UNIVERSITY OF LIVERPOOL

Newcastle University

Centre for Data

ESRC Smart Data Research
UK Strategic Advice Team

REPORT 6: KEY RECOMMENDATIONS AND SUMMARY OF ACTIVITIES FROM THE SDR UK STRATEGIC ADVICE TEAM

Rachel Franklin, Simeon Yates,
Elena Musi, Omar Guerrero, Seth Spielman,
Jeanette D'Arcy, Jessica Crosby

June 2025

DMSI
Digital Media and Society Institute

Centre for Urban & Regional Development Studies
CURDS

CONTENTS

EXECUTIVE SUMMARY

As the SDR UK Strategic Advice Team (SAT) initiative draws to a close, findings from across SAT activities have produced four key recommendations:

1. Create a centralised resource suitable for all career levels and disciplines, providing checklists for legal and ethical considerations, a data discovery portal and standardised documentation.
2. Build relationships for data access and smart data research communities, utilising soft power to broker access, ensuring availability of long-term funding, and facilitating good relationships between industry, government, and academia.
3. Expand upon existing resources to ensure sustainability of safe and secure research environments for smart data.
4. Promote and provide a range of skills, training and outreach, focused on creating clear guidance for upskilling in key areas of smart data research, as well as public awareness and education around the benefits of smart data research.

This report also includes a summary of SAT activities over the last 24 months, and a compendium of resources and contacts that the SAT have gathered in that time.

INTRODUCTION

The Strategic Advice Team: Aims and objectives

As a major investment in the research landscape of the UK, the ESRC Smart Data Research UK (SDR UK) programme (formerly Digital Footprints) has great potential for stimulating innovation and change in the use of smart data for research that has real social impact. The potential to improve people's lives through smart data-driven research in areas such as health, transport, finance, and the environment make the SDR UK's remit one of urgency in the current cost-of-living and climate crises.

The smart data landscape is rapidly transforming as new digital technologies and legislation are coming to fruition, such as the Data Use and Access Bill (at the committee stage at the House of Commons at the time of writing), as well as the Data Protection and Digital Information Bill, both of which make provisions about access to UK customer and business data for the regulation of information standards and disclosure of information. Similar legislation is in the works regarding the advancement of generative AI, a technology dependent on smart data, which has been addressed as part of the Labour government's industrial strategy (Labour, 2024). While it's still uncertain what actions the new government will take, their manifesto laid out intentions to build new data centres, as well as a National Data Library, to consolidate existing research programmes and support data-driven public services. As the programme develops, access to smart data will likely be a key challenge to tackle, especially where social media data is concerned. Ghermandi et al (2023) suggest that issues with accessibility, changes to terms of service, censorship and data manipulation will all pose threats to the sustainability of social media data, particularly in contexts where this data is being utilised for the social good (Ghermandi et al, 2023).

The Strategic Advice Team was proposed as an entity which would aid the SDR UK team in understanding the broader smart data landscape, as well as developing sufficient visibility, leadership, championship and 'buy-in' to the programme. These aims were addressed via a programme of work which took place from March 2023 to January 2025. We convened an experienced and well-connected Strategic Advice Team which developed and maintained a structured, inclusive, and organised mechanism for soliciting and synthesizing expert strategic advice for the SDR UK programme, as well as undertaking 'championing' activities to build interest in, and awareness of, programme goals and opportunities.

To meet the above aims, we delivered the following objectives:

- ***Connecting*** a currently fragmented smart data landscape through an international Advisory Group network that integrated expertise and viewpoints from across smart data communities
- ***Mapping*** the smart data landscape through a Delphi process of semi-structured interviews and questionnaire, anchored in Advisory Group expertise
- ***Amplifying*** the value of ESRC investments through knowledge sharing and networking among current and future smart data grantees, Advisory Group members, and the wider community
- ***Championing*** and highlighting ESRC data infrastructure investments, including smart data Accelerator and Data Services activities, via the visibility of the Strategic Advice Team and Advisory Group, workshop and Data Dive activities, written reports and other types of outputs.

