UK Data Service

Digital trace data management in the UK - legal, technical and practical considerations

Steve McEachern and John Sanderson

UK Data Service

14 January 2026

UKRI Economic and Social Research Council

UKDS Vision

- Our vision is for economic, population and social research to drive an innovative and thriving research and policy-making ecosystem, leading to enhanced knowledge, better decisions, and improved outcomes in society

- Our mission is to catalyse impactful social science research by providing access to high-quality, curated and trustworthy datasets of national strategic importance; fostering data literacy; and contributing to the advancement of knowledge through dynamic data infrastructure and strategic partnerships

Who are UKDS?

- Our services:
  - Data operations: Collections, Curation and Access
  - Macrodata and Census
  - Technical data services
  - Training and user support
  - Communications and User Engagement
  - Impact

Who are we?

- Five partners
- 100+ staff
- In operation as UKDS since 2012
  - 55+ years of combined history (UKDA: 1967)
  - National collaboration since 2003
  - Converging on the 2024-2030 award - £37m

UKDS partners

- University of Essex: Lead partner
  - Lead partner
  - Data operations:
    - Data deposits
    - Curation
    - Data access (including SecureLab)
  - Communications
  - User Support and Training
- JISC:
  - Engagement and Impact
  - Impact stories: UKDS Data Impact blog, co-produced with users
  - Macrodata
  - Census

- Manchester: User Support and Training (UST)
  - User support
  - Training
  - (Training communications)

- Census:
  - Lead: University College London
  - Aggregate data: JISC
  - Spatial data: University of Edinburgh

2024-25 activity

54,427
total registered users

900+
new datasets

102+
staff

85,000+
total downloads, including:
- open+safeguarded
- international microdata
- census data
- census boundaries

