white and grey dots of varying sizes.

The Five Safes framework



Safe people



Safe projects



Safe settings

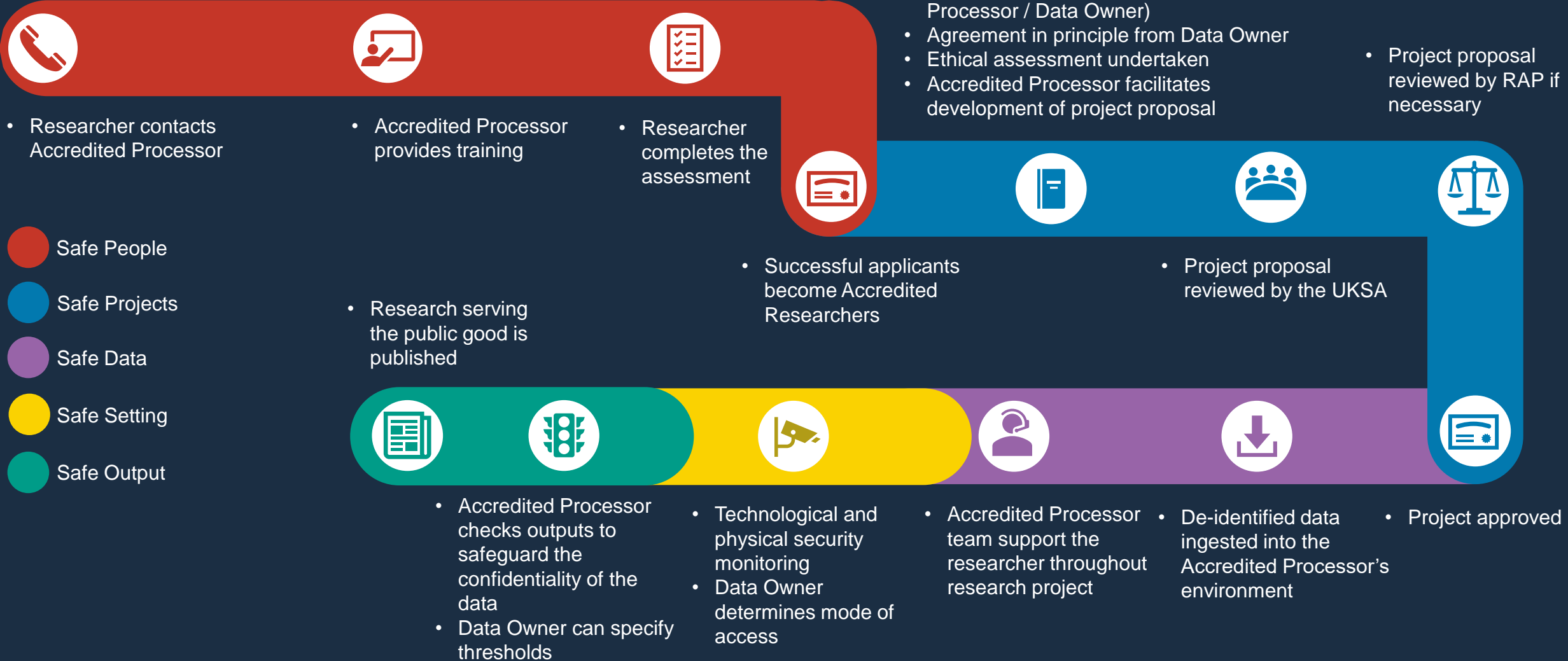


Safe data



Safe outputs

Researcher Journey



SAIL Databank ZOE Case Study

Chris Orton
Programme Manager
Population Data Science
Swansea University Medical School

SAIL Databank

- Trusted Research Environment operated by Swansea University, partnered with the NHS
- Database of anonymised data used for health, social care, and administrative (non-health) research (70+ data sources)
- Created in 2007 to pilot the technical, governance, and operating structure needed to make anonymised health data available for research within a remote access infrastructure
- As of early 2022, SAIL has supported over 500 research projects leading to over 500 publications