SUMMARY OF SAT ACTIVITIES

1 In-Person and Online SAT Workshops

The SDR UK SAT workshops were set out to convene expertise from stakeholders in smart data research across academia, policy, and business. Co-ordinated by The Collective, these interactive sessions worked to establish new networking connections, share interdisciplinary experience, and establish current issues and mandates across the UK's smart data landscape.

- **SAT Launch (January 2023)**

  This workshop, held at the Alan Turing Institute in London, assembled experts from industry, government and academia to highlight the value and promise of smart data, demonstrate existing capacity, and indicate future directions.

- **Workshop 1: Research and Data Priorities (March 2023)**

  This workshop was intended as a broader community introduction to the ESRC Smart Data Research UK programme (then called Digital Footprints) and to establish vital relationships and networks needed to carry out the project's work.

- **Workshop 2: Identifying Researcher Needs (April 2023)**

  This workshop established definitions for the programme's research pillars and proposed data centres that would define the Smart Data Research UK programme and subsequent funding calls.

- **Workshop 3: Learning from International Best Practice (May 2023)**

  This workshop focused on deriving insights on best practice in smart data research outside of the UK, with the aim of establishing relationships and fostering knowledge exchange with our international colleagues. Speakers for this event included: Margaret Levenstein, Director of ICPSR; Deborah Wiltshire, Leader of Secure Data Centre at GESIS-Institute for the Social Sciences.

- **Workshop 4: Research Infrastructure (June 2023)**

  This workshop was focused on computational frameworks for smart data already in operation across the UK, with the intention of learning from best practice and successful business operation. Speakers for this event included: Sebastian Bacon from Open Safely; Rob Baxter and Fergus McDonald from DARE UK; Kieran Jarrett from the ESRC; Nick Bailey, Director of UBDC; James Cheshire from the CDRC.

- **Workshop 5: Meet the Team! (July 2023)**

  This workshop offered an opportunity for the research community to put questions to the SDR UK and SAT teams. This allowed for further detailing of the programme's funding calls and future opportunities in the smart data space. Speakers for this event included: Joe Cuddeford, Director of Smart Data Research UK, ESRC UKRI; Bruce Jackson, Head of Workstreams and Operations, ESRC UKRI; Deborah Kroll, Head of Strategy and Partnerships, ESRC UKRI; Matthew Croston, Senior Policy Manager, ESRC UKRI; Prof. Rachel Franklin, Programme Lead, Newcastle University; Prof. Simeon Yates, Programme Lead, University of Liverpool.

- **Workshop 6: Smart Data Research, Policy, and Practice (December 2023)**

  This workshop focused on hearing findings from the Open Innovation Team's (OIT) rapid review, exploring policy demand and opportunities relating to smart data and public engagement. Speakers for this event included: Jack Hampton from the OIT; Daniel Robinson, Strategic Lead for Policy Partnerships and Impact, ESRC; Josh Hellings and Finn McEvoy from the Economics Observatory; Dr Jonathan Higham, Lecturer at the University of Liverpool.

- **Workshop 7: Data Access and Permissions (May 2024)**

  This workshop was intended to explore issues surrounding data access and permissions, particularly regarding convergences and differences in access requirements between stakeholders in academia, the commercial sector and in government. Speakers for this event included: Casey Weston from LinkedIn's Data for Impact Program; Andrew Gruen, consultant and technology executive from The Future of Privacy Forum; Prof. Gordon Blair, Head of Environmental Digital Strategy at UKCEH.

- **Workshop 8: Ethics and Legalities (October 2024)**

  This final workshop was aimed at offering examples of best practice and how ethical standards are maintained for research in HE settings. This workshop considered different changes and frictions in the smart data landscape that could pose a problem in legal and ethical terms, and the necessity of developing research frameworks that can adapt to evolving digital challenges. Speakers for this event included: Professor Suchith Anand, cofounder of the Ethical Data Initiative; Dr Nina de Cara from University of Bristol's Data Hazards project; Katie Burns & Emilie Olufsen from the PSC.