113
training events

4,444
attendees
at training

229,527
visits to the online
learning hub

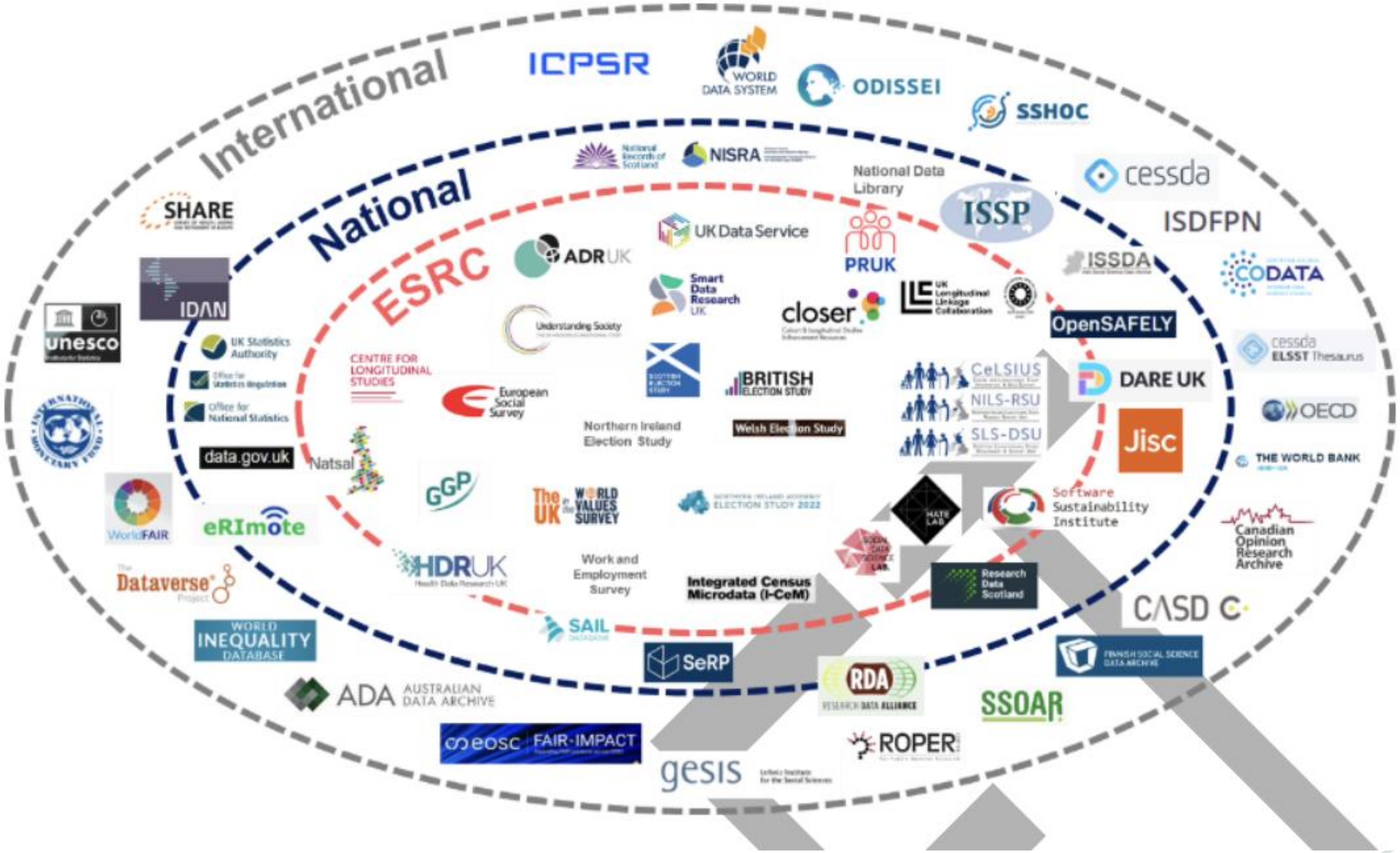
~ 358k
sessions/visits
on UKDS website

3,000
helpdesk requests handled,
ensuring smooth user support

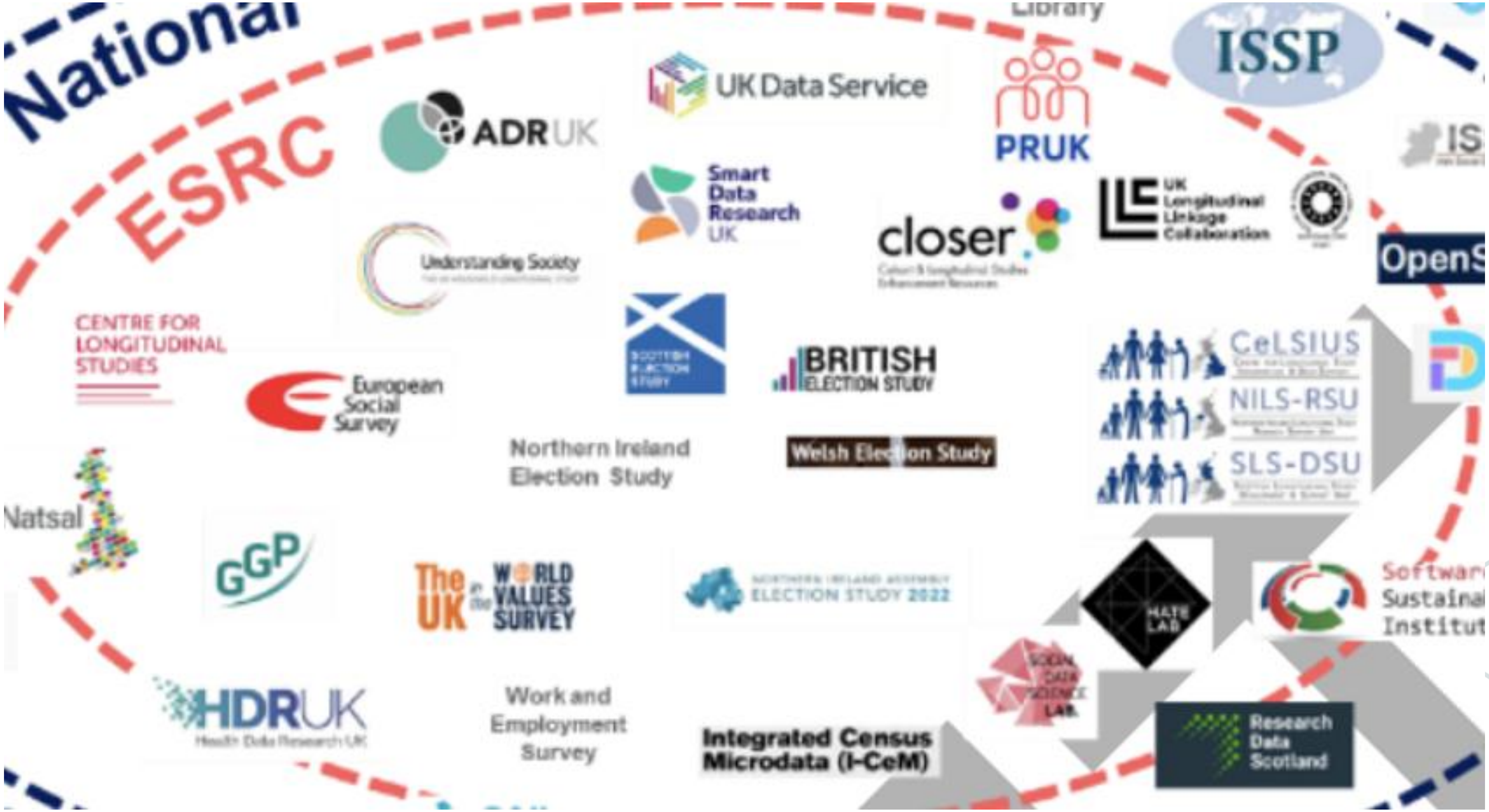
23,357
page views of our
Data Impact blog

(tbc)
people engaged with our news
stories and articles

The social science infrastructure landscape
(Source: Richard Welpton, ESRC)


