Future Data Services:

Where are we now?

Economic and Social Research Council

August 2024

Recap: about Future Data Services

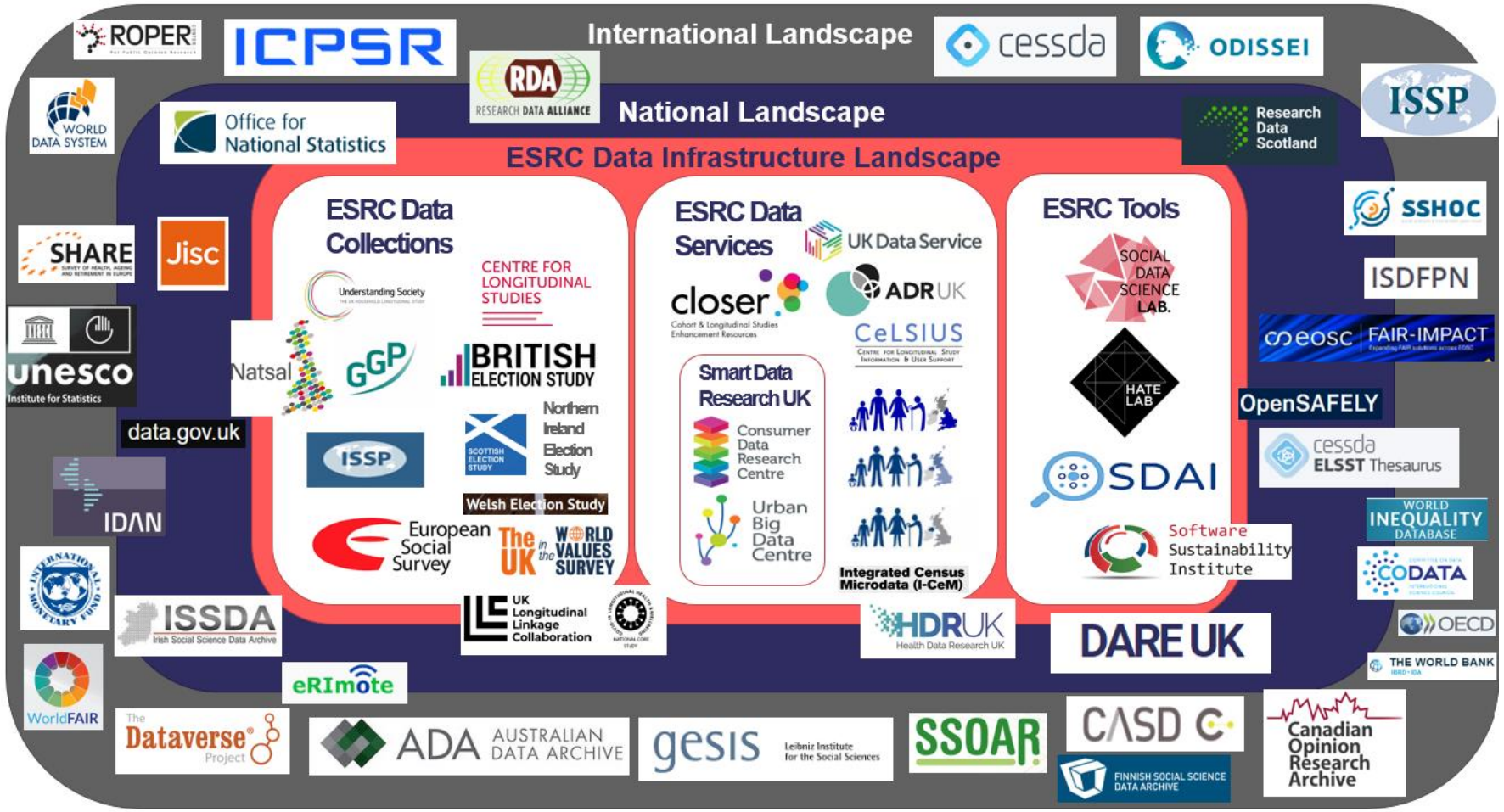
Economic and Social Research Council

Since 2012, the data universe has evolved…..

ESRC invested in new data services….

Other research councils did the same….

Government agencies have invested in data services too…



UKRI Economic and Social Research Council

What is Future Data Services about?

Question:

How can **existing and new data services** maximise the opportunities arising from **changes in data, technology and policy and legal frameworks** to support the research community?

To answer this:

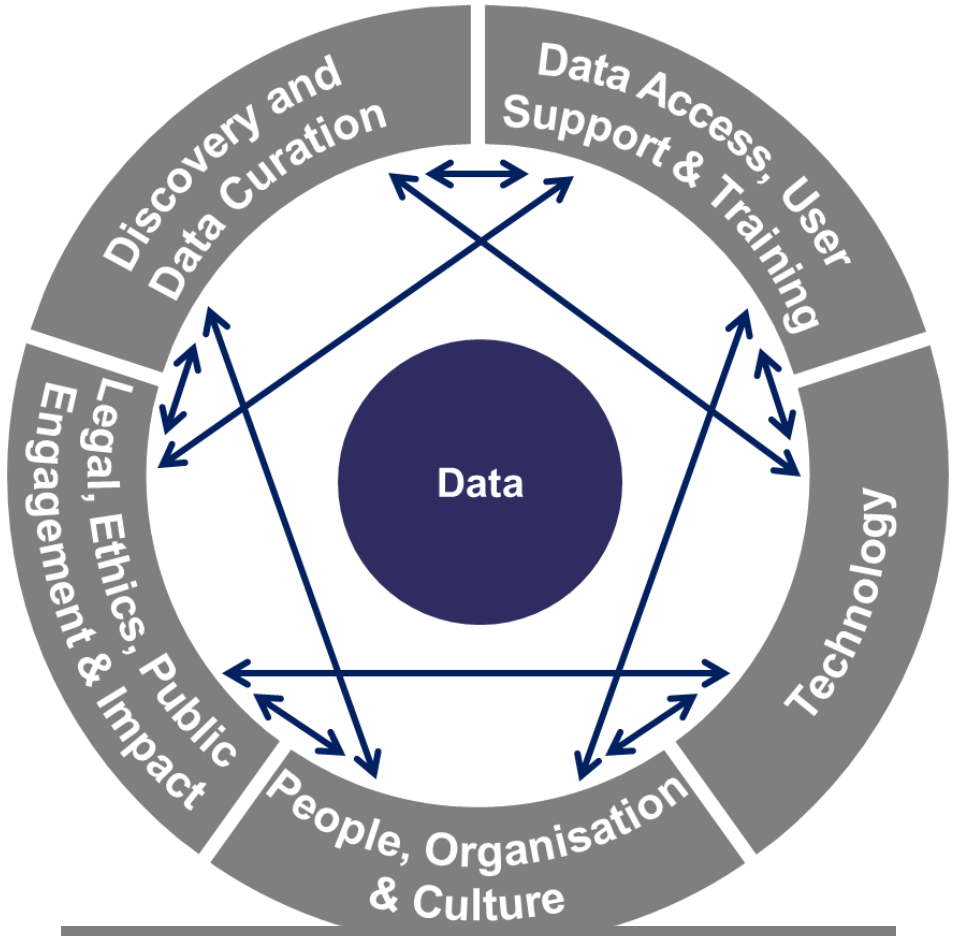
Our review will make a series of recommendations for future data service practices and funding to support the commissioning of new services.