2 Interviews

Twenty-three expert interviews were carried out as part of SAT activities (16 with members of the AG and other academic experts; seven with data owners). These took place from June–August 2023. Coding commenced in August, and the analysis and write-up of the interviews was concluded by October 2023. Headline recommendations for the SDR UK programme include:

- **Infrastructure**
  - Account for interoperability, not only in the initial design of the programme, but thinking of how this will enable sustainability of the programme over time.
  - Approach the building of infrastructure not just with a focus on how data is stored, linked and described, but do so in parallel with a focus on how data will be - and actually is -used by researchers, and how researchers might want to work with data in the future.
  - Create a central 'hub' as a 'first stop' for smart data researchers at all stages of careers, which should:
    - Provide a trusted source of information and guidance on licensing and legal standards.
    - Act as a community of practice and communication.
    - Help produce standard guidance for metadata quality and description.
    - Lead the way in ethical governance, including the production and collation of ethics guidelines and the latest research.

- **Skills, Training and Outreach**
  - Create a clear outreach programme aimed not just at industry and academia, but towards public awareness and education around the benefits of smart data research.
  - Focus on three key areas of training and skills:
  - Encourage computational/data science skills for social science researchers
  - Encourage social science (methodological/ethics) skills for data science researchers
  - Promote Sci-Comms and outreach skills
  - Lead the way in training and expertise relating to AI and Smart Data research.

- **Building Relationships**
  - Focus on building communities of practice and knowledge exchange; this could be done through a forum or peer-to-peer support structures.
  - Focus on creating and developing long-term sustainable Data Sharing Agreements, as well as quality standards and infrastructure that will create sustainable resources.
  - Create and fund 'liaison' roles designed to bridge the gap between industry and academia.
  - Closely consider ethical governance of relationships between industry and government.

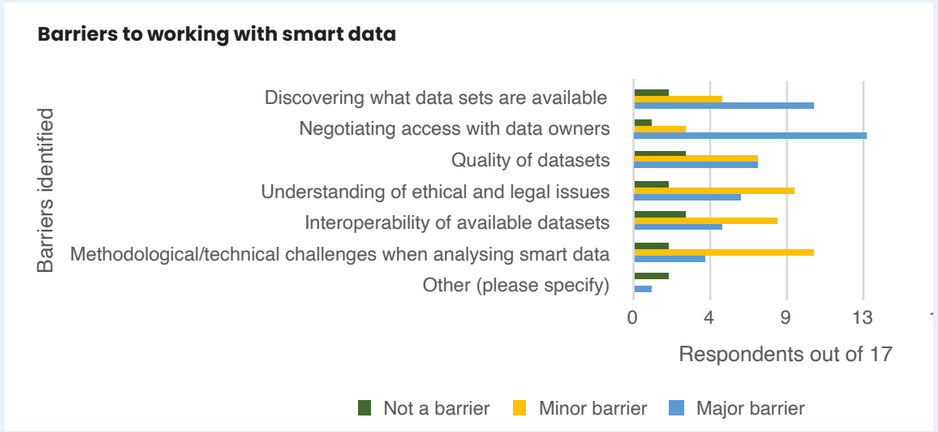
- **Legalities, Licensing and Ethics**
  - Lead the way in focusing on ethical governance, including the production and collation of ethics guidelines and the latest research in this area.
  - Lead the way in producing and collating standardised documentation relating to legalities and ethics in the smart data field.
  - Be mindful of legal precedents taking priority over ethical considerations.

- **Risks and Security**
  - Pursue public policy research.
  - Pursue a tiered approach to secure access, and/or create pathways to 'researcher passports'.

3 Survey

To review and consolidate emergent themes identified during expert interviews, and to provide a baseline in terms of smart data research undertaken in the UK, an online survey was created using Qualtrics and disseminated through research networks, mailing lists, and over social media. After proofing, the survey was published in March of 2024 and closed in August of the same year. The survey attracted a total of 140 respondents, with 61 completing the survey in full. Headline findings include:

- 22% of respondents **did not** currently work with smart data:
  - In response to a question which offered statements to respondents and asked them to tick all that were relevant to them, 41% of those not currently working with smart data said they are unsure whether they would like to work with smart data and would like more information; 35% indicated that they would like to work with smart data but don't have the necessary skills; 35% indicated that they are very unlikely to work with smart data in the future; 12% said they intend to work with smart data in the next year.
  - For those not currently working with smart data, 'negotiating access with data owners' was the most commonly selected 'major barrier', followed by 'discovering what data sets are available':
- Both those who did and those who did not currently work with smart data identified 'negotiating access with data owners' as a **major barrier** to accessing or working with smart data



- Those working with smart data identified 'Rapid development of technologies' as the most common first choice when asked about **key barriers** to accessing and using smart data in terms of **risk and security**
- Most of those who indicated they did currently work with smart data had been working with it for **up to five years**, and mostly in the area of **research and/or teaching in academia**.
- 86% of respondents **somewhat or strongly agreed** with the statement 'Interview participants recommended that the ESRC create a resource centre (i.e., a 'first stop' that would provide guidance on, e.g., ethics, legalities and how to approach the use of smart data').
- In terms of skills gaps, 'responsible research' was identified as the **most important skill area** for researchers, PhD students, and industry/government data providers. It was identified as second most important for those running data services, after 'data engineering'.
- 52% of respondents agreed that **the likely future direction** for smart data in the UK is 'moving towards more open models of access', while 20% felt it will be 'moving towards more commercial models of access'.
- The survey had a completion rate of 43% (61 out of total 140)
- Respondents were **mostly male** (55%), mostly **41-50 years** of age (31%), from a **white British** background (82%)

4 Data Dives

The Data Dives carried out as part of the SAT's remit were intended to both foster knowledge exchange and build a smart data research community and offer insights into practical experiences working with a range of datasets, expertise, and partnerships.

- **Data Dive 1:** Consumer Data Research Centre – University of Liverpool (April 2024)
- **Data Dive 2:** Urban Big Data Centre – University of Glasgow (June 2024)
- **Data Dive 3:** Institute for Manufacturing – University of Cambridge (October 2024)
- **Data Dive 4:** Meta/SOMAR – University of Liverpool (January 2025)

Key takeaways:

- Researchers need clear information on how datasets are collected/compiled/cleaned (i.e., what processes led to the creation of the dataset in its current form). This was frequently raised as important, and often lacking or inconsistent across datasets.
- Complex datasets can be challenging for ECRs and those with less experience. That is, the ability to identify gaps/discrepancies in datasets and/or the use of specific coding skills such as using Python or R both present challenges to researchers who may be lacking in formal training in programming and working with quantitative data.
- Feedback on the data dive methodology was largely very positive, and participants felt that the opportunity for networking, collaboration and learning from peers was unusual and appreciated.

Main recommendations:

- The continued building of data-user communities where members could share experiences and issues and create collaborative networks.
- Providing mechanisms for easing data sharing, and incentives to data holders to share data (e.g. benefits to the business for sharing data)

5 Conferences, workshops, and community engagement

- **Lego workshops:** As part of the SAT's outreach and championing activities, we undertook two Lego workshops utilising an adapted Lego Serious Play methodology, developed by colleagues at the University of Liverpool. These were conducted as part of national and international conferences, including the *Digital Footprints Conference* at Bristol University (May 2024) and the *Association of Internet Researcher's (AoIR) International Conference* (October-November 2024).
- **Data for Policy conference:** At this conference, which was organised around the theme 'Decoding the future: Trustworthy governance with AI?', Professor Rachel Franklin and Dr Omar Guerrero organised a panel discussion titled *Big Tech or Government: Will the Production of Privately Driven Smart Data Crowd Out Publicly Funded Data-Generation Capabilities?* Panellists included representatives from GEOLYTIX, University of Bristol, Financial Times, LSE, and London, GLA. The panel also featured opening remarks from Sam McGregor, Deputy Director of the ESRC Smart Data Research UK (SDR UK) Programme, and Mark Birkin, Director of the Urban Analytics Programme at the Alan Turing Institute (July 2024).
- Dr Jessica Crosby and Dr Jeanette D'Arcy attended an **online workshop organised by SDR UK with Google Health UK**, titled *Using internet search data to improve health research*. This workshop explored new data portability products that make it easier for users to consent to share data with authorised third parties, and the impact of these changes. (July 2024)
- Professor Simeon Yates and Dr Jeanette D'Arcy attended the **British Academy SHAPE workshop** in Sheffield, focused on the understanding the challenges and solutions related specifically to online data access from digital media companies (Meta, YouTube, etc) for SHAPE researchers, who are to deliver societally important insights that have the potential to inform crucial policymaking and regulation (Oct 2023).
- Dr Jeanette D'Arcy and Dr Jessica Crosby attended the **Turing Novel Data Linkages for Health & Wellbeing Interest Group workshop**, focused on how digital footprints data can be used for public benefit (Nov 2023).
- Dr Jeanette D'Arcy attended the **DARE UK Scientific Use Cases workshop**, a collaborative workshop designed to explore how to better understand the research potential that can be unlocked through improved data connectivity, develop use cases for joining up different types of sensitive and non-sensitive data at scale, and identify the most exciting opportunities for enhanced data connectivity. This included contribution to their final report (Feb 2024).
- Four conference sessions on novel data and methods ("The Interesting Data and Methods" sessions) were organised for the **Annual Meeting of the American Association of Geographers (AAG)** in Detroit and included presentations on a range of smart data types and applications, including mobile phones, real estate advertisements, street view imagery, mortgage data, and social media (March 2025).