International
National
ESRC
ICPSR
WORLD DATA SYSTEM
ODISSEI
SSHOC
cessda
ISDFPN
CODATA
cessda ELSST Thesaurus
OECD
THE WORLD BANK
Canadian Opinion Research Archive
CASD
FINNISH SOCIAL SCIENCE DATA ARCHIVE
SSOAR
ROPER
RDA RESEARCH DATA ALLIANCE
gesis
eosc FAIR-IMPACT
ADA AUSTRALIAN DATA ARCHIVE
WORLD INEQUALITY DATABASE
The Dataverse Project
WorldFAIR
unesco
SHARE
IDAN
National Records of Scotland
NISRA
National Data Library
ISSP
ISSDA
OpenSAFELY
DARE UK
Jisc
Software Sustainability Institute
Research Data Scotland
SeRP
UK Statistics Authority
Office for Statistics Regulation
Office for National Statistics
data.gov.uk
eRImote
UK Data Service
ADR UK
Smart Data Research UK
Understanding Society
PRUK
closer
UK Longitudinal Linkage Collaboration
CENTRE FOR LONGITUDINAL STUDIES
European Social Survey
BRITISH ELECTION STUDY
Northern Ireland Election Study
Welsh Election Study
CeLSIUS
NILS-RSU
SLS-DSU
Natsal
GGP
The UK in the WORLD VALUES SURVEY
NORTHERN IRELAND ASSEMBLY ELECTION STUDY 2022
HATE LAB
SOCIAL DATA SCIENCE LAB
HDRUK Health Data Research UK
Work and Employment Survey
Integrated Census Microdata (I-CeM)
SAIL DATABANK

ESRC funded infrastructure



UK Data Service

What about digital trace data infrastructure?

Incentives to get, use and share data

- Do data owners REALLY want to share their data?
- Do researchers REALLY want to share their data?

- **Control over access enables control over intellectual property**

- For social media companies
  - Continued revenue streams
  - Control over criticism
- For researchers
  - Competitive advantage in the research market
  - Continued citation streams