How it works

- SAIL hosts datasets for research from multiple data providers – primarily NHS Wales – but also local and national government, academic institutions, and charities
- Robust anonymisation and governance procedures to ensure fair use and data protection
- Application for data access for research scrutinised by an independent panel of lay people, clinicians, and scientists as to the project's suitability (SAIL Information Governance Review Panel)
- If approved – users access project-specific views of the approved datasets via a virtual desktop infrastructure accessed via two-factor authentication

NHS Wales data – an example

- Governance – data sharing agreements between NHS Wales and Swansea established to set the sharing rules, legal bases, and data protection obligations
- Technical – identifiable data always stays within the NHS. We partner with Digital Health and Care Wales to link and pseudonymise data before sharing it to SAIL, including necessary data suppression and alteration

NHS Wales data – an example

- Technical – within SAIL data is DQ/DA assessed, ensuring compliance with anonymisation policies
- Technical – patient IDs are always encrypted to ensure database specificity, and encrypted again for project specificity
- Technical – data is always accessed as a project-specific view, either restricted by period, scope, or variables, within the SAIL Gateway virtual desktop infrastructure

ZOE COVID Study – an example

- Data generated from ‘citizen science’ – during the pandemic, it was a prevalent source of publically generated information on how infection was spreading, in addition to NHS and government data also being generated
- Permitted the tracking of COVID-19 at local and national level before routine testing, track and trace etc. were all fully established
- >4 million participants to date

ZOE COVID Study – an example

- SAIL as the BREATHE TRE, partnered with Zoe to make available their anonymised data to the NHS, local authorities, national government agencies, and academic researchers to for ongoing COVID-19 research and tracking
- As with the NHS Wales example, we followed the same path to set up data ingest, with some slight differences to aid the pandemic response

ZOE COVID Study – an example

- Governance – agree the DSA – pre-anonymised data (ZOE ensure this before sharing)
- Technical – create automated daily transfer mechanism for millions of rows of data to transfer (including all additions, withdrawals, and schema changes)
- Technical/governance – review the data arriving in to SAIL to ensure it conforms to anonymisation policies

ZOE COVID Study – an example

- Governance – team setup to ensure review of all applications for the ZOE data even before submission to the IGRP (scoping). Emergency response projects identified as any NHS body, local or national authority, etc. needing data for COVID-19 monitoring, strategy and policy purposes – 2/3 day turnaround
- Governance/access – users sign data access agreements, send CVs and ensure compliance with safe research
- Access to data via the SAIL Gateway (VDI), using mobile authenticator as rapid 2FA solution to access for this dataset alone

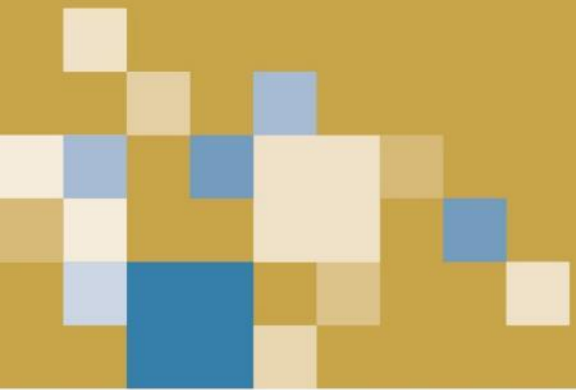
ZOE COVID Study – an example

- Technical – data provisioned automated to update every few days to ensure it
 - a) did not compromise ongoing work
 - b) ensure data was up to date in content and participant presence
- Governance – all outputs subject to SAIL's DEA compliant output checking process, regardless of origin (public sector/academia etc.)



AD | ARC

Administrative Data
Agricultural Research Collection



So what is AD|ARC?

- The AD|ARC (Administrative Data | Agricultural Research Collection) project aims to integrate the **human dimension** with data on farming activities.
- AD|ARC brings together experienced data linkage researchers and experts in agricultural affairs with partner organisations.
- Research will investigate relationships between farm households characteristics and recent farming and land use activities with a view to improving the design of future policies.