- **determined by evidence** gathered from the research and data community
- underpinned with **theoretical principles** for delivering effective data services.

UKRI Economic and Social Research Council

ESRC will decide how to take these recommendations forward to create new funding opportunities from 2025 onwards.

5 FDS Themes



Economic and Social Research Council

Approach

Stakeholders

- Data infrastructures
- Data owners
- Researchers
- Policymakers
- Funders

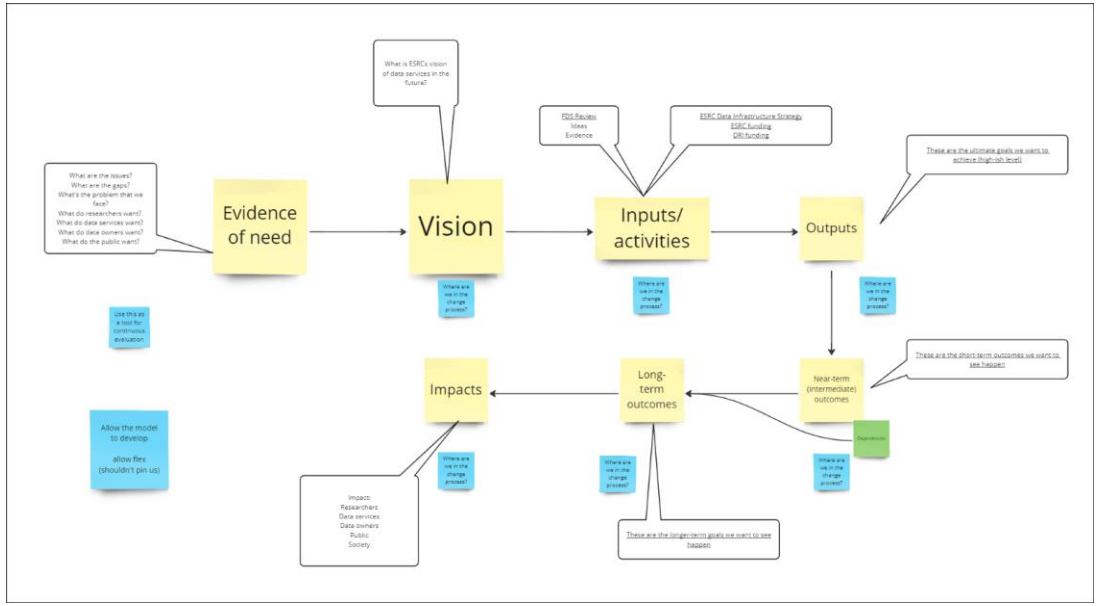
Theory of change concept

Activities

- Survey
- Interviews
- Roundtables
- Workshops (online via miro)
- Workshops (in person)
- Visits
- Slack expert groups
- UKRI meetings
- Monitoring new service developments

Our approach

- Theory of Change approach:
  - Ascertain evidence of need
  - Develop vision
  - Determine short-term and long-term outcomes (change)
  - Impact: does it meet vision and address evidence of need?



Economic and Social Research Council

Our progress

Economic and Social Research Council

Theme reports:
Challenges, Opportunities and Gaps

Economic
and Social
Research Council

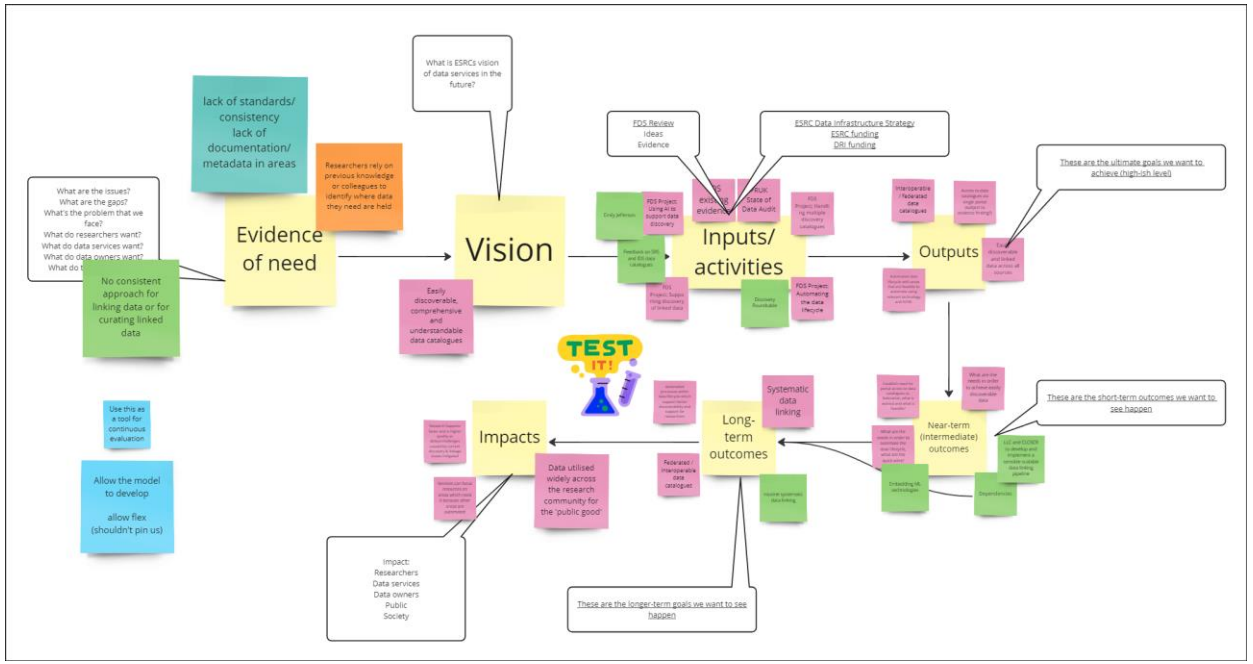
Theory of change: Discovery and Data Curation

