Smart Data Research UK

UKRI Economic and Social Research Council

Strategy Engagement Exercise

August 2023

Contents

1. About this engagement exercise 3
Summary 4
Questions 5
How to respond 5
Upcoming events 6
Key milestones 6
2. The data opportunity 7
Challenges in using smart data 14
Building on strong foundations 16
3. Aim and objectives 17
Aim 18
Objectives 18
Thematic pillars 19
Theory of change 20
4. Delivery approach 21
Programme structure 22
Equality, diversity and inclusion 26
Monitoring and evaluation 27
End notes 28

1. About this engagement exercise

Summary

Smart Data Research UK is a new UK Research and Innovation investment. It will deliver a step change in the use of new forms of data generated through everyday interactions with digital devices – including mobile apps, digital marketplaces, social media, wearables, satnavs, sensors, and smart technology.

Smart data, sometimes referred to as 'digital footprints data', can provide new insights about major social challenges such as energy security, the cost of living, climate change, emergency response, health, wellbeing and social mobility.

Developing research partnerships to use this data requires dedicated resources, interdisciplinary expertise, robust oversight, and input from the public. Our programme builds on over 10 years of learning from related projects including the **Big Data Network**, **ADR UK**, **HDR UK** and nearly 60 years of Economic and Social Research Council funding for research infrastructures like **Understanding Society** and the **UK Data Service**. Feedback from our 2021 consultation was instrumental in developing the case for Smart Data Research UK, and we are grateful for the on-going support of our stakeholder community.

Smart Data Research UK has been shaped by researchers, data experts, industry leaders and policymakers. We want your input into our strategy, and we invite your responses to the guiding questions below.

Your feedback, challenge, and ideas will help us build a better programme, securing the benefits of smart data for society in the most responsible and sustainable way. Thank you.

Questions

Q1 **Do you agree with how we have defined the data in scope for the programme?**

Q2 **What advice can you share regarding the challenges and opportunities in delivering this programme?**

Q3 **Do you agree with how we have framed the aim and objectives of the programme?**

Q4 **What do you think of the four proposed thematic pillars?**

a. Will they provide the right balance to address key social, economic and environmental challenges, and engage a wide range of research disciplines?

b. Will they provide a helpful basis for partnerships with organisations who hold smart data?

Q5 **Do you think our delivery approach is the right one to meet our objectives?**

a. What mechanisms should we put in place to ensure the separate data services are integrated and coherent from a user perspective?

b. What are the main risks to successful delivery of this model?

Q6 **What support can you provide to us in achieving our objectives?**

How to respond

Responses are invited via the UKRI Engagement Hub by the 6 October 2023 **at this link**. If you would like to discuss the strategy, partnership opportunities, or your own research ideas with us, please reach out via email to: **smartdataresearch@ukri.org**

We will also be seeking indications of interest from research organisations who may be interested in submitting a proposal to deliver parts of Smart Data Research UK. For more information, please refer to the **UKRI Funding Finder**.

We also welcome any other comments you would like to submit.

Upcoming events

Our Strategic Advice Team convenes an advisory group and conducts workshops and other engagement activities to provide the strategic hub with independent advice and prime the community to participate in future activities and funding calls.

The team and advisory group reflect a diversity of disciplines, skills, and approaches across the social sciences and beyond, as well as representatives from business, the third sector, and government.

Open workshops will be held this autumn on a range of topics related to Smart Data Research UK. Event details and registration links will be provided **here**.

Key milestones

2023/24

- ESRC Strategic Hub team established
- Strategic Advice Team and advisory group established
- Strategy engagement published
- Partnership development underway
- Funding call for data services
- First accelerator scheme awards made

2024/25–2025/26

- Data services commence
- New industry partnerships in priority sectors established
- Second accelerator call (methods, tools and pilots)
- Ethics and legal hub in place
- Public engagement activities commence
- First research publications and impact case studies
- First accelerator projects complete

2026/27–2027/28

- Deepening of industry partnerships, increased multilateral licences
- Third accelerator call
- Second accelerator projects complete
- Federation of data services develops
- Mid-point evaluation

2028/29

- Well established partnerships in place
- Third accelerator projects complete
- Gateway 5 review completed
- End point evaluation
- Sustainability plan in place

2. The data opportunity

The data opportunity

Smart data offer a powerful opportunity to understand society, improve lives, and sustainably grow our economy. But there are commercial, methodological, and technical challenges in using such data for research. Smart Data Research UK will build on the strong foundations laid by previous investments in this field to deliver a step change in the use of smart data for research and innovation.

The growth of smart data, brought about by the rapid adoption of digital technologies, has created a global opportunity. This opportunity was accelerated by the expansion of our digital lives during the pandemic, with the shift towards online services and the increasing adoption of digital transactions.

Data within our scope

Physical environment, mobility and infrastructure

- Navigation systems
- Environmental sensors
- Public transport
- Satellite observations
- Connected vehicles

Consumer and retail

- Online shopping activity
- In-store behaviour
- Loyalty cards
- Checkout scanner data

Social media

- Networks
- Cultural participation
- Mis- and dis-information
- Recommendation systems

Finance and economy

- Electronic payments
- Card transactions
- Job websites
- Real estate apps

App-derived data

- Smart phone use
- Media consumption
- Games
- Gambling
- Workplace tools

Smart devices and sensors

- Wearable technology
- Medical IoT
- Smart agriculture
- Connected and smart homes