- **Need a model where the benefits of sharing outweigh the costs**

Legal and ethical considerations

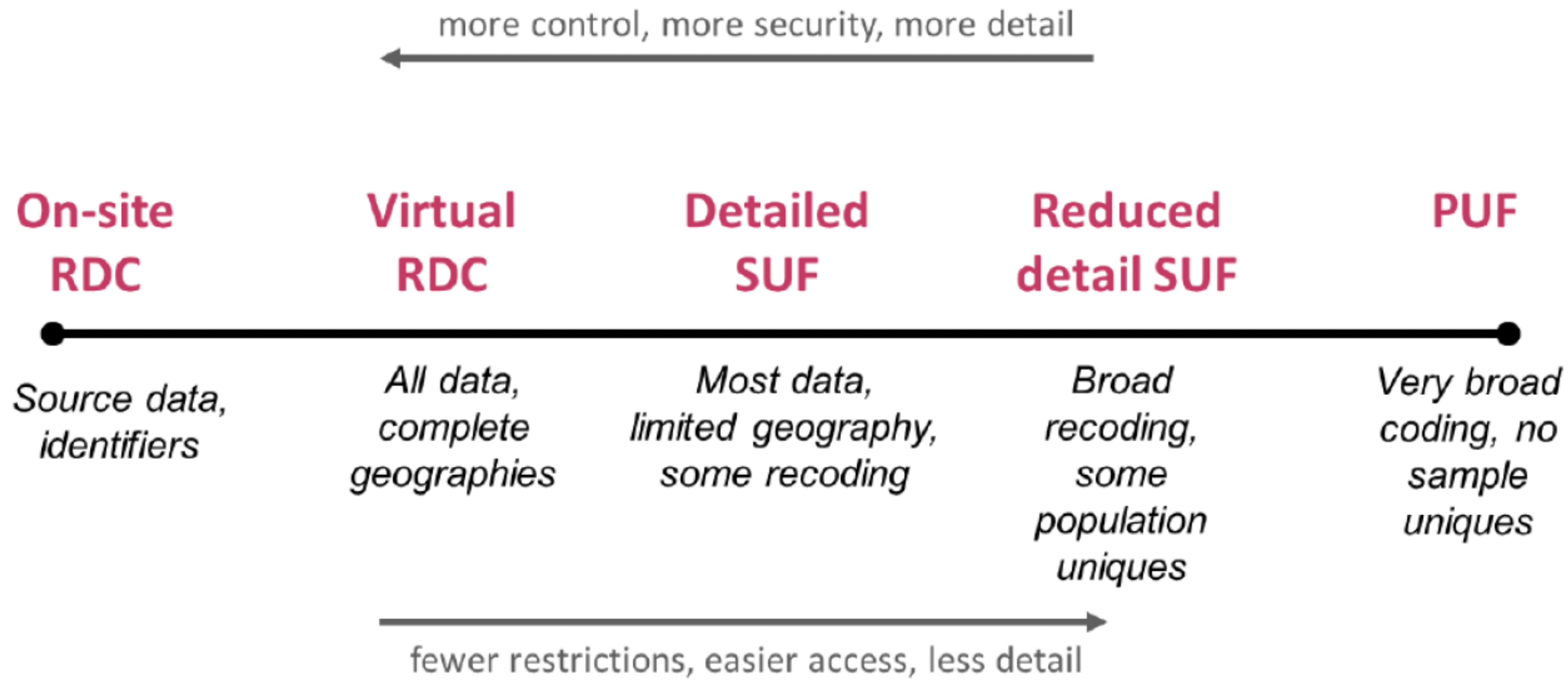
Intellectual property

- Ownership and control:
  - Who owns the content?
  - Who determines use of the content?
- What do the Terms of Use say about:
  - **Getting**: How do I go about getting the data
  - **Using**: Once I have the data, what can I do with it?
  - **On-sharing**: if I have the data, can I share with others?

Privacy and data protection

- Consent
- (Re)identification
- Anonymisation processes

Data access models



more control, more security, more detail

On-site RDC

Virtual RDC

Detailed SUF

Reduced detail SUF

PUF

Source data, identifiers

All data, complete geographies

Most data, limited geography, some recoding

Broad recoding, some population uniques

Very broad coding, no sample uniques

fewer restrictions, easier access, less detail

Figure 1 Data access spectrum from Green and Ritchie, 2016[1]

- Ritchie, F. & Kendal, C. (2024) FDS BN06 The data access spectrum. University of the West of England. https://uwe-repository.worktribe.com/output/13507443

The Five Safes

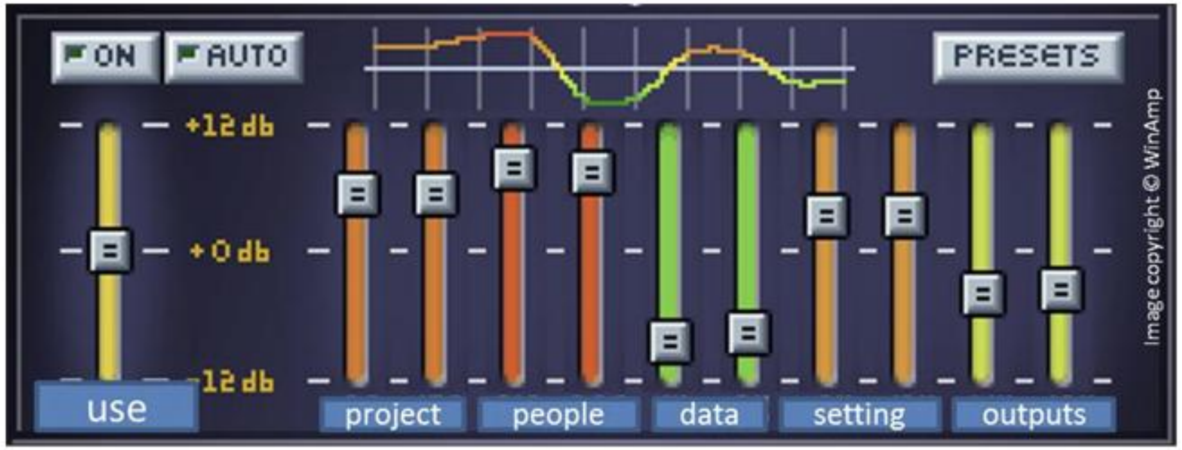
The Five Safes (Desai et al, 2015)

- People
- Projects
- Data
- Settings
- Outputs

Plus… (McEachern et al., 2021)

- Groups
- Organisations

- Instantiated in the UK for government data through the Digital Economy Act
- Accredited by the UK Statistical Authority



Desai T., Ritchie F. and Welpton R. (2015) “The Five Safes: designing data access for research” https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf
McEachern, S. et al.. (2021). CADRE Five Safes Framework - Conceptualisation and Operationalisation of the Five Safes Framework (1.0). Australian Data Archive, Australian National University. https://doi.org/10.5281/zenodo.5748611