50 draft recommendations to be refined:

- Resources
- New technologies
- Search functionality

EXAMPLES

From…..	To….
No consistent approach for linking data	Systematic data linking across the landscape
Reliance on personal experience of colleagues to source data	Federated data catologues based on AI
Lack of documentation/metadata	Automation of data lifecycle processes



Economic and Social Research Council

Discovery & Data Curation

Challenges

- Ensuring proper citation of data
- Lack of funding streams for data curation activities
- Data spread out (& duplicated) across lots of locations
- Need for longitudinal search functionality
- Lots of catalogues – particular issue for new users
- Lacking consistent standards and documentation

Opportunities

- Licences by user types
- Better signposting from a central location
- Harmonising parts of the data pipeline inc. standards and search terms
- Researching metadata functionality
- Set expectations for sharing and reusing data
- Research ready data
- Incentivise citation of data
- UX testing of services

Gaps

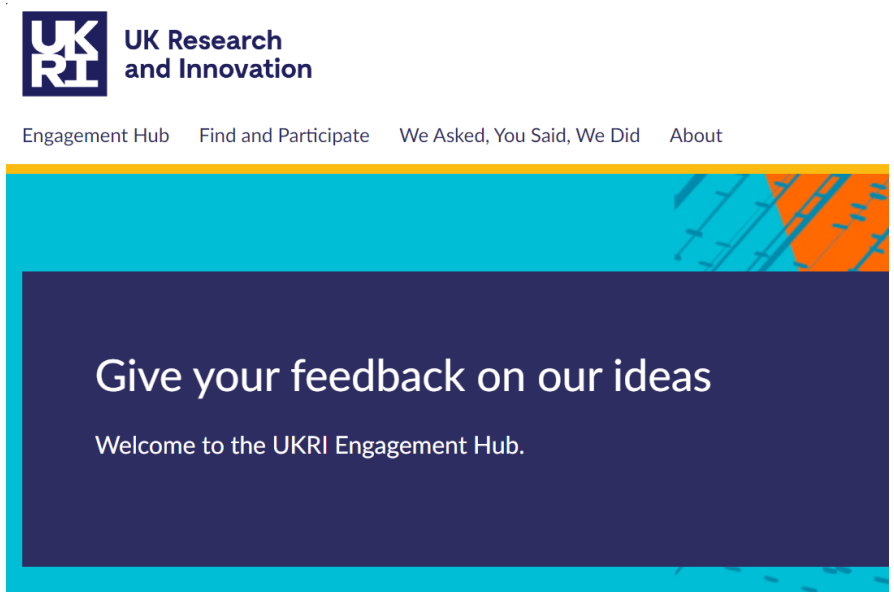
- How much time do studies spend creating and implementing data?
- What metadata research is happening or planned?
- What linkage exists and what falls through the gaps?
- Do we need a one stop shop for data?
- How do people find their data?
- How do we harmonise standards in a useful way?

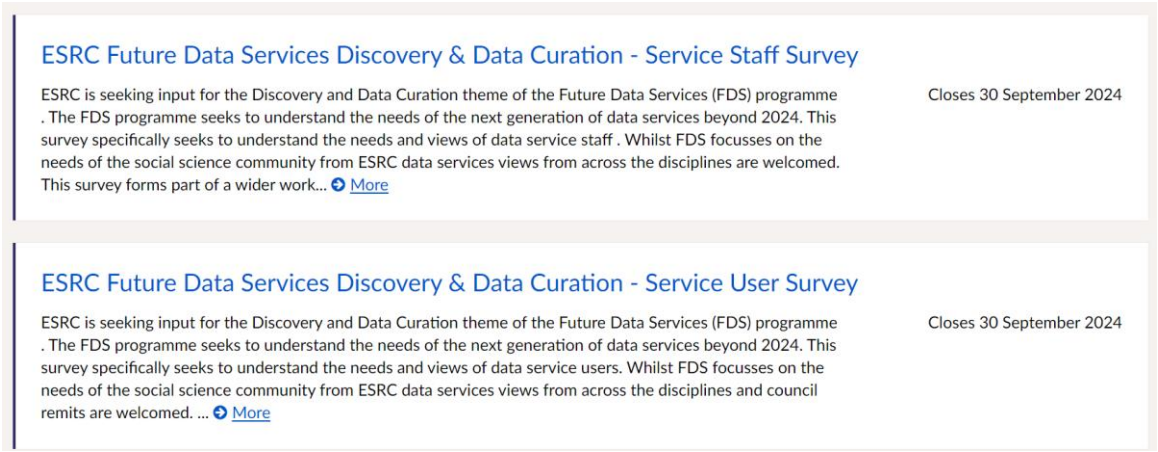
UKRI Economic and Social Research Council

Discovery consultation

1 – for data service staff

2 – for researchers





Economic and Social Research Council

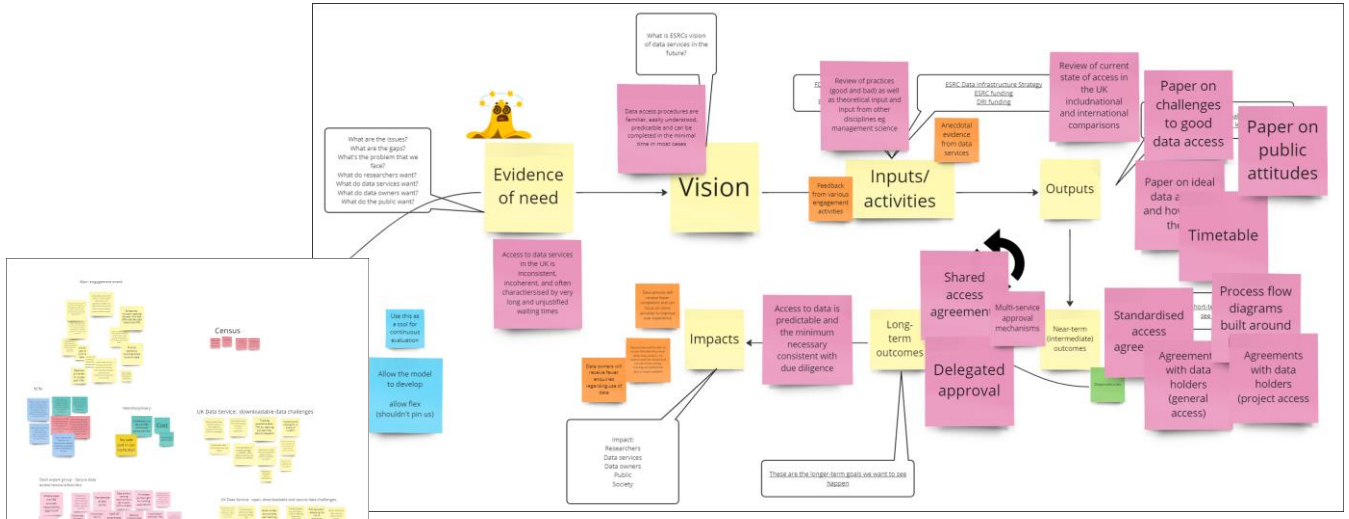
Theory of change: Data Access, User Support, Training