KEY RECOMMENDATIONS

1 Create a centralised resource

Throughout SAT activities, researchers identified the need for a 'one-stop' resource to find guidance and signposting. This central resource should include:

1. An easy to navigate data discovery portal, including signposting to training and upskilling modules.
2. A 'checklist' document to identify what should be considered for all smart data research projects in terms of legalities, ethics, ownership, etc., as a step towards standardising legal requirements and ethical approaches to smart data.
3. Resources that cater for those at the beginning of their career in smart data as well as more established researchers.

Key challenges identified across SAT activities centred around difficulties in discovering available datasets and knowing how to approach this data in a legally compliant and ethical manner. Input from workshops and interviews stressed the difficulties involved in trying to leverage support and training for researchers both at the start of their smart data careers, as well as those who had previous expertise in working with smart data, but ongoing demands for training and upskilling. Resource allocation should therefore account for such upskilling, both for researchers and those running the services.

Amongst survey respondents, responsible research was identified as the most important skill gap to address amongst doctoral students, researchers, industry and government providers. For those running the data services, responsible research was identified just below data engineering as a skill gap that needed addressing.

There is a wide range of potential legal considerations for researchers to address when engaging with smart data, as well as a lack of available ethical guidance in this space, so standardisation of smart data resources was identified as a key priority, particularly as many participants discussed concerns over legal precedents for smart data use taking priority over ethical considerations. Eighty-six percent of survey respondents strongly agreed that SDR UK should create a resource centre to help centralise, and therefore work towards standardisation of, guidelines for engaging with smart data for research purposes. Participants in all activities frequently recommended that the programme avoid 'reinventing the wheel', taking advantage of existing resources to help build beneficial social norms, creating and adhering to agreed standards. This would avoid creating unnecessary duplication (or abstraction) of existing technical and social practices.

Participants ranked potential resources that could be offered and identified a data discovery portal as the most important, along with access and/or signposting to training. It was stressed that working towards a set of agreed standards for smart data research would help to develop beneficial community norms and best practice and set rigorous examples for future smart data infrastructure.

The programme can therefore lead the way in relation to:

- Ethical governance, including the production and collation of ethics guidelines and the latest research in this area.
- Production and collation of standardised documentation relating to legalities and ethics in the smart data field.
- Equitable access, metadata management, and governance that ensures outputs are reproduceable and transparent.
- Creating and developing long-term sustainable Data Sharing Agreements, as well as quality standards and infrastructure that will create sustainable resources.

2 Build relationships for data access and smart data research communities

Across SAT activities, researchers were keen to discuss opportunities for building better relationships between academia and industry, data owners, and government, as well as creating a robust researcher community in the smart data space.