Usage by access mode

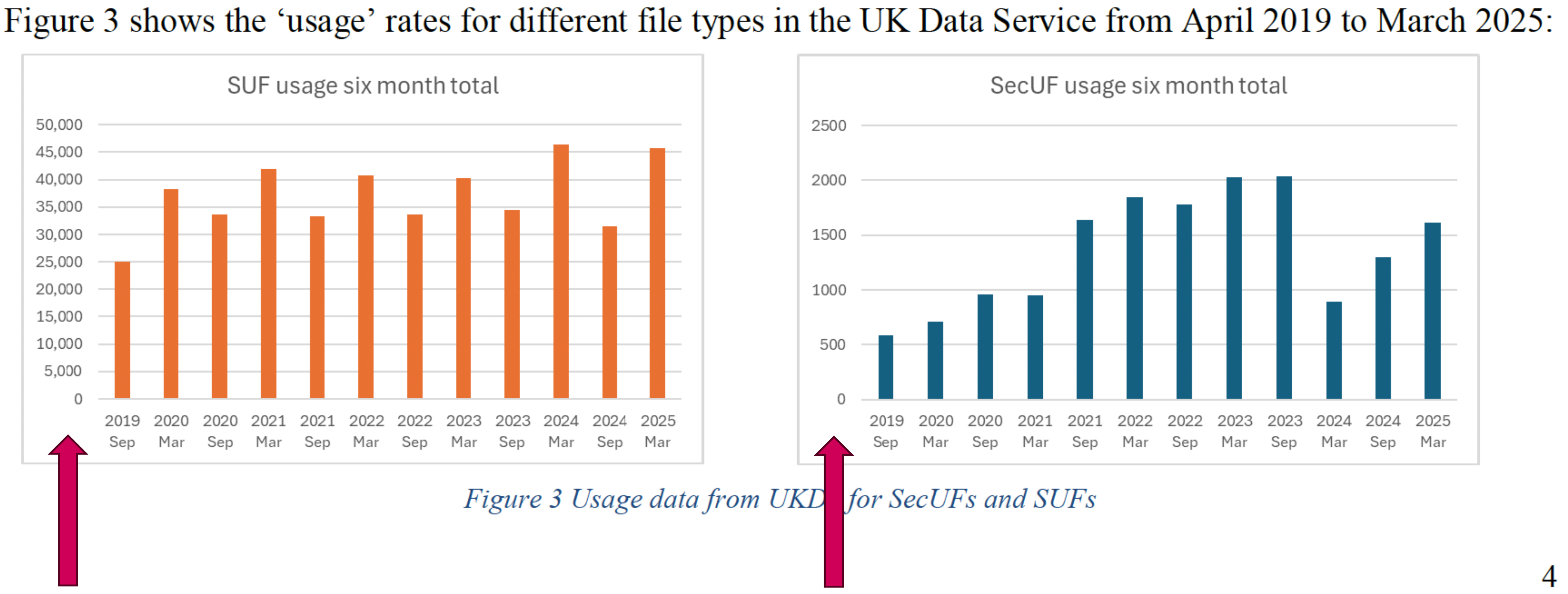
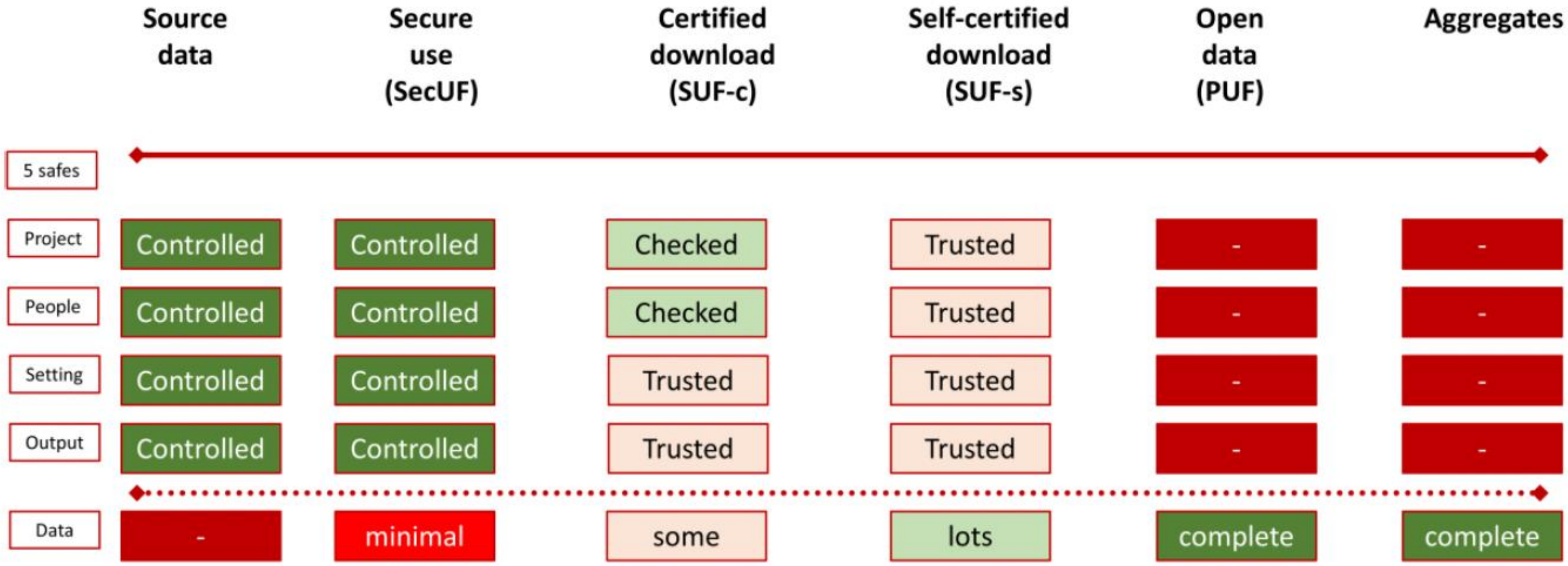