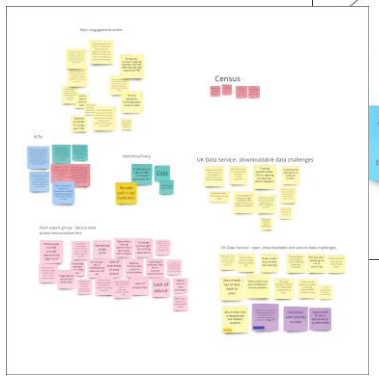
20 draft recommendations for testing:

- Cultural and system change
- New service commissioning
- Interventions to increase information and reduce friction

EXAMPLES

From…..	To….
Data access is inconsistent and incoherent	Standardised policies, tools and implementation
Accreditation of service facilities illustrates ‘pick and choose’	Universal and meaningful accreditation
Access to data presents a large cost to doing research	Access to data is quick and predictable





Data Access, User Support & Training

Challenges

- Data in multiple locations with different means of access
- Complex data = complex outputs
- Delays to access
- Political barriers to access
- Varying (sometimes high) costs of training
- User support can be resource intensive
- Accessing locations for secure data
- Culture problem

Opportunities

- Greater use of online training opportunities
- Centralised information regarding what can be accessed from where
- Coordinated negotiation for access e.g. with government departments
- Coordinating & committing to terminology to reduce confusion
- Scaled up support resources
- Training for data owners

Gaps

- What does and doesn't work for user support?
- How could training the trainer work e.g. creating 'super users'?
- What needs to be provided to free up more time for specialist support requests?
- What is required to agree this terminology and what should be included?
- Accessing qualitative data
- Tools for access

UKRI Economic and Social Research Council

Data access: Culture
DRAFT RECOMMENDATIONS

Awareness

- Support for data owners to increase their understanding of how researchers use data
- Data services should coordinate functions
- International data access should be an objective

Approach

- Consistent accreditation frameworks
- Standardised data access agreements
- Move from sequential to parallel processes
- Transparent decision making
- Better expectation of duration to approve
- More automation of processes

Aptitude

- Data owners should focus on outcomes, not processes
- Need for bespoke, technical expertise
- Better understanding of researchers through developing communities

UKRI Economic and Social Research Council

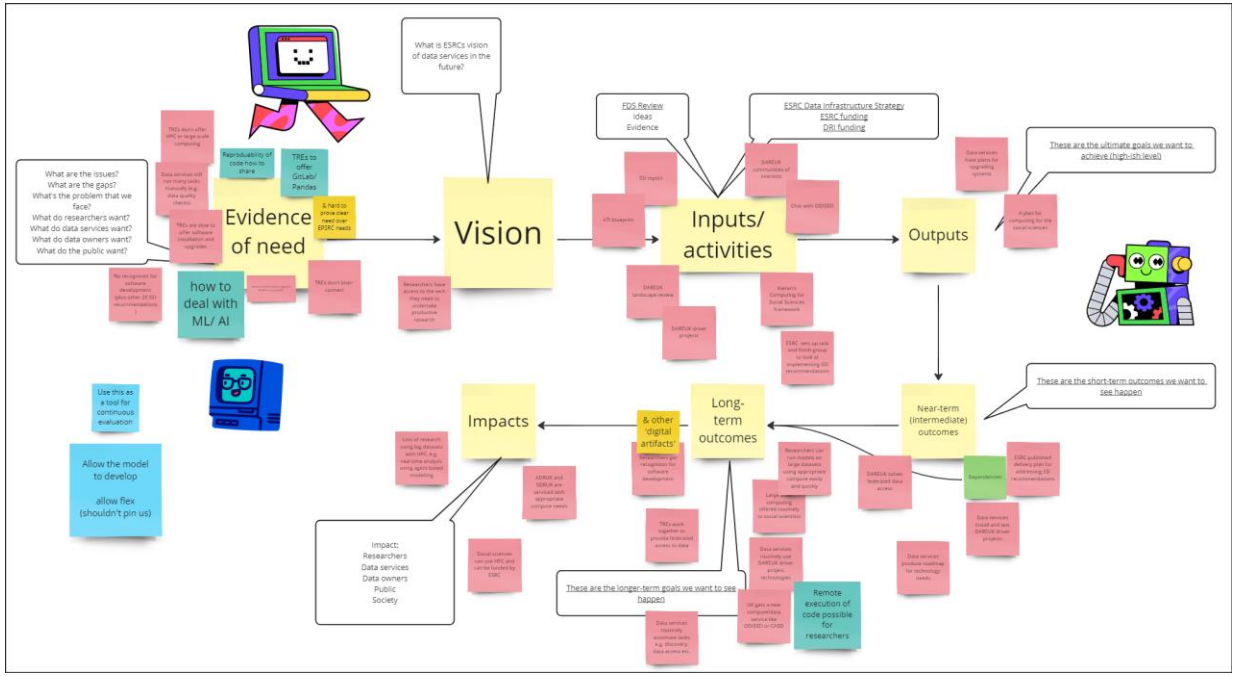
Theory of change: Technology

36 draft recommendations to be refined:

- Users and access
- Compute architecture
- Trusted Research Environments

EXAMPLES

From…..	To….
Compute is limited	Researchers have access to the compute they need
Outdated backend processes	Data services work efficiently
User requirements not met	Data services plan ahead to provide the tech researchers need



Economic and Social Research Council

Technology

Challenges

- Procurement & access to up-to-date technologies
- Environmentally sustainable technologies and systems
- Access to suitable ‘beyond desktop’ compute
- Too many accreditation standards
- Lack of coordination of software requirements
- Lots of TREs, often lacking capability
- Knowing which tech to use