There was broad support for ESRC to leverage its negotiating power to make access to data as open as possible, for example by applying soft pressure to ease issues around licensing and proprietary data. Researchers emphasised the importance of building relationships with data owners, to foster access and standardise processes and legalities across different data services. Potentially risk-averse data providers were often discussed as a challenge to accessing the data required for research, and therefore the need for the programme to focus on building trusted relationships with industry and government. This will take time and require sustained, long-term planning and funding. One possible solution suggested by experts and reiterated by survey respondents was to create and fund 'liaison' roles designed to bridge the gap between industry and academia.

When promoting relationship-building activities, the programme should emphasise their mutually beneficial nature to the data providers themselves. There is opportunity for developing iterative processes to help providers understand more about what data they have, its value, and how to realise its potential in commercial terms and regarding its use for social good. However, it should be noted that researchers also raised the importance of considering issues of ethical governance in relationships between academia and industry. Balancing commercial incentives with ethical imperatives will be key in building relationships that encourage responsible practices.

Peer-to-peer support networks are currently rare in the smart data space, especially cross-disciplinary spaces like those the programme wishes to encourage. There is an opportunity to fill a gap in existing smart data resources around communities of practice and information exchange. This could be addressed through, for example, provision of a forum or peer-to peer support structures.

The programme can therefore lead the way in relation to:

- Utilising soft power to broker access to datasets for UK smart data research.
- Ensuring long-term funding is available with an understanding that relationships take time to establish and maintain.
- Forging trusted relationships with data owners based on mutual benefits and clear governance.
- Facilitating good relationships between UK academia, government and industry.
- Creating a forum for UK smart data researchers to come together and engage in knowledge exchange and peer support.

3 Build upon existing resources to ensure safe, secure environments for smart data

Most participants felt that, while safety and security must be key considerations in the smart data space, there are already procedures, policies and technologies that mean this is not a major challenge. They again emphasised that the SDR UK programme should not attempt to 'reinvent the wheel', instead making use of existing security measures and tools. For example, many suggested the use of cloud-based resources already available, which were considered useful as augmentation and updating would be easy, quick and relatively inexpensive as new innovations in technologies are introduced.

Security and safety were raised as key issues for smart data researchers, particularly when dealing with the possibility of reidentification and/or data leakage in sensitive data sets, as well as the risk that over-zealous security measures may result in siloing and make access even more difficult. This suggests a balanced approach would be welcome, w. Researchers were keen to see these issues addressed through investment that also takes a balanced approach, including funding for the necessary computing power and technologies to maintain safe and secure research environments, as well as long-term, sustainable funding for projects to allow for relationship building, community and public outreach, and ethical expertise. Suggestions in this area included following the example of existing Trusted Research Environments (TREs); adopting a tiered approach to access, allowing the most sensitive data to remain secure, while less sensitive datasets are more open; and the development of procedures such as a 'researcher passport', which would allow those holding such accreditation quicker and easier access to SDR data sets.

The programme can therefore lead the way in relation to:

- Adopting a balanced approach which provides open access as far as possible, while keeping safety and security at the forefront of developments.
- Developing and standardising procedures and policies to encourage innovative work in secure research environments, including tiered approaches to secure access, and/or pathways to 'researcher passports'.
- Using existing resources to ensure safe and secure environments for UK smart data research.
- Accounting for interoperability, especially how this will enable sustainability of the programme over time.

4 Promote and provide a range of skills, training, and outreach

Interdisciplinary research was flagged by many as a key area for development in the smart data space, with a concurrent concern over friction between qualitative and quantitative skills, methodologies and ways of working. Participants spoke about the importance of investing in people as well as technologies, pointing out the need for diversity in skills, training and outreach, especially in interdisciplinary areas. For example, knowledge and skills in data science were felt to be underdeveloped for those in the social sciences, and *vice versa*. It is important to develop productive interdisciplinary working between social and data/computer sciences, particularly in light of the rise of AI technologies. The programme should be flexible in its design to allow for intersections in research between AI and smart data, preparing for the development of materials and training as this field advances.

In line with discussions around a lack of a 'universal' infrastructure for smart data, it was reiterated by participants that the fast-paced nature of advancing technologies in this field means there is no 'one-size-fits-all'. For example, data formatting and coding languages will be varied and change over time. The need for flexibility in design was considered very important, to accommodate differences that exist between data sets but also the changes that will need to be made to adapt to a shifting smart data terrain.