Figure 3 shows the ‘usage’ rates for different file types in the UK Data Service from April 2019 to March 2025:

Figure 3 Usage data from UKD for SecUFs and SUFs

Magder, Ritchie, Sanchez and Welpton (2025) In defence of Scientific Use Files. UNECE Conference of European Statisticians. Expert Meeting on Statistical Data Confidentiality. Barcelona, Spain, 15-17 October 2025. https://unece.org/sites/default/files/2025-10/SDC2025_Ritchie_D_0.pdf

The five safes and the data access spectrum



	Source data	Secure use (SecUF)	Certified download (SUF-c)	Self-certified download (SUF-s)	Open data (PUF)	Aggregates
5 safes						
Project	Controlled	Controlled	Checked	Trusted	-	-
People	Controlled	Controlled	Checked	Trusted	-	-
Setting	Controlled	Controlled	Trusted	Trusted	-	-
Output	Controlled	Controlled	Trusted	Trusted	-	-
Data	-	minimal	some	lots	complete	complete

Figure 6 Five Safes controls applied to the data access spectrum

- Ritchie, F. & Kendal, C. (2024) FDS BN06 The data access spectrum. University of the West of England. https://uwe-repository.worktribe.com/output/13507443

Technical considerations

Data requirements

- Data formats
- Data management
- Data immutability
- Data processing

- Is the data more like:
  - Social science?
  - Climate science?

Infrastructure requirements

- Compute (processing power)
- Storage

- AI/ML:
  - Am I using AI models
  - Can I use the data in an AI model?
  - If I use the data in an AI model – is it used for future training of the model?
  - Is this compliant with the Terms of Use of the DATA and of the MODEL?

Transfer of content

UKDS EULA

- *Opt-in model*
- 5.4 To abstain from using any Online Data Tools in connection with your use of the Data Collection(s), **unless explicit written permission is granted** by the Data Service Provider. The Registered User acknowledges that the **use of such tools may compromise data security and enable data transfer to unauthorised parties** in violation of clause 4.
- Source: https://ukdataservice.ac.uk/app/uploads/cd137-enduserlicence.pdf (emphasis added)

ChatGPT (OpenAI) terms of use

- *Opt-out model*
- You can opt out of training through our privacy portal(opens in a new window) by clicking on "do not train on my content." To turn off training for your ChatGPT and Operator conversations, follow the instructions in our Data Controls FAQ(opens in a new window). **Once you opt out, new conversations will not be used to train our models**.
- Source: https://openai.com/policies/how-your-data-is-used-to-improve-model-performance/ (emphasis added)

Practical considerations

Data and technical

- Can/should we really store this much data?
- Do specified extracts have reuse benefits beyond their original purpose?
- How will access environments be scaled?
  - Number of virtual machines
  - Volume of storage
  - Separation of storage by project

Governance

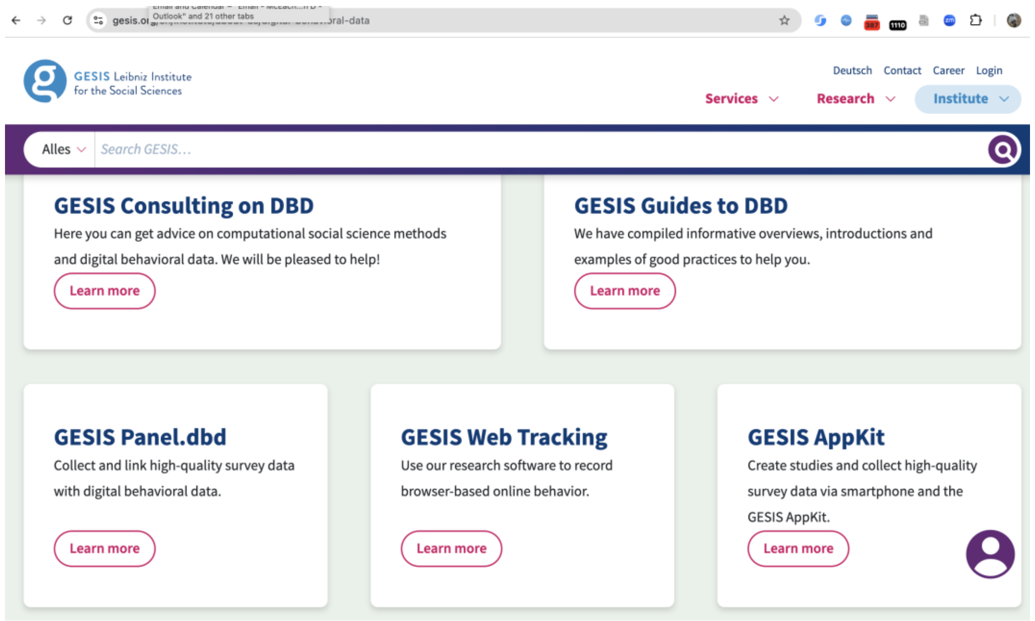
- How will data access requests be handled?
- How will access requests be scaled?
- What happens to risk and privacy when we integrate more and more data sources?
- Open research vs data protection

Examples of DTD infrastructures and services

- GESIS (Germany)
- Digital Behavioural Data
- https://www.gesis.org/en/institute/about-us/digital-behavioral-data

ICPSR (USA)

- SOMAR: Social Media Archive
- https://www.icpsr.umich.edu/sites/somar/home



GESIS Consulting on DBD

Here you can get advice on computational social science methods and digital behavioral data. We will be pleased to help!