Opportunities

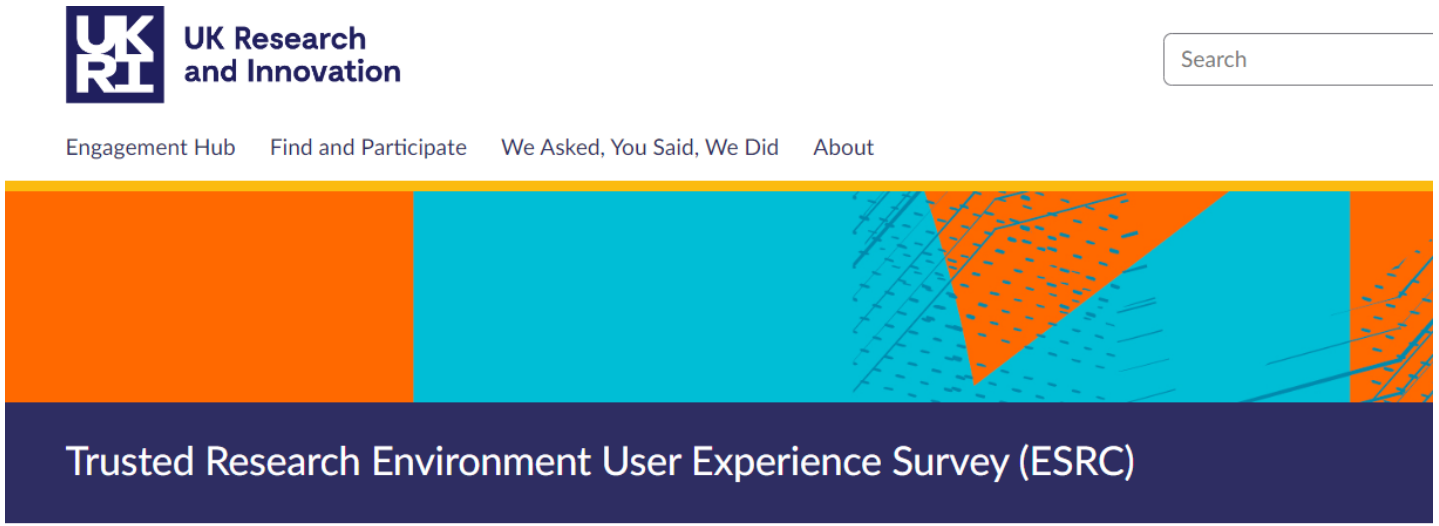
- Help from users to design systems
- Knowledge Exchange
- Breaking down of disciplinary silos to integrate useful user feedback
- Support to decide which service best fits your need
- Examine the benefits of federated/distributed services
- BYOD/BYOS
- Serving non expert users

Gaps

- What technologies do data services need for their own processes?
- What technologies do services provide to users?
- Why do people choose one facility over another?
- How to we verify software suitable for TREs?
- How are other countries supporting social research and compute?

UKRI Economic and Social Research Council

TRE survey



KEY TOPICS

- Applying to access data
- Safe Researcher Training
- Onboarding experience
- Easy of use
- User support
- Output release

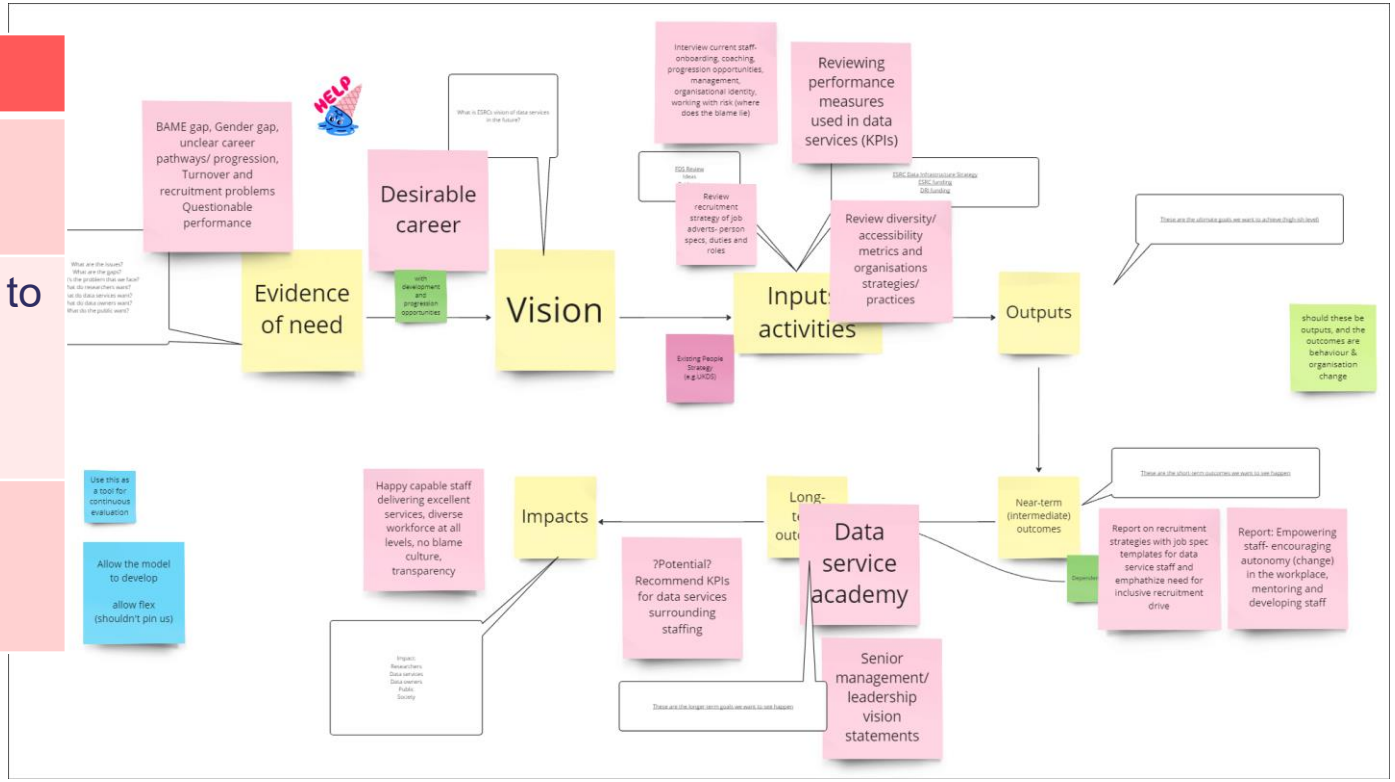
Theory of change: People, Organisations and Culture