When speaking to data owners, it was clear that they consider outreach and Sci-Comms skills to be paramount, citing the need to 'tell the story' of smart data more effectively, both in the public sphere and throughout industry. This was felt to be crucial in several contexts, as it would help to reassure the public that their data is being used responsibly, help those in industry see the value of allowing access to their data for research, and help to make research itself more accessible and transparent.

The programme can therefore lead the way in relation to:

- Creating a clear outreach programme aimed not just at industry and academia, but public awareness and education around the benefits of smart data research.
- Creating clear guidance, signposting and potentially provision for upskilling of researchers.
- Training and expertise relating to AI and smart data research:
    - Computational/data science skills for social science researchers;
    - Social science (methodological/ethics) skills for data science researchers;
    - Sci-Comms and outreach skills.

5 Conclusion

The work carried out over the two years of the Strategic Advice Team's activities has provided insights that should allow the SDR UK programme to shape itself effectively in an ever-changing smart data landscape. The recommendations derived from SAT activities should prove invaluable as the programme moves into its next phases. For instance, the creation of a centralised resource with standardised procedures and documentation will aid researchers in navigating future policy and legislative changes relating to smart data usage such as bills currently moving through parliament, like the Data Use and Access Bill, and the Data Protection and Digital Information Bill. As the current government has signalled a desire to engage with evidence-based approaches, with effective use of data at their core (Labour, 2024), the programme has opportunity to build and develop upon relationships between academia, government and industry, using insights provided by SAT activities. This will be especially important as a foundation for developing expertise in areas identified by SAT activities as being particularly challenging, such as the rapid development of AI technologies, and the increasingly constricted access to social media datasets (Ghermandi, 2023).

These kinds of agile, flexible approaches recommended by the SAT's gathering of expert opinions will require the SDR programme to carefully consider how best to allocate funding to create long-term and sustainable infrastructure. This should include approaches that invest not only in technologies, but in people, given that a range of interdisciplinary skills will be required, both in terms of running and using data services, and in terms of creating the sci-comms outreach that experts have identified as vital for building public trust in smart data use. To this end, the programme should focus on offering sources of funding which are longer term, allowing for the time-consuming business of building and maintaining relationships as well as upskilling and outreach.

Programme activities have been well-received, with feedback from Data Dives being particularly positive. Data Dive participants emphasised that the opportunity to collaborate with others in the smart data space and brainstorm ideas and opportunities with peers was extremely valuable. This suggests that SDR UK might want to carry out similar events in the future, as these proved useful in facilitating the building of long-term research relationships.

6 Compendium A-Z

This section provides information about projects, institutions and individuals, either taking part in or encountered during SAT activities, that the ESRC may wish to be aware of as the SDR UK programme develops.

AoIR Ethics Guidance

AoIR have ethics guidance documents which could be useful for researchers, the latest and previous iterations can be found here: https://aoir.org/ethics/

AI for Social Good

The AI For Social Good (AI4SG) project, which is sponsored by the National Science Foundation, engages students from all disciplines to propose or develop AI powered solutions that address social issues. https://www.aiforsocialgood.org/

Ethical Data Initiative

The Ethical Data Initiative (EDI) seeks to tackle significant issues such as data literacy, transparency, and trust-building within data ecosystems. By fostering open dialogue and collaboration, the initiative sets out to shape ethical data practices and pave the way for a more transparent and trustworthy data landscape.

OpenSAFELY

OpenSAFELY is an open-source secure platform for analysis of electronic health records data. All platform activity is publicly logged and code for data management and analysis is shared, under open licenses and by default, for scientific review and efficient re-use. It can be used for making a Trusted Research Environment (TRE) or deployed as a privacy-enhancing layer on an existing secure database or TRE.