Learn more

GESIS Guides to DBD

We have compiled informative overviews, introductions and examples of good practices to help you.

Learn more

GESIS Panel.dbd

Collect and link high-quality survey data with digital behavioral data.

Learn more

GESIS Web Tracking

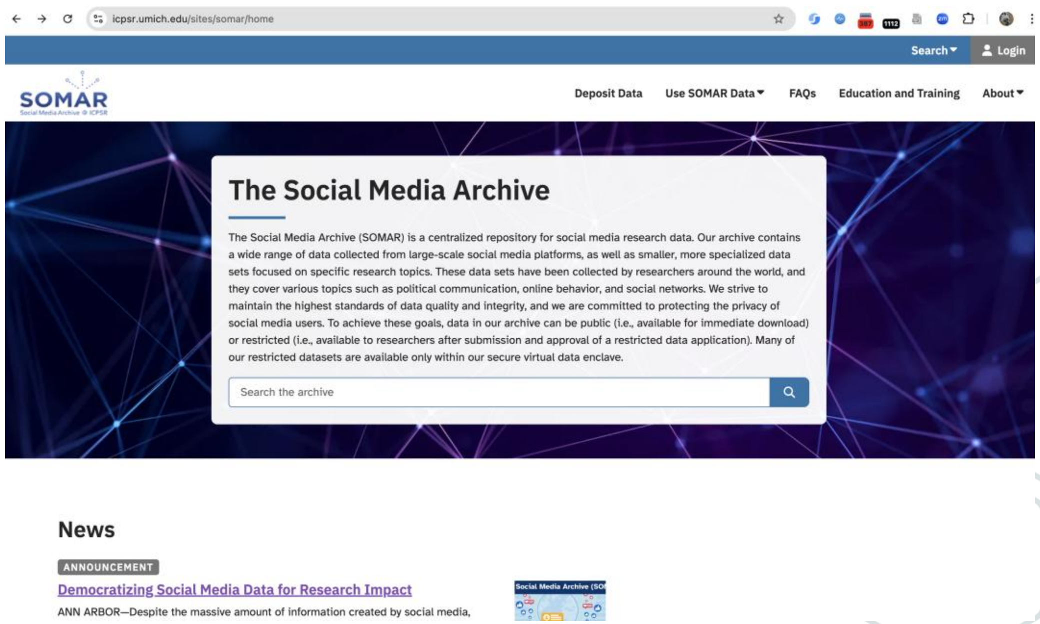
Use our research software to record browser-based online behavior.

Learn more

GESIS AppKit

Create studies and collect high-quality survey data via smartphone and the GESIS AppKit.

Learn more



The Social Media Archive

The Social Media Archive (SOMAR) is a centralized repository for social media research data. Our archive contains a wide range of data collected from large-scale social media platforms, as well as smaller, more specialized data sets focused on specific research topics. These data sets have been collected by researchers around the world, and they cover various topics such as political communication, online behavior, and social networks. We strive to maintain the highest standards of data quality and integrity, and we are committed to protecting the privacy of social media users. To achieve these goals, data in our archive can be public (i.e., available for immediate download) or restricted (i.e., available to researchers after submission and approval of a restricted data application). Many of our restricted datasets are available only within our secure virtual data enclave.