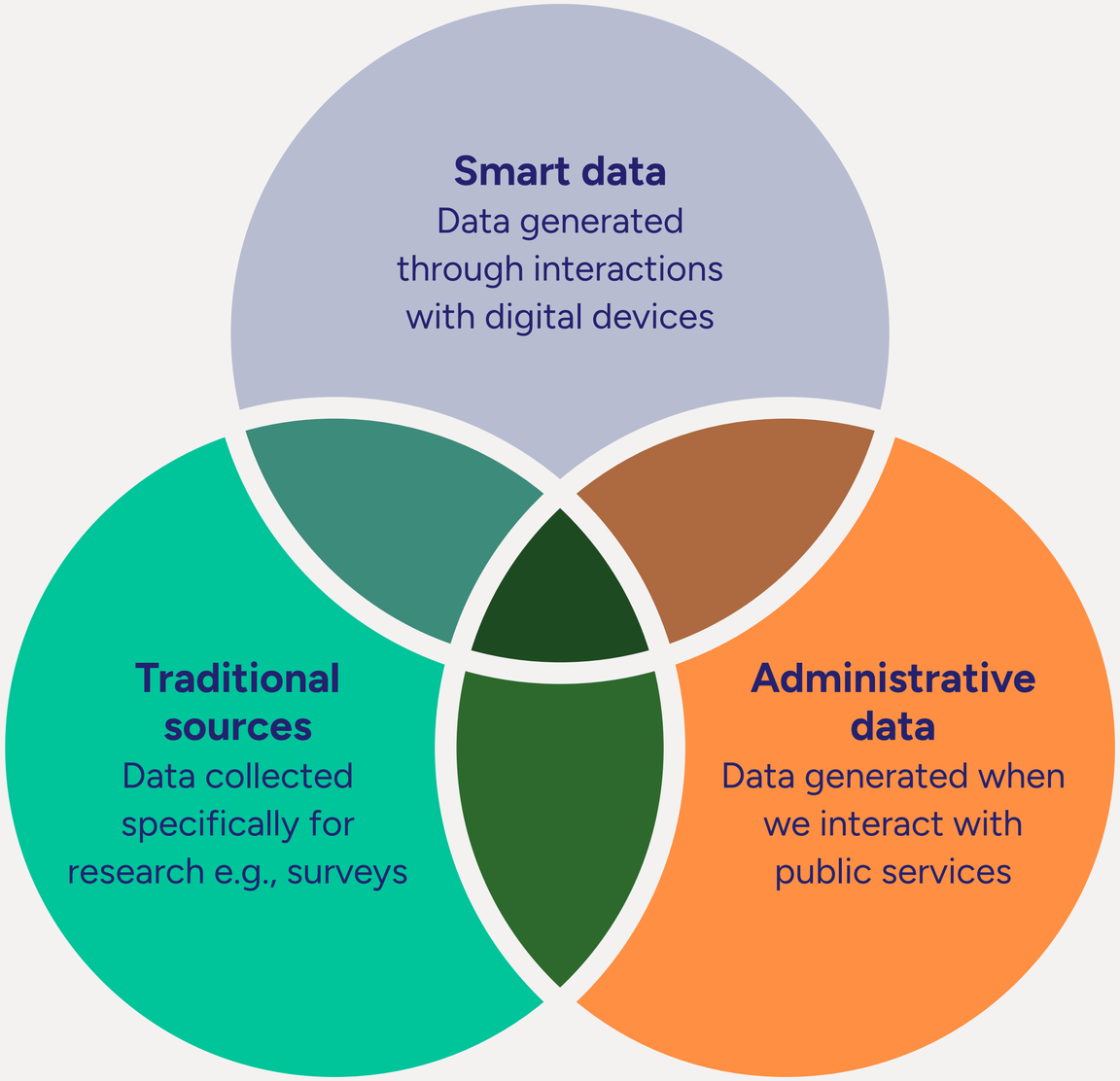
Q1 **Do you agree with how we have defined the data in scope for the programme?**

What is smart data?

We define smart data as the data generated through engagement with digital systems, devices and sensors[1]. These traces of everyday interactions with the digital world are sometimes referred to as 'digital footprints data'. Smart data comes from diverse sources, including mobile apps, navigation systems, social media, sensors in consumer devices and the environment, and digital transactions. It can be about individuals, organisations, systems, or the physical environment.

Smart Data Research UK comes at a time of increased recognition of the power of data. Many advances in Artificial Intelligence (AI) are reliant on access to immense stores of data combined with significant computing power and expertise. AI now offers a powerful tool to derive new insight from data, provided sufficient capabilities and safeguards are in place. The new infrastructure, services, tools and methods developed by Smart Data Research UK will unlock new research and innovation possibilities across all UKRI's research disciplines – building upon existing pockets of excellence as well as enabling new foundations to be laid.

The analytical power of smart data is increased when researchers combine it with traditional sources and administrative data. Smart Data Research UK is informed by, and will work to promote interoperability with, other forms of data served by UKRI investments.



Four data opportunities

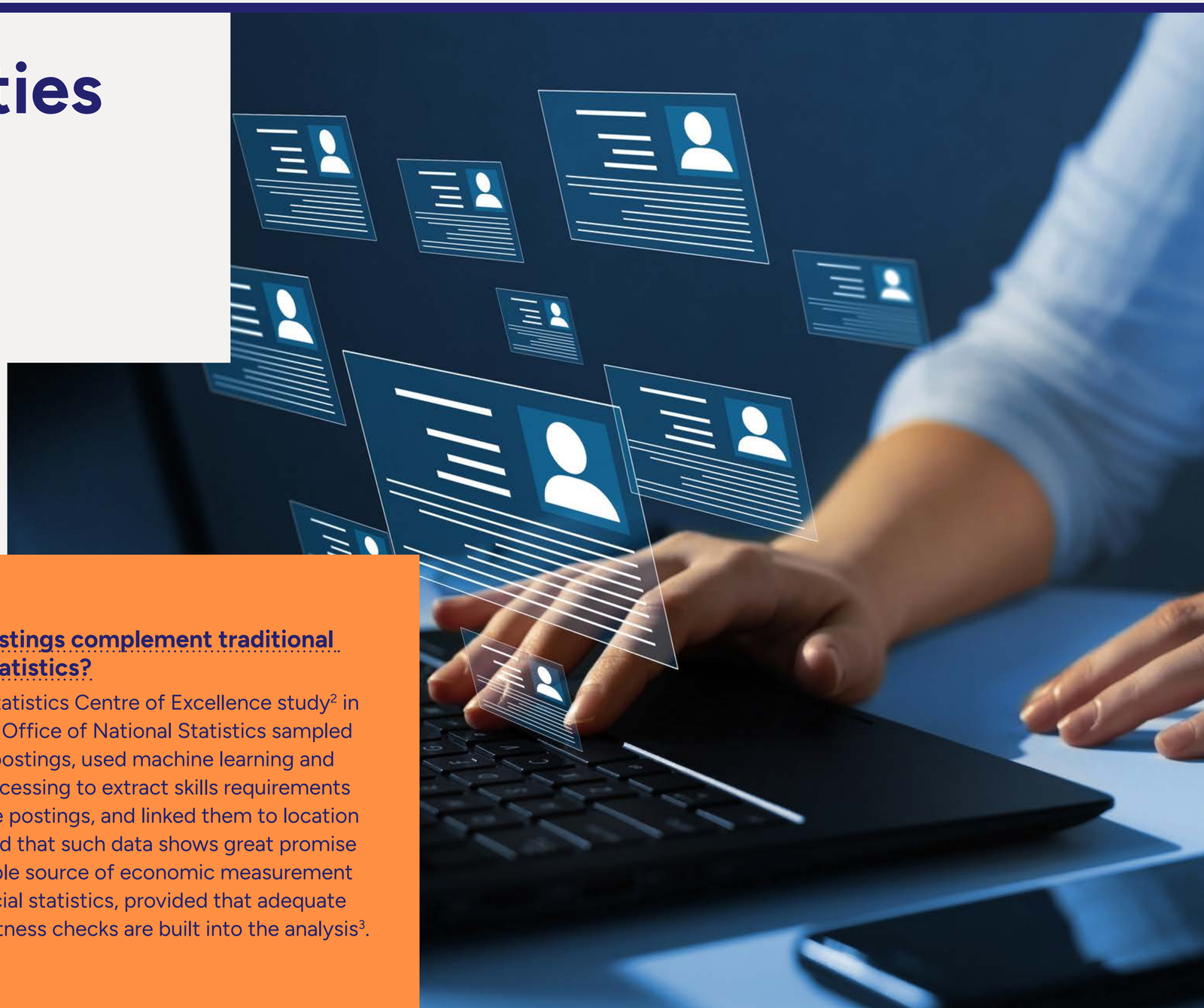
Opportunity 1:
Productivity and prosperity for all

Data generated through sources as diverse as professional social networks and card transactions present opportunities to address research questions on how to increase UK productivity and deliver prosperity for all. Such data sources can allow us to study how people, firms, and financial institutions interact and allow us to study links between finance, productivity and wages. This new knowledge can inform efforts to boost productivity and prosperity across the UK, underpinning research across a wide range of UKRI programmes.