DARE UK

Data and Analytics Research Environments UK (DARE UK) aims to establish a network of Trusted Research Environments (TREs) for approved researchers to access sensitive data to advance research for the public good. DARE UK's latest report examined how linked sensitive data in the UK could be brought together to enable research that benefits the public: https://zenodo.org/records/14025303

UBDC

The Urban Big Data Centre (UBDC) acts as a research hub and data service for the UK, championing the use of smart data to improve urban life and inform policymaking.

UK Centre for Ecology and Hydrology

The UK Centre for Ecology & Hydrology (EKCEH) is an independent research institute that hosts long-term environmental data used to inform policymaking and drive research and innovation to tackle urgent environmental challenges.

Institute for Manufacturing/Supply Chain Artificial Intelligence Lab

The Supply Chain Artificial Intelligence Lab (SCAIL) at Cambridge University's Institute for Manufacturing (IfM) studies complexity science, emergent AI technologies and agent-based computing techniques to create methods to discover hidden patterns in data and provide insights for improving supply chain operations.

Public Service Consultants (PSC)

The Public Service Consultants (PSC) work with the UK public sector to improve public services.

CDRC

The Consumer Data Research Centre (CDRC) uses consumer data for research to provide insight into a broad range of societal issues and economic challenges. They work with industry and the public sector to provide data access and expertise to better understand consumer behaviour, as well as offering training and education. The new Geographic Data Service will integrate a wide range of data sources to generate new insights into equitable and sustainable growth, bringing focus to geographic disparities, barriers to opportunity and the circumstances of vulnerable populations.

Data Hazards project

The Data Hazards project, based at Bristol University, aims to mitigate downstream harms of data science tools and research by prompting researchers and developers to reflect on potential ethical issues and mitigate them where possible. The project provides resources to bring in a wide variety of views on these issues, and help researchers reflect about the ethics of their work. https://datahazards.com

Open Innovation Team (OIT)

The Open Innovation Team (OIT) works across UK government, reviewing and analysing evidence, developing policy and evaluating impact.

The Economics Observatory

Based in Bristol, the Economics Observatory (ECO) aims to

bridge gaps between policy, research and the public. They analyse data and publish articles, videos and charts to help policymakers and the public better understand social and economic challenges.

ICPSR, University of Michigan, US

The Inter-University Consortium for Political and Social Research (ICPSR) is an international consortium of academic institutions and research organizations. They specialise in training and leadership in data access, curation, and methodologies for the social science research community, as well as maintaining a data archive of social and behavioural sciences research.

Digital Footprints Lab

The Digital Footprints Lab, based at the University of Bristol's Population Health Sciences centre, focuses on transaction data (e.g., banking and loyalty cards) to improve population health. They run the annual Digital Footprints conference.

GESIS

The GESIS Leibniz Institute for the Social Sciences provides fundamental research-based services for the social sciences focused on digital behavioural data, including Research Data Centres (RDCs) and the Integrated Survey and Data Infrastructure (IEDI).

REFERENCES

DARE UK (Data and Analytics Research Environments UK). (2024). Scientific use cases for cross-domain sensitive data research in the UK. Zenodo. https://doi.org/10.5281/zenodo.14025303

Ghermandi, A., Langemeyer, J., Van Berkel, D., Calcagni, F., Depietri, Y., Egarter Vigl, L.E., Fox, N., Havinga, I., Jäger, H., Kaiser, N., Karasov, O., McPhearson, T., Podschun, S., Ruiz-Frau, A. Sinclair, M., Venohr, M. and Wood, S.A. (2023) Social media data for environmental sustainability: A critical review of opportunities, threats, and ethical use, *One Earth, 6*(3), 236-250. https://doi.org/10.1016/j.oneear.2023.02.008

House of Lords. (2024). *Data Protection and Digital Information Bill*. https://bills.parliament.uk/bills/3430

House of Lords. (2024). *Data (Use and access) Bill*. https://bills.parliament.uk/bills/3825

Labour. (2024). *Change: Labour party manifesto 2024. https://labour.org.uk/change/*

DMSI
Digital Media and Society Institute

Department of Communication & Media
University of Liverpool
School of the Arts
19 Abercromby Square
Liverpool
L69 7ZG

Newcastle University
Centre for Data

Centre for Urban & Regional Development Studies
CURDS

THE ORIGINAL REDBRICK