Over 50 interviews with staff (varying seniority) across organisations, UK and beyond

- Career development
- Talent acquisition and retention
- Management and leadership

EXAMPLES

From…..	To….
Limited career development	Clear and desirable career paths
Varied and sometimes poor and inconsistent performance	Data service ‘academy’ to train and promote excellence and consistency
Lack of leadership and vision	Senior management ‘leadership and vision’ strategies



People, Organisation & Culture

Challenges

- Retention of data professionals, promotion and management of staff turnover
- Lack of defined progression opportunities
- Establishing culture of good work life balance
- Access to communities/networks
- Competitive compensation & benefits
- Closed door culture

Opportunities

- People & Leadership strategies for services
- Focussing on leadership
- Cross-service career frameworks
- Clear routes to progression
- Benefits such as access to communities/ networks/ training
- Evidencing contributions to data services in traditional academic metrics e.g. ref.
- Mentoring

Gaps

- What career progression frameworks exist?
- How do we make roles attractive? What additional benefits are needed?
- What is the relationship between data professionals and users? How can it be improved?
- What role could the Technicians Commitment (or similar) play in this space?

UKRI Economic and Social Research Council

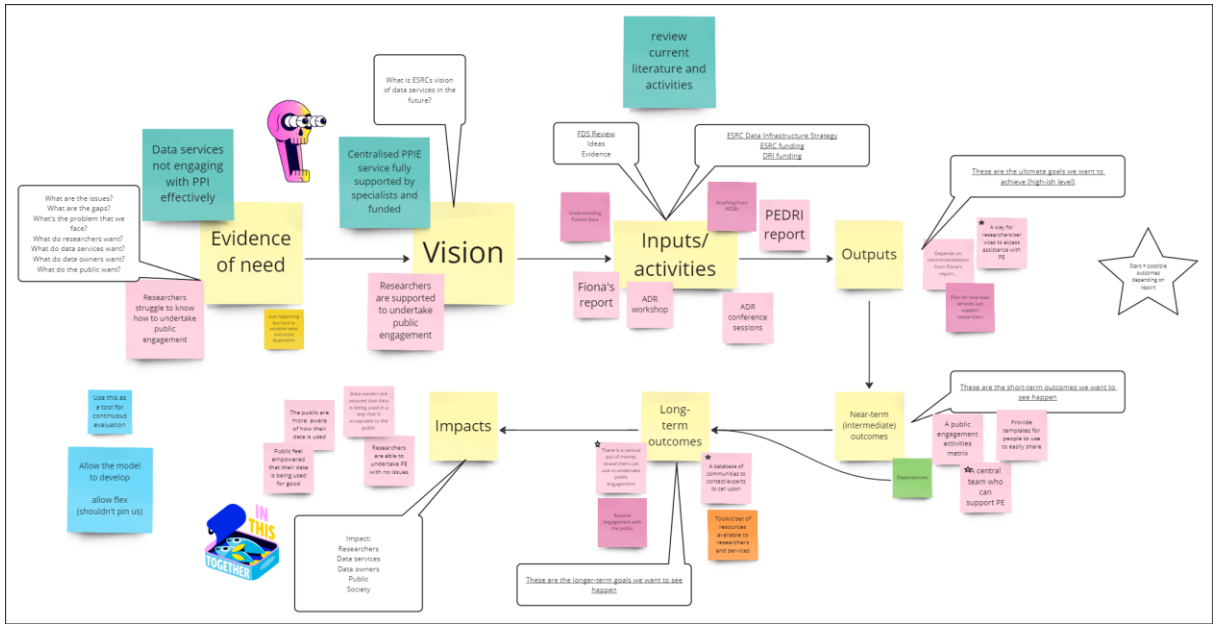
Theory of change: Public Engagement

29 draft recommendations to be refined:

- Doing public engagement
- Resources
- Talking about data

EXAMPLES

From…..	To….
Data services do not engage with the public	Data services regularly involve the public in their work
Researchers do not know how to get started with public engagement	Clear & accessible resources for all researchers
Good public engagement initiatives exist but not consistent or widespread	



Legal, Ethics, Impact & Public Engagement

Challenges

- Multiple processes
- Changes to the regulatory environment
- Lots of separate contracts
- Ensuring robustness of data sharing licenses
- Lack of public engagement
- Risks and perception regarding ethics e.g. of webscraping
- Inclusivity & representation in methods and data

Opportunities

- Standardised agreements
- Raising public awareness of data services and their role
- Provide researchers the resources to engage the public more
- Utilise existing public engagement opportunities
- Utilise community catalysts
- Exploring international best practice

Gaps

- How do we communicate differently to different audiences?
- What best practice examples exist for standardising contracts and agreements?
- How might passporting (approval applying to multiple locations) work?
- What would a gold standard of ethics require?

UKRI Economic and Social Research Council

Human-Centred Health

RECOMMENDATIONS TO SUPPORT PUBLIC ENGAGEMENT

Report for ESRC Data Strategy and Infrastructure

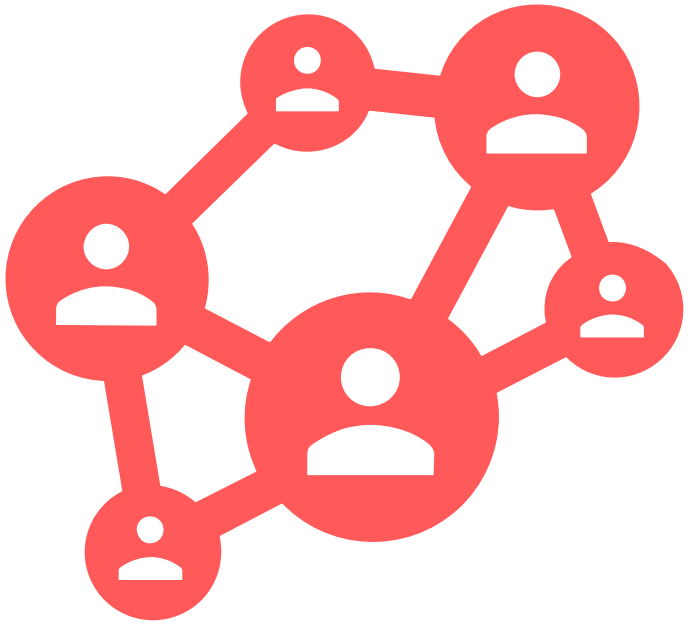
Recommendations

1. Agree on, or develop, a clear definition for public engagement, including scope
2. Define clear public engagement expectations for investments
3. Develop and share guidance for undertaking public engagement
4. Offer learning and support opportunities for ESRC staff and investments
5. Develop funding opportunities for innovative public engagement in investments
6. Continue to support and promote initiatives that encourage investments and organisations to work together for better public engagement
7. Role model leadership in public engagement through ESRC programmes and projects