Case Study

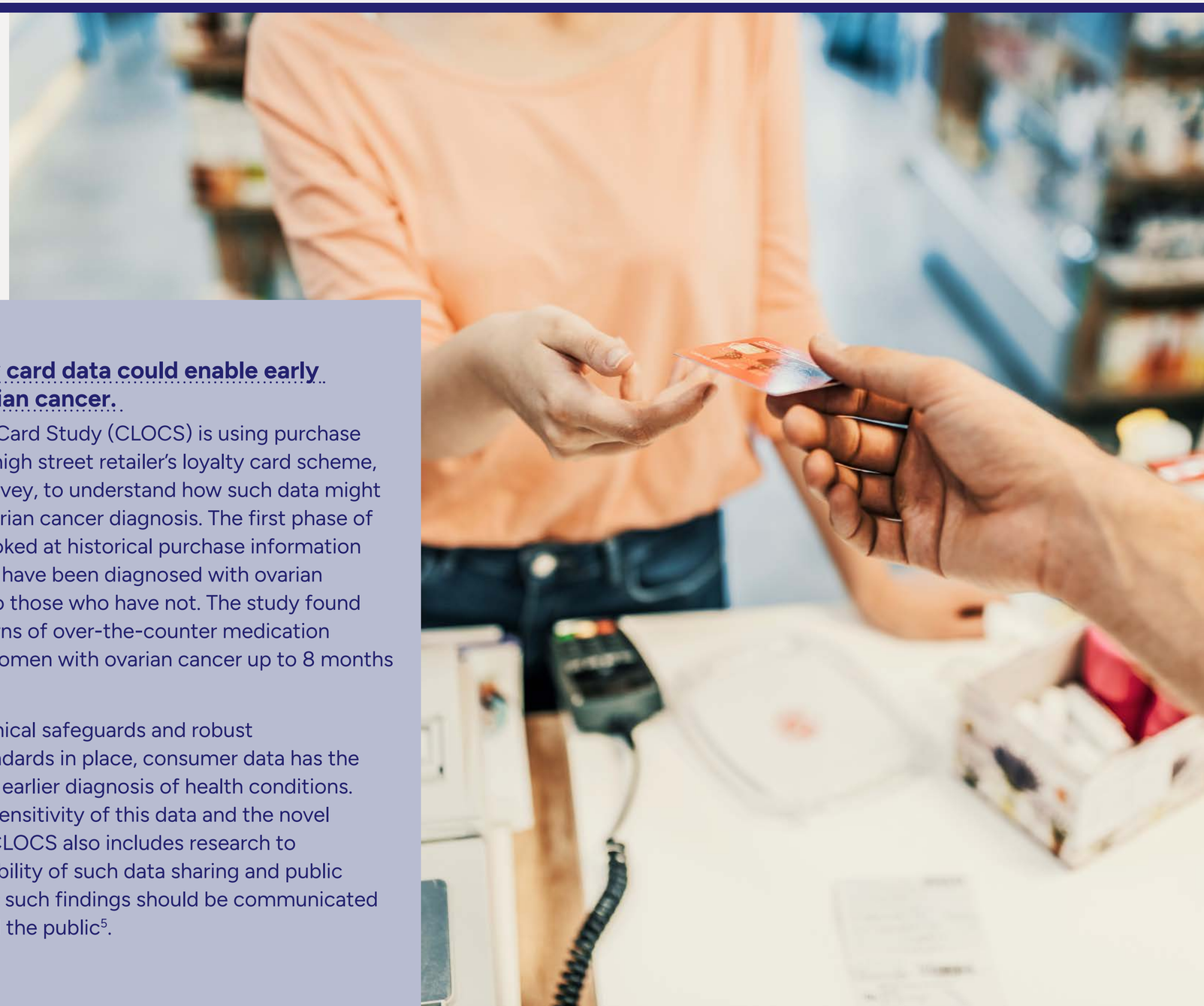
Can online job postings complement traditional labour demand statistics?

A 2022 Economic Statistics Centre of Excellence study[2] in partnership with the Office of National Statistics sampled 140,000 online job postings, used machine learning and natural language processing to extract skills requirements and salaries from the postings, and linked them to location data. The study found that such data shows great promise as a timely and flexible source of economic measurement to complement official statistics, provided that adequate validation and robustness checks are built into the analysis[3].



Opportunity 2:
Health and wellbeing

Our understanding of the complex relationships between health, wellbeing and specific behaviours, circumstances, and places can be greatly enhanced by data from sources like financial transactions, consumer loyalty cards, social media, and wearable devices. Such data sources can provide insights to improve population health and tackle inequalities, understand drivers of the mental health crisis, and help us live longer, healthier lives.



Case study

Pharmacy loyalty card data could enable early diagnosis of ovarian cancer.

The Cancer Loyalty Card Study (CLOCS) is using purchase information from a high street retailer's loyalty card scheme, combined with a survey, to understand how such data might reduce delays in ovarian cancer diagnosis. The first phase of this study (2023) looked at historical purchase information among women who have been diagnosed with ovarian cancer, compared to those who have not. The study found differences in patterns of over-the-counter medication purchases among women with ovarian cancer up to 8 months before diagnosis[4].

With appropriate ethical safeguards and robust methodological standards in place, consumer data has the potential to support earlier diagnosis of health conditions. In reflection of the sensitivity of this data and the novel methods involved, CLOCS also includes research to understand acceptability of such data sharing and public preferences for how such findings should be communicated back to patients and the public[5].

Opportunity 3:
Digital society

Data from social media, search queries, and online consumption can play a crucial role in helping us understand the challenges and opportunities posed by rapid advances in technology. They can provide insights related to topics such as online safety, misinformation, the digital divide, and the future of work. Research on these topics is often intertwined with other research areas, like health and productivity.