The data



NI
EU FSS
Rural Payments
Census
IDBR
Passed Feasibility test

Wales
EU FSS
Rural Payments
Census
Health data
IDBR
Education data
Construction of RRD
underway



Scotland
EU FSS
Rural Payments
Census
Health data
IDBR
Education data
PBPP forms being
finalised

England
EU FSS
Rural Payments
Census
Health data
IDBR
Education data
Waiting for provision
of data in SRS

Key
Linkage complete
Linkage underway
Governance in progress
Issues to be resolved

What are we going to do with the data?

Conduct research that will be relevant to farming households, businesses and policy makers.

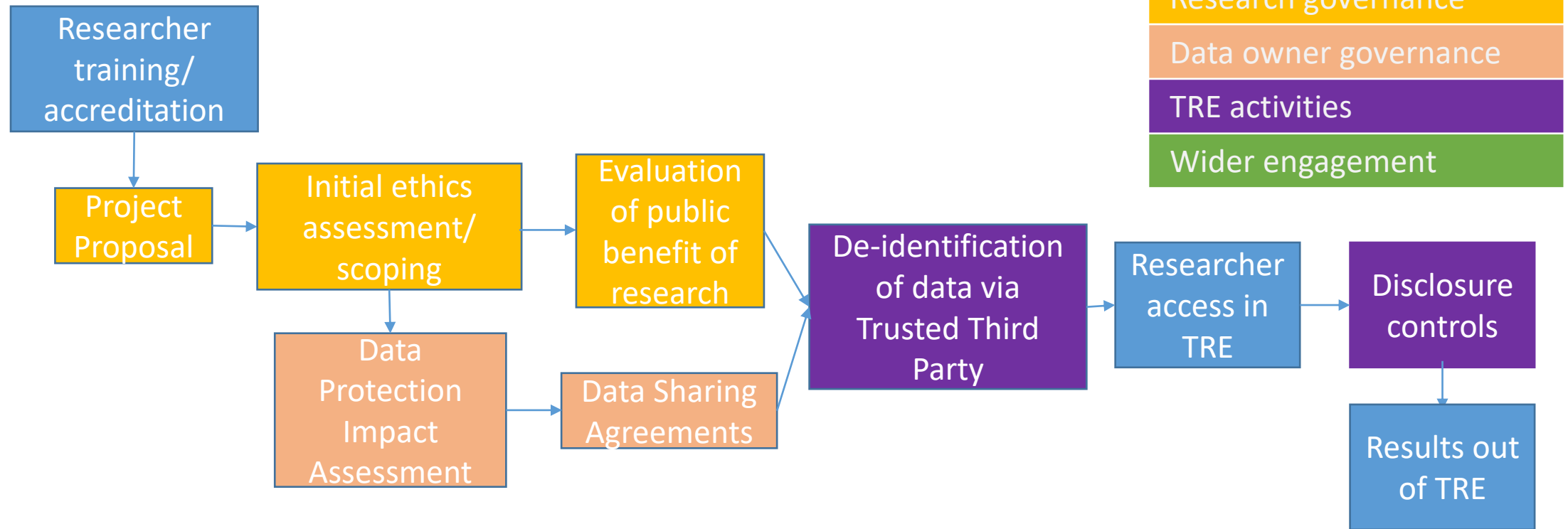
Research work streams:

- Household socio-economic characteristics
- Health and wellbeing
- Prosperity and business resilience
- Environment and place

The different UK Nations

Country	Trusted Research Environment	Trusted Third Party	Application process
England	Secure Research Service	Office for National Statistics	Research Accreditation Panel
NI	Secure Research Service	Northern Ireland Statistics and Research Agency	Research Accreditation Panel
Scotland	National Safe Havens	National Records of Scotland	Public Benefits and Privacy Panel
Wales	Secure Anonymised Information Linkage	Digital Health and Care Wales	Information Governance Review Panel

What is required of Researchers



Consultation with stakeholders, members of the public, government analysts, policy makers and academic experts