UKRI Economic and Social Research Co

Theme X: Other Infrastructure Needs

- Dedicated service provision & support for qualitative data
- Promotion of ‘investment management’ based on theme/activity
- More ‘communities of practice’
- Shared goals, ambitions, strategic outlook
- Joint delivery of data service outcomes: moving away from silos to the grid….connection and harnessing



UKRI Economic and Social Research Council

Data service pilots 2024/25

Economic and Social Research Council

Data services: pilot funding 2024

£3m awarded to 9 pilots for 12 months from 1st April 2024

Includes £1.7m from UKRI Digital Research Infrastructure fund

- Data service federation
- AI and machine learning for data discovery
- Skills and capacity building for data service staff

- 2-3 touch points meetings planned (funders and data infrastructure to be invited)

Economic and Social Research Council

Pilot title	Lead organisation	Project Lead
A roadmap and rich metadata catalogue for the analysis of federated sensitive data	University of Edinburgh	Jano van Hemert
CORDIAL-AI: Innovating with Large Language Models at the UK Data Service for Census Information Discovery and Retrieval using Natural Language	University College London	Vassilis Routsis
Data discovery made easy: Applying ML to a diverse social science database	University of Portsmouth	Humphrey Southall
Enhancing Data Accessibility and Security through Innovative Data Synthesis (EDASIDA).	University of Manchester	Mark Elliot
Enhancing Data Services: A Comprehensive Training Program for Trusted Research Environment Staff in Social Science Data	University of Edinburgh	Atul Anand
Extraction and Utilisation of Metadata from Non-machine-actionable Documents to Improve Data Curation and Discovery	University College London	Jon Johnson
Harmony: A natural processing approach to data discoverability and harmonisation	University College London	Bettina Moltrecht
Optimizing Data Professional Success: Identifying Skills, Career Trajectories, and Training Requirements for Enhanced Data Service Delivery	University of the West of England	Elizabeth Green
Talk data to me! Evaluating the potential for large language models to enhance data discoverability across federated data services	University of Liverpool	Mark Green

4th July Workshop

FDS is generating a number of outputs, including reports and recommendations, to be published within 2024.

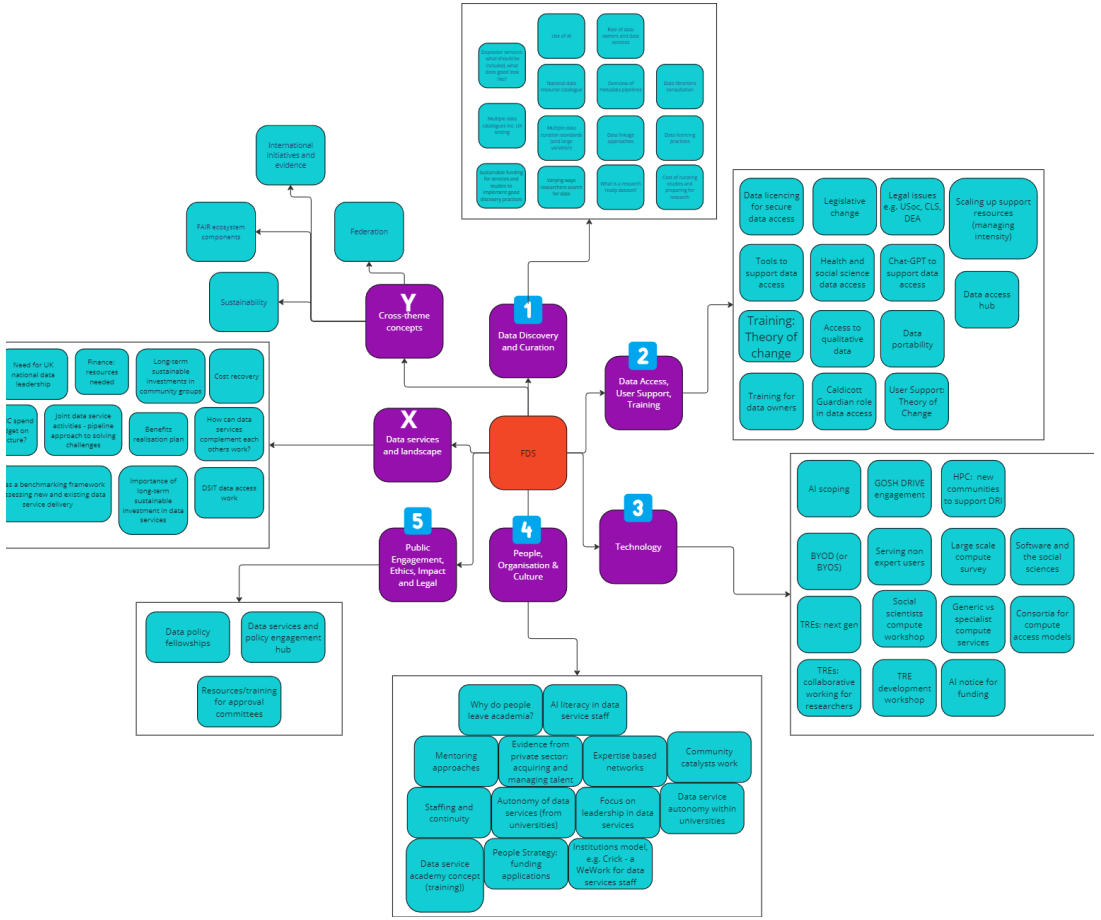
Main objective: review and **prioritise** of remaining challenges across all FDS themes.

- Issues were classified as High, Medium or Low Priority
- Issues which were not in scope of FDS were flagged
- Missing issues were identified

Economic and Social Research Council

Areas for further investigation

- Throughout our work, we have identified issues for each thematic area which we haven't been resourced to investigate further.
- On July 4th 2024, we held a workshop with UKRI colleagues to identify which we should prioritise.