Case study:

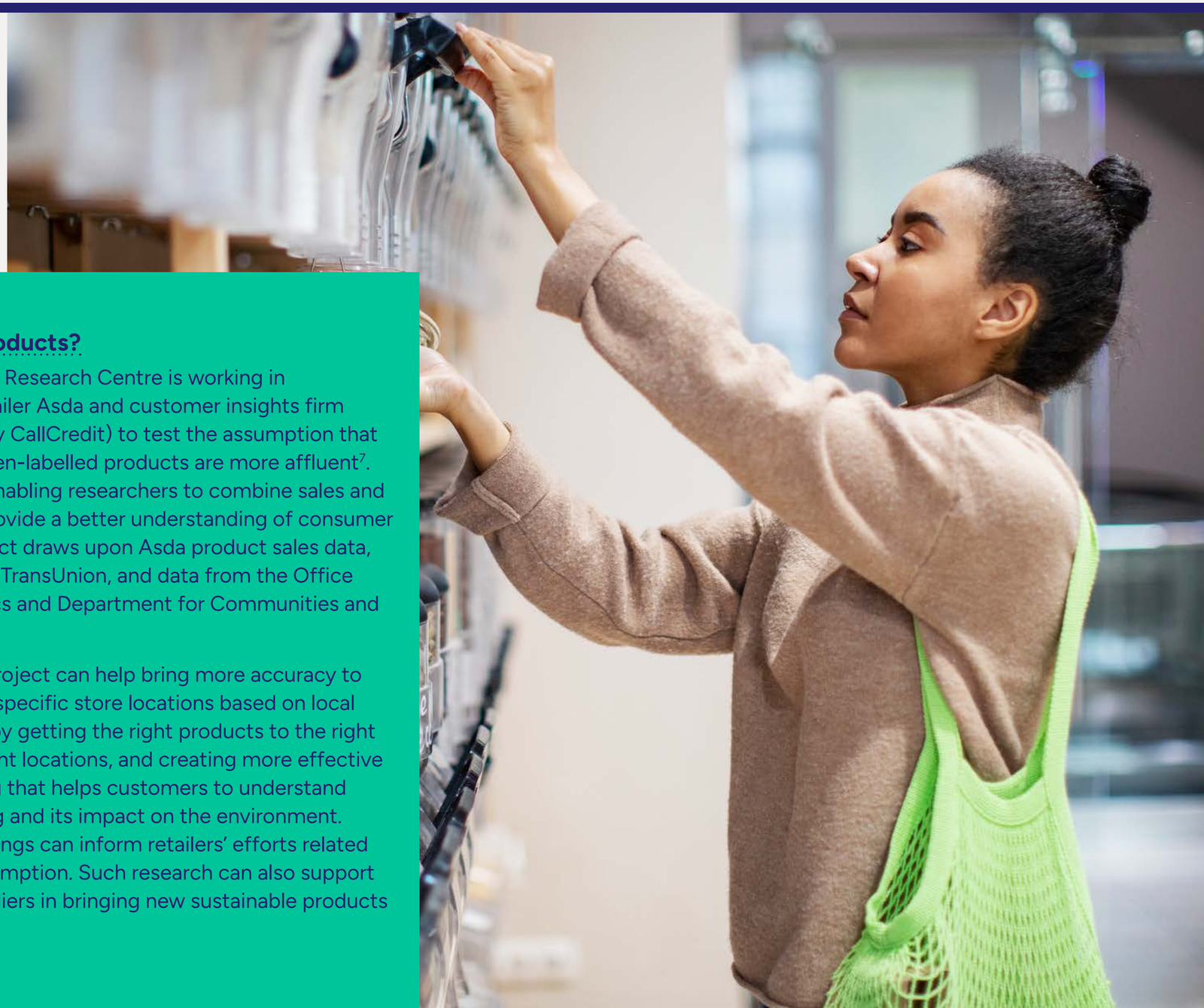
‘Suggested for you’? Understanding algorithmic recommendation systems.

Researchers and policymakers around the world are increasingly concerned with how online platforms promote harmful content or drive certain behaviours through algorithmic recommendation systems, shaping everything from how we feel about vaccines to the music we play. AI-powered chatbots like ChatGPT and Bard are bringing new influencing powers to the fore that are even less understood.

Interdisciplinary research using online data has a vital role to play in understanding algorithmic recommendation systems and their impacts on individuals, societies, and marketplaces. Researchers have already produced important foundational work related to recommendation systems[6]. However, further research in this area is a key priority for policymakers and regulators and can provide significant value to platform owners.

Opportunity 4:
Sustainability

Data generated through our use of devices and the data gathered by sensors have huge potential to support research related to topics like environmental sustainability, green growth, and the target of achieving net zero in the UK by 2050. UKRI has committed to building a portfolio that will fast-track the solutions necessary for meeting our net zero targets, and Smart Data Research UK has a key role to play in this effort, providing data, methods, and the required interdisciplinary capacity to keep the UK at the forefront of the green industrial revolution.



Case study:

Who buys eco products?

The Consumer Data Research Centre is working in partnership with retailer Asda and customer insights firm TransUnion (formerly CallCredit) to test the assumption that people who buy green-labelled products are more affluent[7]. This partnership is enabling researchers to combine sales and customer data to provide a better understanding of consumer behaviour. The project draws upon Asda product sales data, customer data from TransUnion, and data from the Office for National Statistics and Department for Communities and Local Government[8].

Findings from this project can help bring more accuracy to sales projections at specific store locations based on local demographic data, by getting the right products to the right customers at the right locations, and creating more effective labels and packaging that helps customers to understand what they are buying and its impact on the environment. Together, these findings can inform retailers' efforts related to sustainable consumption. Such research can also support innovators and suppliers in bringing new sustainable products to market.

Challenges in using smart data

Although we have seen examples of pioneering interdisciplinary research using novel data sources, many datasets cannot currently be used to their full potential for research: researchers are limited by insufficient data access and technical infrastructure, as well as underdeveloped methodology and ethical procedures.

Challenge 1: Protecting sensitive data and safeguarding public trust

For the UK to capitalise on the social and economic benefits that smart data can drive, the public must trust researchers to use their data responsibly.

People share their data with different organisations every day, under the protection of strict data protection regulations. In its raw form, data can include personal information about individuals, including their habits, health, location and behaviours. It is often precisely this detail which gives data its commercial value – for instance, enabling market analysts to understand consumer behaviour, to develop new products and tailored services.

For data to be shared with academic researchers, rigorous safeguards must be in place to protect personal information. For instance, the **Consumer Data Research Centre** (CDRC) and the **Urban Big Data Centre** (UBDC) use de-identification (along with secure research environments for storage and analysis), so that data provided to researchers does not include any information that could directly identify individuals. In research using smart data, de-identification may need to be accompanied by additional safeguards to prevent re-identification. The growing field of Privacy Enhancing Technologies (PETs) has the potential to support the use of these sources while protecting sensitive data.

To earn public trust, the public needs to be actively heard, engaged and involved in decisions around how their data is used. We will put this at the centre of our work through a programme of public involvement and engagement.

Challenge 2: Access to privately held data

Despite areas of significant progress, smart data can be very difficult for researchers to find and access because the majority of sources are owned by private companies, who face their own barriers to sharing data for research. Significant time can be spent negotiating on a case-by-case basis, leading to duplication of effort and a lack of capacity for industry to engage with multiple requests.

We have engaged with industry to understand the challenges they experience in sharing data with researchers. Companies must consider commercial sensitivities, their own resource costs in engaging with researchers, concerns about public perceptions, and – importantly – a lack of clarity about data protection legislation and how the law enables sharing for research purposes.

But we have found that significant incentives exist, as well. The work of teams who have successfully developed data sharing partnerships has shown that industry collaboration with researchers can generate business value in many ways: by validating and improving the quality and representativeness of data; improving products and supporting R&D; and by creating a pipeline of highly valued skills into a company.