Researcher activities

- Researcher accreditation
 - Most TREs now accept central ONS Digital Economy Act Accreditation
 - Takes a couple of weeks to come through
- Researcher access in TREs
 - Metadata patchy
 - SAIL now accessible remotely
 - Some TREs accessible via safe rooms only
- Results out
 - Straightforward

Time taken for AD|ARC: Weeks

Research Governance

- Project proposal
 - Useful exercise in feasibility but varies significantly around the UK
- Ethics assessment
 - Helped by ONS standardisation
 - Varies by institution
- Evaluation of public benefit of research
 - Variable definition of “public good” across the different panels
 - Timescales getting shorter but still Covid backlog
 - If anything challenged, can take a long time

Time taken for AD|ARC: months

Data owner Governance

- Data protection Impact Assessments
 - Different formats between all data owners
 - Have to go through many internal teams with risk averse attitudes
 - Different interpretations of same legislation
- Data sharing agreements
 - Often slowed by legal technicalities
 - Complications of who is responsible for what
 - Often need high level sign off from very busy people

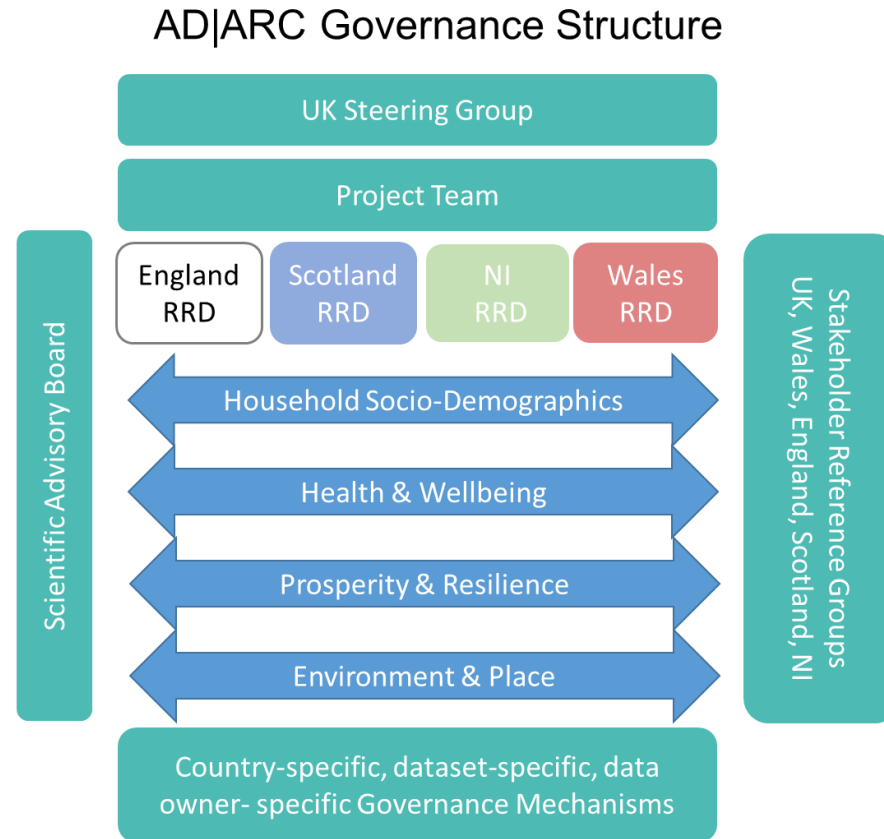
Time taken for AD|ARC: TBC (still not completed but years!)

TRE Activities

- De-identification of data
 - Very resource dependent
 - Inconsistent matching techniques across TREs
 - Can be a “black box”
- Disclosure controls
 - Reasonably quick
 - Inconsistent rules
 - Manual process

Time taken by AD|ARC: Months

Wider engagement



Time taken by AD|ARC: N/A

How DARE can help

- Building common understanding of legislation/policies
- Increase transparency of administrative data research
- Increase consistency across TREs
- Develop technical solutions to work between TREs
- Support metadata development



UK Research
and Innovation



DARE UK

Thank you for listening

www.dareuk.org.uk

[@DARE_UK1](https://twitter.com/DARE_UK1)