Economic and Social Research Council

Workshop prioritisation results

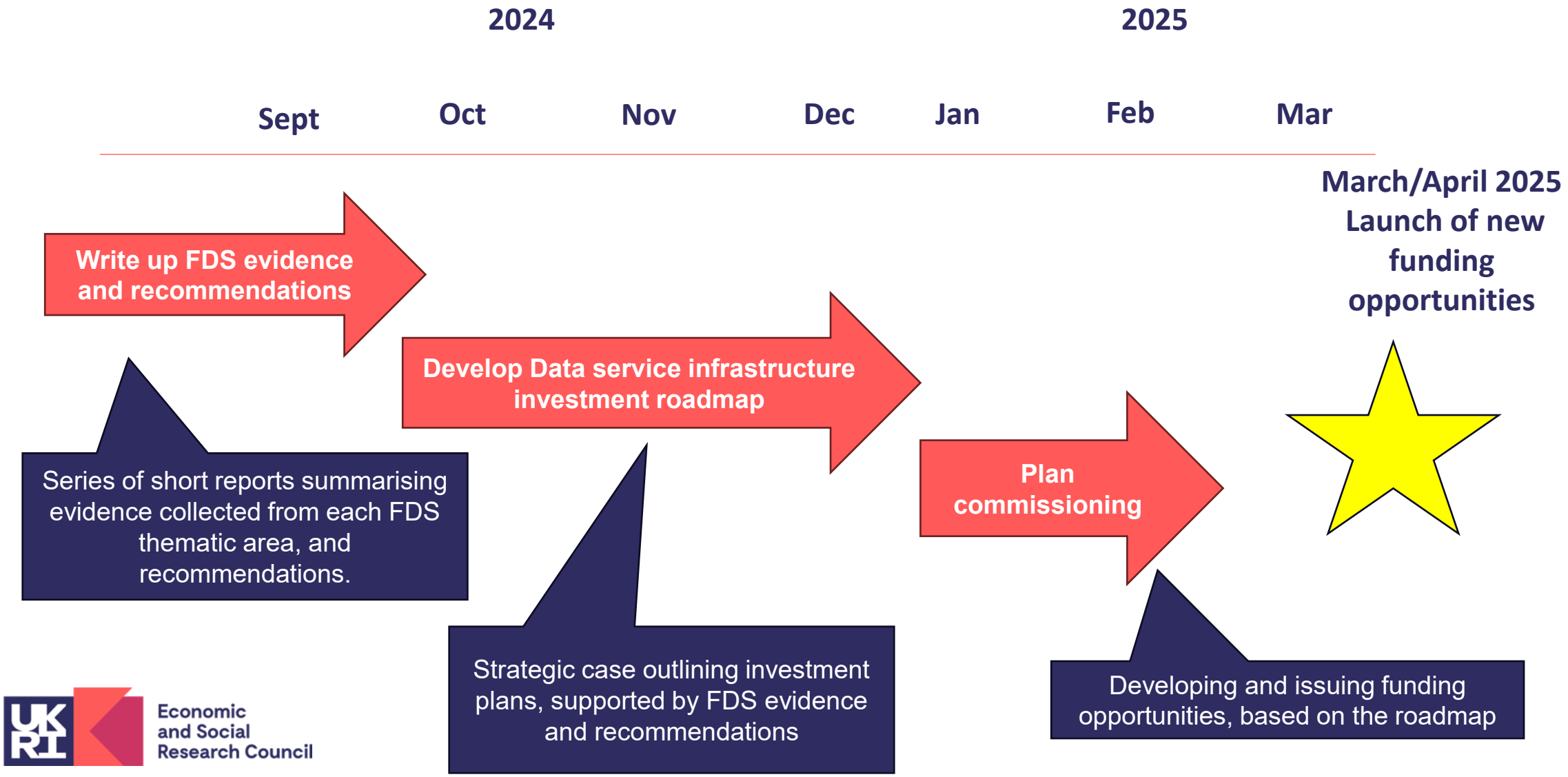
	Data Discovery and Curation	Data Access, User Support and Training	Technology	People, Organisation, Culture	Public Engagement, Legal, Ethics, Impact	Landscape and cross-theme concepts
High priority for FDS to review	• Varying ways researchers search for data • Multiple data curation standards (and large variation) • Sustainable funding for services and studies to implement good discovery practices • Depositor services: what should be included, what does good look like?	• Training: Theory of change • Scaling up support resources (managing intensity) • Data access hub • User Support: Theory of Change • Health and social science data access	• Bring your own device (BYOD) • Bring your own software (BYOS)	• Staffing and continuity • Focus on leadership in data services • Data service academy concept (training) • People Strategy: funding applications • AI literacy in data service staff • Write up output academy as a Proof of concept for other staff training		• FDS, ESRC and international connections • Need for UK national data leadership • Importance of long-term sustainable investment in data services • Sustainability
What's missing	• Continuous flow of data • Discovery tools working through different tools to get data on to a single platform • Highlighting use and growth of existing catalogues not just creating new ones • Grading datasets e.g. scoring based on coverage of data to assess utility • Ensuring we don't lose the sensitivity and credibility of a data source through federation and learning	• Commission something around legislative change & legal issues - low resource, high impact • Tools including software, synthetic data and a focus on streamlining access • Differing approaches to disclosure across government departments	• TREs from the user perspective • Technical issues affecting users of social science data and/or data service providers • Approaching AI from the skills angle – can people identify the right tool to fit their requirements?	• Focus on retaining skilled people within the sector and building a route for development • Look across investments not just within investments • Developing talent in the face of AI as the landscape is continuously changing and adapting – how do we adapt with it?	• Identifying legal frameworks that impede data use • Identifying good practice and case studies from those supporting applicants and services e.g. university ethics teams, policy engagement hubs • Emphasis on secondary data use • Active engagement and use of responsibility to support those applying for ethical approval to understand the full impact	• Pipeline approach to solving challenges • Options analysis to understand best quality approach to data services • Benefits realisation plan • Environmental sustainability of data infrastructure • Aligning progress with policies e.g. open access

What's Next?

- Discovery & Data Curation – *synthesis of user and service staff surveys, report collating evidence and refined recommendations*
- Data Access, User Support & Training – *data access report collating evidence and refined recommendations, theories of change for user support and training*
- Technology – *synthesis of TRE survey, refinement of existing reports (e.g. compute framework) and refinement of recommendations*
- Public Engagement – *review of report and approaches to application*
- FDS Pilots – *ongoing investment management and planning for follow on*
- Future Commissioning – *Utilising prioritisation exercise and recommendations to establish themes for commissioning 2025 and onwards*
- Data Infrastructure Planning – *collating a blueprint for way forward and inputting to ESRC workstream activities ahead of ESRC Council in November 2024*

UKRI Economic and Social Research Council

FDS timeline



UKRI Economic and Social Research Council

Economic and Social Research Council

Thank you

ESRC: Economic and Social Research Council

@ESRC

Economic and Social Research Council