Challenge 3: Methodology and quality

The characteristics and origins of smart data create research challenges distinct from other forms of data. The following brief examples[9] illustrate just some of the challenges that may arise:

1. **Representativeness:** Datasets are susceptible to both under- and over-coverage because only people interacting with a service are represented, potentially multiple times, rather than all individuals of research interest.
2. **Algorithmic confounding:** Many systems use algorithms to select content or induce specific user behaviours, which can produce confounding patterns in datasets.
3. **Unstructured data:** Sources such as social media, image, and sensor data are unstructured, or differently structured to traditional social science data. These datasets cannot be easily analysed in a relational database, and they require specialised techniques and data preparation to allow researchers to extract insights from them.

Challenge 4: Technical infrastructure

The data in our scope require technical infrastructure that is fit for purpose. This includes meeting stringent security requirements, the capacity to handle varied complex data types and, in some cases, the ability to manage very large volumes of data at speed and provide effective analytical environments for these. Capability to serve researchers with these data already exists in the UK, but existing resources need to be connected, supported, and complemented by additional capacity where gaps are present. Approaches such as the use of aggregated datasets also have a role to play in enabling and streamlining data access, while reducing the financial and environmental costs of storing and working with large datasets.

Building on strong foundations

Overcoming the barriers to using new forms of data is a significant undertaking, but we are not starting from scratch. Since 2013, UKRI has made a series of investments focused on improving access to data for impactful research and innovation[10]. **UBDC** and **CDRC** have proven that a 'ladder of engagement' model can build trust with industry through small, low-commitment projects building towards broader, multilateral data access agreements. **ADR UK** and **HDR UK** have demonstrated the value of connecting data resources and organisations on a national scale. The UK also has a range of highly capable institutes with dedicated social data science programmes[11], and infrastructure for safeguarding data through secure research environments.

A recent review by ESRC of the UK's existing data infrastructure for research highlighted many strengths, including increasing numbers of researchers using such data (in academia, business and government) and innovation in data supply, methods development and computational abilities. However, the UK's capability requires enhancement in the following areas:

- Data gaps: Many new forms of data lack availability to researchers, e.g. social media and population mobility data.
- Efficiency: Approvals procedures for researchers could be faster and simpler.
- Methods: Methodological frameworks for smart data research are in their infancy.
- Interdisciplinarity: Smart data research demands greater interdisciplinary collaboration across the social, natural and computational sciences.
- Skills: Capability and capacity building is required, including in computational, data appraisal and data management skills paired with social science expertise.
- Multilateral licencing: Data access agreements are often exclusive to projects or institutes.
- Reproducibility: Reproducible research practices (e.g. sharing workflows, models and algorithms) have not been broadly established, especially in work with confidential data.

Our work to build these capabilities through Smart Data Research UK has already begun, through the work of CDRC and UBDC, our Strategic Advice Team, and a **£2.5 million accelerator scheme.** Projects funded by the accelerator scheme will be assessed in two rounds, in December 2023 and January 2024.

Q2 **What advice can you share regarding the challenges and opportunities before us?**

3. Aim and objectives

Aim

Smart Data Research UK will unlock the power of new forms of data for research and innovation that solves social and economic challenges. We will do this by providing secure data access, safeguarding public trust, and building capability for cutting-edge research.

Objectives

Access

Provide secure data access

- Build long-term partnerships with data owners
- Deliver secure, effective digital research infrastructure

Trust

Safeguard public trust

- Demonstrate responsible research practices
- Champion deliberative public engagement in smart data research

Innovation

Build capability for cutting-edge research

- Lead a thriving and skilled interdisciplinary research community
- Solve methodological and technical challenges

Impact

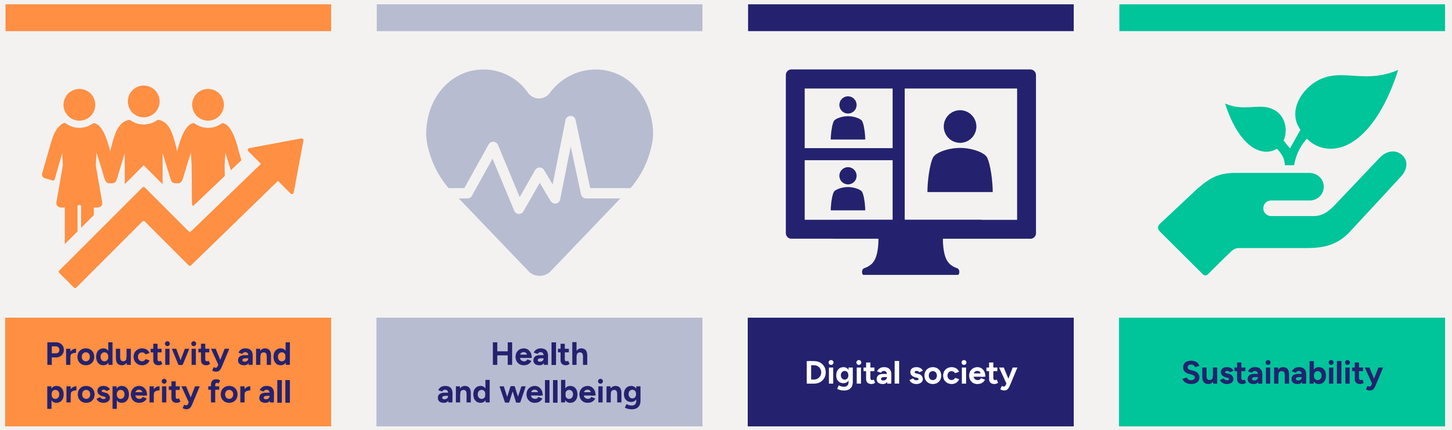
Generate social and economic benefits

- Support research that addresses significant social and economic challenges
- Track and publicise benefits from smart data research

Q3 **Do you agree with how we have framed the aim and objectives of the programme?**

Thematic pillars

Four thematic pillars will guide our priorities for data acquisition, collaboration, and partnerships, ultimately leading to research findings that are impactful and aligned with peer projects and priorities across UKRI, government, and beyond.



These pillars are designed to be broad, creating space for cross-disciplinary research and use of a wide range of data types. We have defined these through our Strategic Advice Team's consultation with the research community, alongside a review of government and UKRI priorities. These pillars will evolve over time, allowing Smart Data Research UK to be agile and resilient.

We will use the pillars to ensure that the infrastructure we develop delivers a critical mass of impactful research targeting UKRI priorities. The pillars will shape priorities for the data services, create grounds for partnerships with other UKRI programmes, government departments, and co-funders, and open doors to data acquisitions based on shared interests.

Q4 What do you think of the four proposed thematic pillars?

- Will they provide the right balance to address key social, economic and environmental challenges, and engage a wide range of research disciplines?
- Will they provide a helpful basis for partnerships with organisations who hold smart data?

Theory of change

We will meet our goals through activities delivered and supported by a range of programme participants. Our role and the wider community's role, including advice and input, are described below in section 4. The following diagram illustrates in summary how we and our partners will deliver to meet our goals.