News

ANNOUNCEMENT

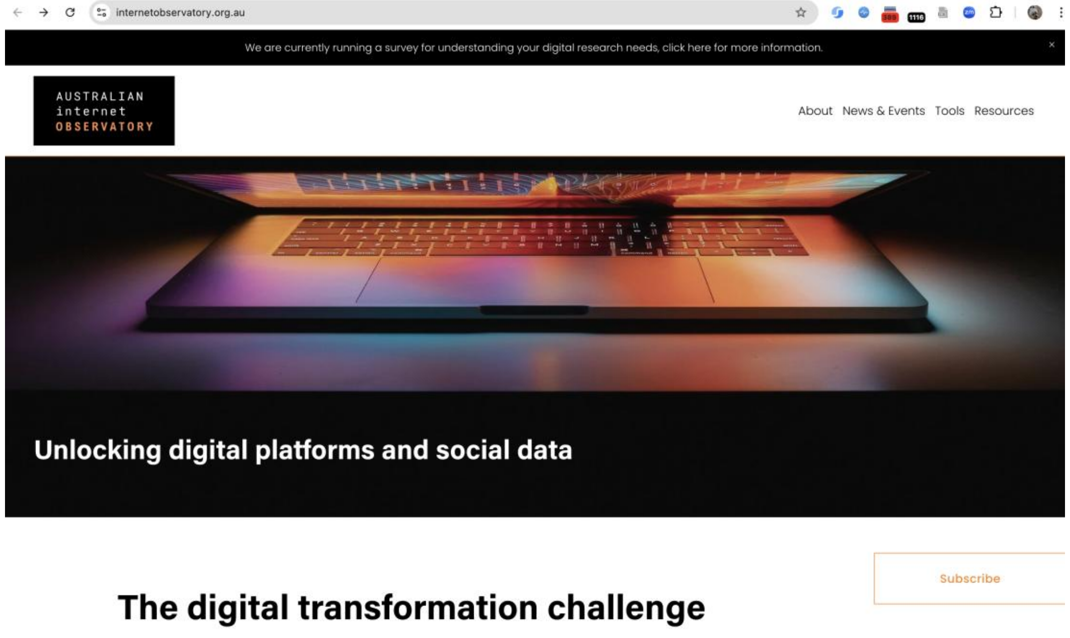
Democratizing Social Media Data for Research Impact

ANN ARBOR—Despite the massive amount of information created by social media,

Examples

Australian Internet Observatory

- Part of NCRIS national strategy
- https://internetobservatory.org.au/



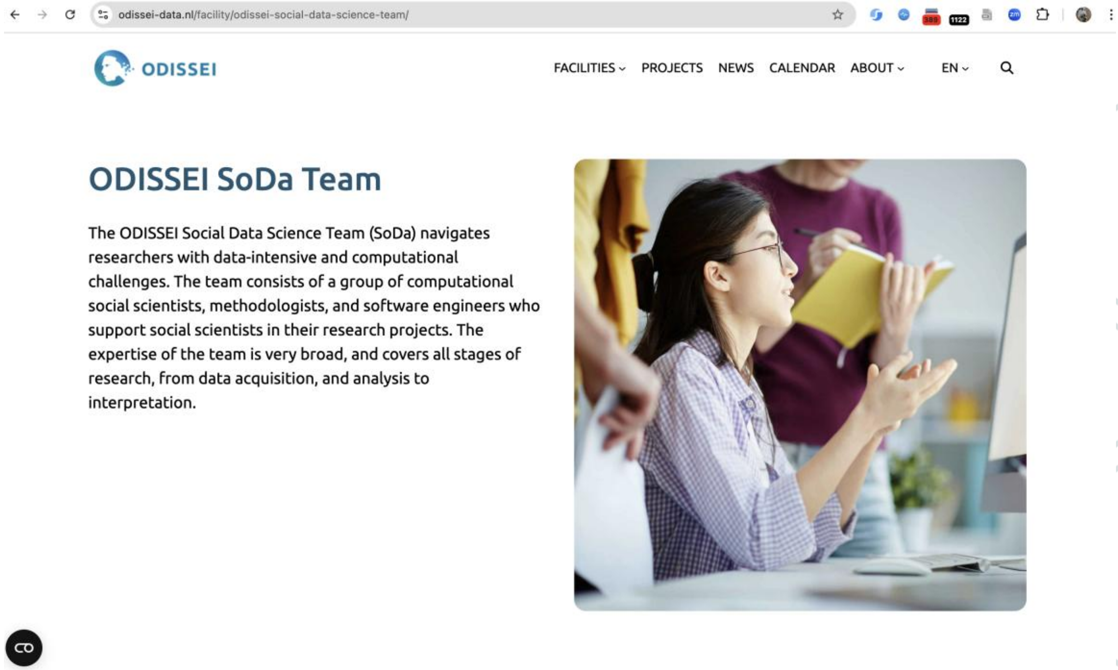
We are currently running a survey for understanding your digital research needs, click here for more information.

Unlocking digital platforms and social data

The digital transformation challenge

ODISSEI (Netherlands)

- Social Data Science team (Utrecht)
- https://odissei-data.nl/facility/odissei-social-data-science-team/



ODISSEI SoDa Team

The ODISSEI Social Data Science Team (SoDa) navigates researchers with data-intensive and computational challenges. The team consists of a group of computational social scientists, methodologists, and software engineers who support social scientists in their research projects. The expertise of the team is very broad, and covers all stages of research, from data acquisition, and analysis to interpretation.

Questions for an infrastructure

- What do you want to do?
  - What are your research questions?
  - What are your methodological approaches?
- What data do you want to do it?
  - Social media
  - Web tracking
  - Data donations
  - Survey data
  - …
- What do you want to do with the data?
  - Pre-process and clean
  - Integrate
  - Analyse
  - Integrate
- What do you want to do with the outputs?
  - Publications
  - Dashboards
  - Models
  - Data

How will you coordinate all of this?

Thank you!!

Questions?