Foundations

- Grants
- Multidisciplinary research groups
- Public engagement
- Ethics and legal support
- Scientific advice
- Partnerships
- Communicating our impact
- Robust governance
- Evaluating our work

Data access and stewardship

- Interoperable data services
- FAIR data
- Research-ready data and data products
- Integrated data catalogue
- Linking with other data sources
- Privacy preservation
- Secure researcher access
- Ethics review
- GDPR compliance

Making the most of our data

- New methods and tools
- Supporting communities of practice
- Proof of concept projects
- Increasing researcher capacity
- Increasing equity in data access
- Expanding reuse of data and replicability
- UKRI partnerships for further research and skills development

Enabling new insights to address major challenges

- Research grows in rigour and breadth
- Researchers have more capacity to use data in responsible ways
- New evidence-based findings on key policy issues are enabled
- We show how trustworthy research using smart data benefits society

Research findings and a strengthened research landscape support policy decisions and innovation

- Policymakers are supported in data-driven decision-making
- Data readiness increases across the academic, public and private sector
- Long-term academic/ industry partnerships support further innovation, collaboration and growth
- Other UKRI programmes produce further benefits by working with smart data

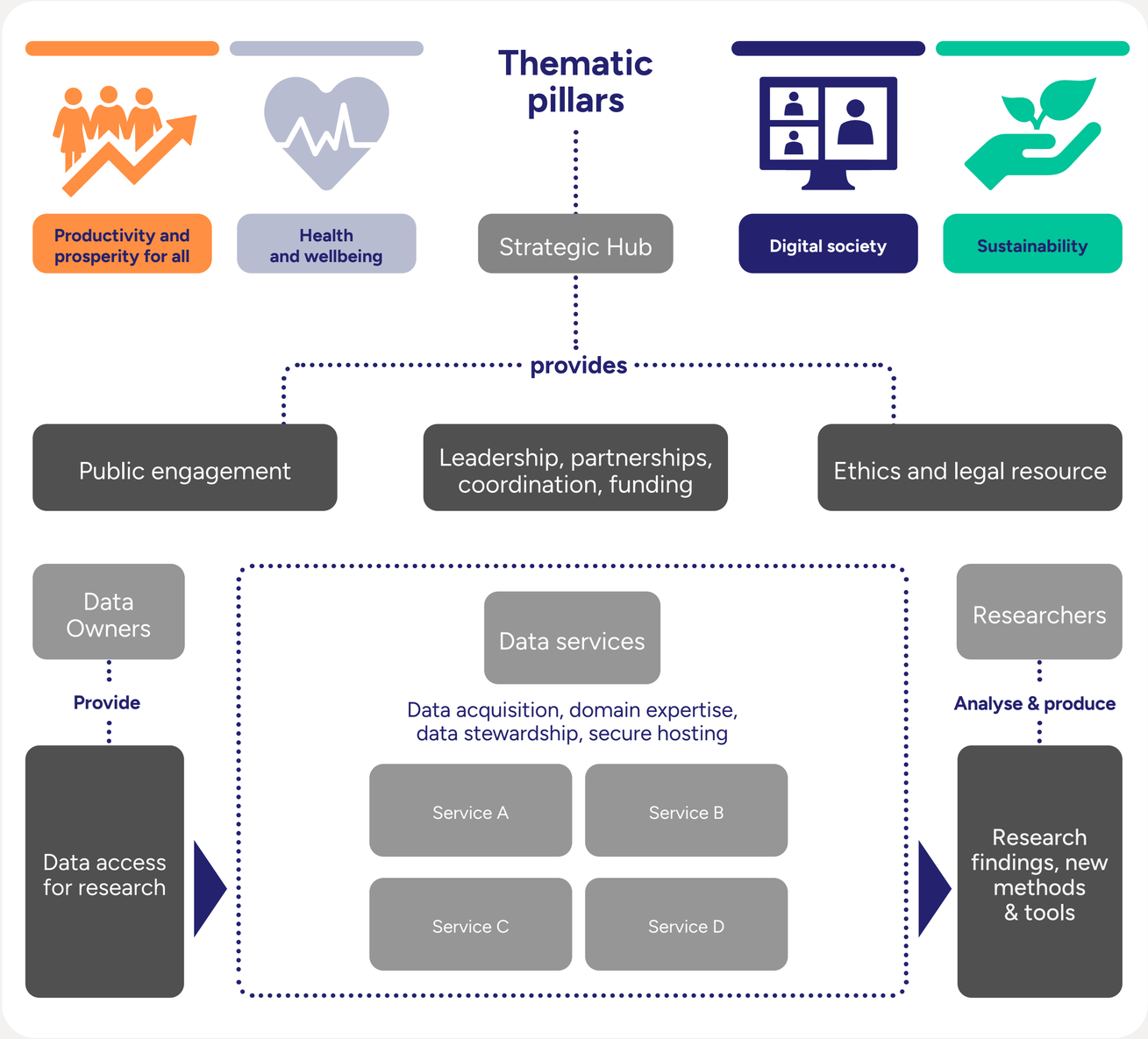
4. Delivery approach

Programme structure

Smart Data Research UK will see an investment of £49.3 million from the **UKRI Infrastructure Fund** from 2023 to 2029.

Building on lessons from **ADR UK** and other data infrastructure programmes and pilots, the programme will be delivered through a **centrally coordinated, federated infrastructure** comprising:

- A **strategic hub**, embedded in ESRC, providing **leadership, coordination and oversight,** a programme of **citizen-centred involvement and engagement** and an **ethics and legal resource** to provide support and assurance to the public, researchers and partners on responsible research.
- Federated **data services** cultivating long-term partnerships with data owners, providing curated access to data for priority areas of research interest. We will work iteratively to develop integration across the data infrastructures, working towards a federated model.
- Cross-cutting **accelerator funding calls** for projects to:
  - Develop new tools and methods
  - Build researcher capacity to work with smart data
  - Pilot access to and use of new data sources
  - Support interdisciplinary and cross-sector collaborations focused on pressing national challenges



Strategic hub

The Smart Data Research UK strategic hub will work with partners across the UK and internationally, leading and coordinating our programme through the following means:

- We will provide leadership, creating a strong network including businesses, government, academia, and the third sector to cultivate mutually beneficial partnerships and augment the data acquisition efforts of our data services.
- We will leverage ESRC's experience in overseeing digital, administrative, and survey research infrastructures to ensure effective governance for our data services and other activities.
- We will demonstrate to businesses that Smart Data Research UK should be their partner of choice for sharing data with researchers, and we will safeguard public trust through responsible practices.

We will lead collaboration with the data services to define and operationalise our approach to data standards, interoperability, governance, and the end user experience. This will allow the programme to achieve a balance between pursuing opportunities to integrate services, while keeping a degree of space for services to tailor their approaches to specific data types and research requirements.

Maintaining public trust, through demonstrating responsible stewardship and tangible benefits to society, will be essential for the success of the programme. The strategic hub will lead a programme of public involvement and engagement and will establish a dedicated ethics and legal resource. This will enable the programme to discuss key questions regarding the public interest, trust and transparency, and provide expert advice to researchers on ethics and privacy.

Federated data services

Smart Data Research UK will embark on an ambitious data acquisition programme through a cohort of grant-funded data services. The teams delivering these services will build long-term partnerships with data owners and deliver research-ready datasets to accredited researchers. We envision these services as embedded in existing or new centres of capability in the UK research landscape. The strategic hub will provide leadership and coordination to enable the data services to develop federated capabilities incrementally.

The diagram on the following page describes the capabilities and activities a data service is likely to provide. Specific capabilities and activities will vary based on a service's area of focus and maturity of the research landscape in accessing and working with the data being targeted.

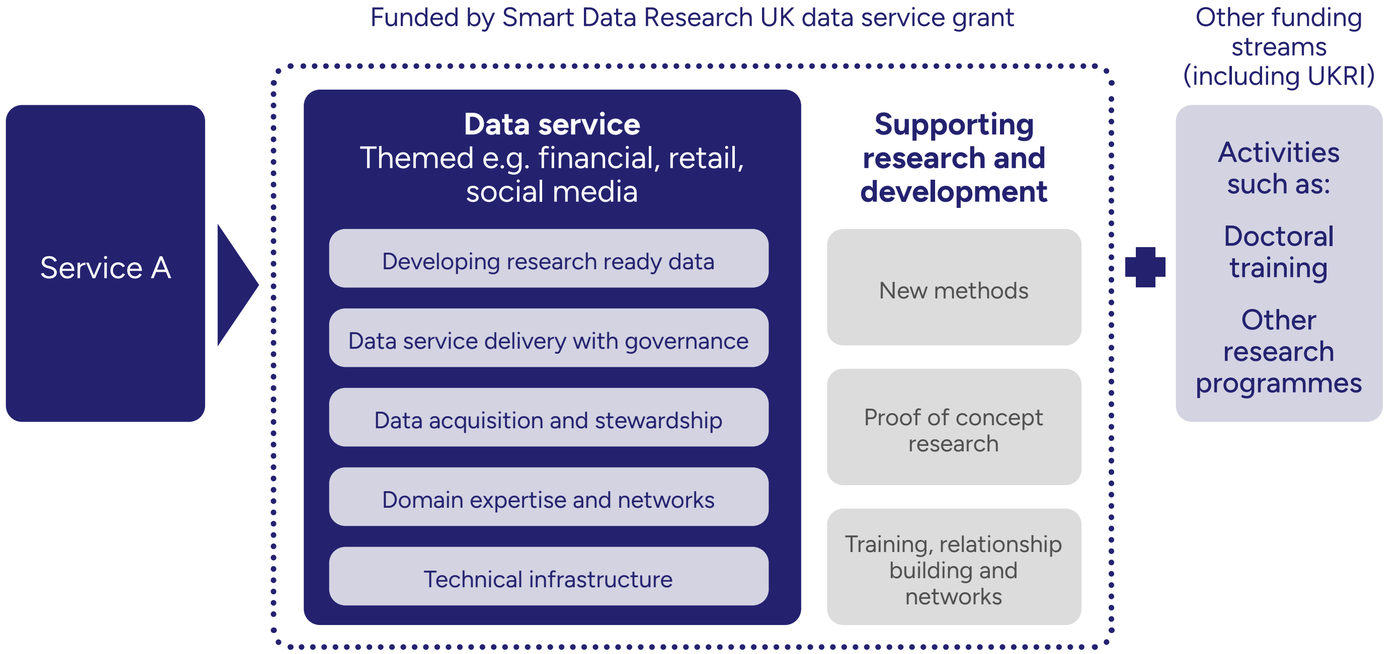
Data services will partner with other infrastructure providers to build on existing digital research infrastructure, ensuring funding can be devoted to filling gaps in the landscape and meeting technical needs arising from newly sourced data with unique privacy, security, and access requirements.

The data services will work collaboratively, coming together with the strategic hub to work on discovery and curation, data access, user support and training, processes and governance, and technological and methodological development. The following principles will guide decisions about data standards and interoperability, governance, and the end user experience:

- **Use and impact driven:** We will focus on addressing the needs of users, including researchers and data owners, and prioritising use cases where there is evidence of the greatest opportunity for impact in line with our thematic pillars.
- **Agile:** We will learn through small-scale testing and pilots, and we will be open to new ideas from the wider digital research infrastructure landscape.
- **Collaborative:** Through strategically aligned partnerships and close working, we will connect data, people and systems.
- **Holistic:** Technology is only one part of the picture: leadership, culture, institutions, legal frameworks and skills matter too.
- **Effectively governed:** We will assure that security, privacy, and licensing standards are managed in a consistent way across partners.

The full funding call for data services will open in early 2024. We will seek indications of interest from the research community through a process separate to this engagement exercise. The call for indications of interest will be posted on the **UKRI Funding finder.**

Data service capabilities and activities



Q5

Do you think our delivery approach is the right one to meet our objectives?

- What mechanisms should we put in place to ensure the separate data services are integrated and coherent from a user perspective?
- What are the main risks to successful delivery of this model?

Future accelerator funding calls

Smart Data Research UK will periodically publish funding calls, open to the research community for projects to:

- Develop new tools and methods, addressing major challenges around representativeness, uncertainty, provenance, bias and understanding of latent/proxy measures or causal insights
- Build researcher capacity to work with smart data
- Pilot access to and use of smart data sources
- Support interdisciplinary and cross sector collaboration focused on pressing national and international challenges

We will seek to collaborate with other UKRI programmes, working in partnership to drive use of smart data sources, methods, and infrastructure to foster research and policy impacts related to the pillars.

Case study:

Consumer data casts light on place-based food insecurity

The **Priority Places for Food Index** was developed by CDRC Leeds in partnership with **Which?** in response to the 2022 cost of living crisis. It is constructed using open data to capture complex and multidimensional aspects of food insecurity.

Building on the **CDRC e-Food Desert Index,** but with additional domains relating to fuel poverty and family food support, the goal of the Priority Places for Food Index is to identify neighbourhoods that are most vulnerable to increases in the cost of living and which have a lack of accessibility to affordable, healthy, and sustainable sources of food.

This project helps identify places within the UK where people are most likely to need support in accessing affordable food and also to understand why they have been so identified, whether because of a lack of retail provision, poor access to online supermarket deliveries, or high levels of deprivation and need.

Equality, diversity and inclusion

Smart Data Research UK is committed to promoting equality and participation in all its activities, including with external stakeholders, in funding decisions, and as an employer.

We will embed equality, diversity and inclusion across our activities, including:

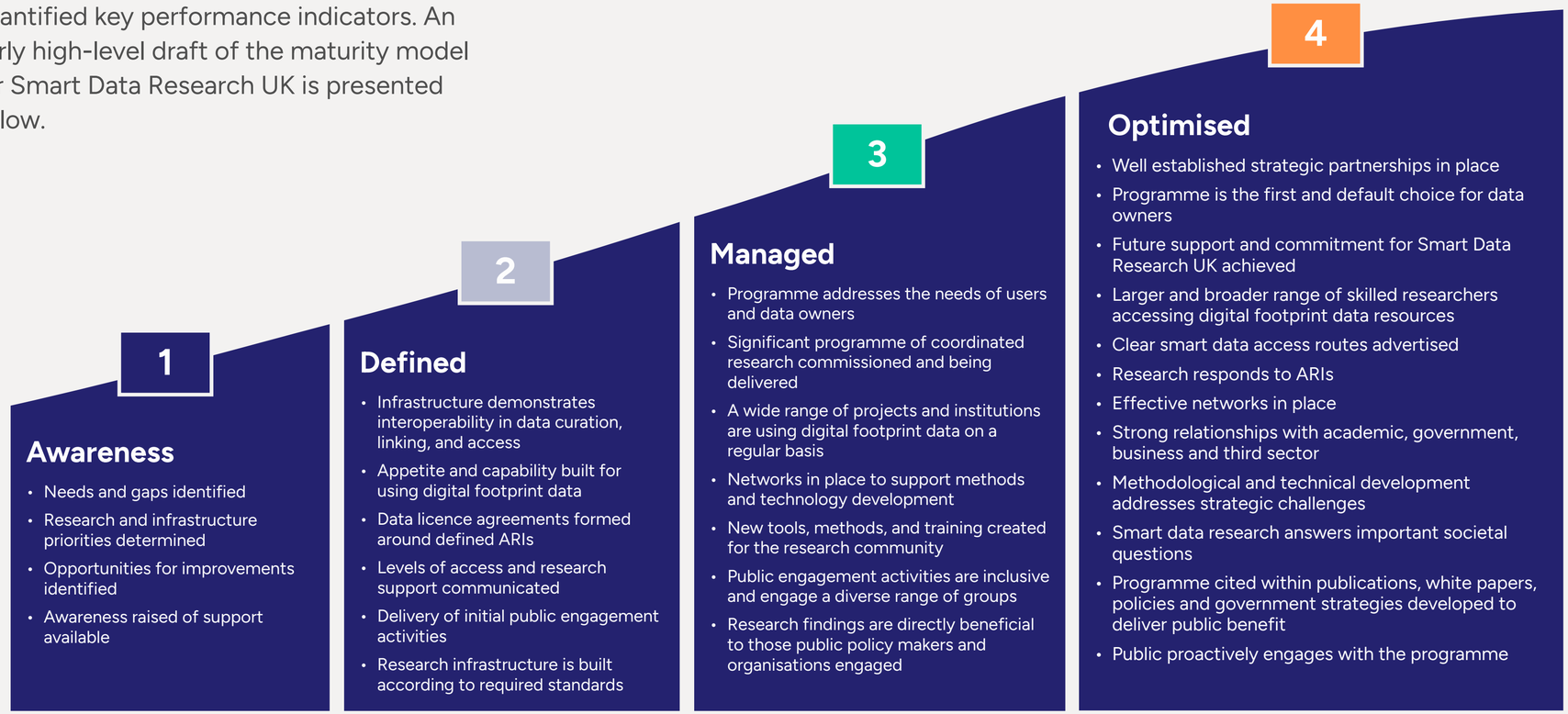
- Working from the **UKRI EDI strategy** and **ESRC's EDI living action plan** to ensure that our teams of researchers and data scientists are as representative as possible, including taking commissioning measures to foster inclusive practices and objective and fair decision-making.
- Enabling equitable researcher access through data services that serve researchers on equal terms.
- Unlocking access to new data sources that can provide evidence about hard to capture groups sometimes missing from administrative and other data sources.
- Developing rigorous methods to assess and evaluate bias and data gaps and to improve representativeness.

Monitoring and evaluation

We will evaluate progress against our strategy regularly. Targets for key performance indicators have been established, and we will track our progress in meeting these targets on a quarterly basis.

We have committed to an external evaluation of Smart Data Research UK, to ensure that the impact of the programme can be independently demonstrated, and to identify lessons learned from delivery. Given the innovative nature of our programme, a rigorous evaluation will be particularly valuable in creating a new evidence base on the impact of data infrastructures.

Sitting alongside formal evaluation, maturity modelling will assess programme performance towards each of the strategic objectives. Maturity modelling will provide a nuanced qualitative lens for assessing the overall performance, complementing the quantified key performance indicators. An early high-level draft of the maturity model for Smart Data Research UK is presented below.



Q6 **What support can you provide to us in achieving our objectives?**

End notes

1 Our definition of smart data extends the definition developed by the **Smart Data Working Group**, which is "the secure sharing of customer data with authorised third party providers, upon the customer's request". Our definition covers a broader set of data types (including non-consumer data) with our focus being on the reuse of de-identified data for academic research.

2 Romanko, O, O'Mahony, M. The Use of Online Job Sites for Measuring Skills and Labour Market Trends: A Review. ESCoE Technical Report No. 2022-19.

3 The Urban Big Data Centre currently provides job advertisement data from Adzuna, which was used in this study, to researchers in the UK: **https://www.ubdc.ac.uk/data-services/data-catalogue/employment-data/adzuna-data**

4 Brewer HR, Hirst Y, Chadeau-Hyam M, Johnson E, Sundar S, Flanagan JM. Association Between Purchase of Over-the-Counter Medications and Ovarian Cancer Diagnosis in the Cancer Loyalty Card Study (CLOCS): Observational Case-Control Study. JMIR Public Health Surveill. 2023 Jan 26;9:e41762.

5 Hirst Y, Stoffel ST, Brewer HR, Timotijevic L, Raats MM, Flanagan JM. Understanding Public Attitudes and Willingness to Share Commercial Data for Health Research: Survey Study in the United Kingdom. JMIR Public Health Surveill. 2023 Mar 23;9:e40814.

6 For example, Ada Lovelace Institute. (2022). Inform, educate, entertain… and recommend? Exploring the use and ethics of recommendation systems in public service media. Available at: **https://www.adalovelaceinstitute.org/report/informeducate-entertain-recommend**

7 **https://www.cdrc.ac.uk/research/retail/do-only-affluent-consumers-buy-green-labelled-products/**

8 Barkemeyer R, Young CW, Chintakayala PK, Owen A. 2023. Eco-labels, conspicuous conservation and moral licensing: An indirect behavioural rebound effect. Ecological Economics. 204 (Part A) **https://eprints.whiterose.ac.uk/192787/**

9 Examples taken from a longer list in 'Bit by Bit: Social Research in the Digital Age', by Matthew J. Salganik. 2017. Princeton University Press.

10 This includes ESRC investments such as Digital Economy Hubs (2014 – 2015), **HomeSense: digital sensors for social research** (2016 – 2019) at the National Centre for Research Methods, Big Data Network (2013 – 2018), Business & Local Government Data Centres (2014 – 2020), New and Emerging Forms of Data Policy Demonstrators (2017 – 2018), and funding for the **Social Data Science Lab** and the **Data Analytics and Society Centre for Doctoral Training**. The Strategic Priorities Fund has also built capacity in this area, including through the Global Open Finance Centre of Excellence (now the **Smart Data Foundry**).

11 See, for example, the **Alan Turing Institute**, **Oxford Internet Institute**, **Web Science Institute** at the University of Southampton, **Data Science Institute** at Imperial College London, and the **Office of National Statistics Data Science Campus**.

12 **digitalfootprintscommunity.org**

© Crown copyright 2023

This publication is licensed under the terms of the Open Government Licence v3.0
except where otherwise stated.
To view this licence, visit **nationalarchives.gov.uk/doc/ open-government-licence/version/3**

9718