

**ESRC Smart Data Research  
UK Strategic Advice Team**

**REPORT 2:  
RESEARCH  
INFRASTRUCTURE  
& LEARNING FROM  
INTERNATIONAL  
BEST PRACTICE**

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

**July 2023**

**DMSI**

Digital Media and Society Institute

Centre for  
Urban & Regional  
Development Studies

**CURDS**

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Executive Summary and Headline Recommendations</b>	<b>5</b>
<b>3</b>	<b>Summary: Learning from International Best Practice Workshop</b>	<b>6</b>
3.1	Executive highlights	6
3.2	Overview of event	6
3.3	Emerging Themes	6
3.3.1	Legal infrastructure	6
3.3.2	Technical and human infrastructure	7
3.3.3	Access	7
3.3.4	Quality and trust	7
3.3.5	Best practice examples	7
3.4	Outstanding issues and open questions	7
3.4.1	Standardisation	7
3.4.2	Linkage and interoperability	8
<b>4</b>	<b>Summary: Research Infrastructure workshop</b>	<b>9</b>
4.1	Executive highlights	9
4.2	Overview of Event	9
4.3	Emerging Themes	9
4.3.1	Flexible design	9
4.3.2	Interoperability and linkage	9
4.4	Outstanding Issues	10
4.4.1	Skill Gaps	10
4.4.2	Inclusivity and accessibility	10
<b>5</b>	<b>Summary: Advisory Group Meeting</b>	<b>11</b>
5.1	Executive highlights	11
5.2	Overview of event	11
5.3	Emerging Themes	11
5.3.1	Technological infrastructure	11
5.3.2	Problems may not be easily solved through technological solutions	11
5.3.3	Legalities	11
5.3.4	Qualitative research	11
5.3.5	Best practice exemplars	11
5.4	Outstanding issues and open questions	13
5.4.1	Resource allocation	13
5.4.2	Regulations and governance	13

# INTRODUCTION

Please note: This preliminary report was produced in July 2023. Since the writing of this document, the name of the programme and its attendant terminology has changed from 'Digital Footprints' (DF) to 'Smart Data' (SD).

This report was commissioned by Smart Data Research UK – an Economic and Social Research Council (ESRC) data infrastructure programme. The purpose of this report is to provide context, input, and early-stage recommendations regarding the overall structure of the Digital Footprints Phase 2 Investment. It was prepared by Jeanette D'Arcy at the University of Liverpool and Jessica Crosby at Newcastle University, members of the Strategic Advice Team funded by Smart Data Research UK to give independent strategic advice between 2022 and 2024.

## Disclaimer

The findings and conclusions presented in this report are solely those of the researchers and do not necessarily reflect the views of Smart Data Research UK or the Economic and Social Research Council.

This report offers an overview followed by three separate summaries of the recent events in the Digital Footprints series: two workshops, Learning from International Best Practice and Research Infrastructure, and the June meeting of the Advisory Group, at which the two group discussion topics were User Needs and Data Owner Needs.

# EXECUTIVE SUMMARY AND HEADLINE RECOMMENDATIONS

Given the recurrent themes and issues that have arisen from workshops and advisory group discussions, the SAT:

1. concurs with ESRC's existing goal to set an example and help build beneficial social norms by creating and adhering to agreed standards (e.g., setting expectations for inclusive and accessible data processing; instigating governance that ensures outputs are reproducible and transparent)
2. sees broad support for ESRC leveraging its negotiating power to make access to data as open as possible (i.e., applying soft pressure to ease issues around licensing and proprietary data)
3. recommends an iterative, flexible, and diverse approach to building up the data services that allows for ongoing evolution across the DFD landscape
4. proposes a long-term view on the allocation of resources for the Digital Footprints programme that considers not just the initial creation of the data services, but requirements for their ongoing maintenance and future sustainability. Resource allocation should also take into account upskilling of both researchers and those running the services.

Where prospective data service funding applicants are concerned, expressions of interest and bids should ideally demonstrate:

- understanding of existing data infrastructure landscape and how their own service's approach will link up with other services and infrastructures to maximise access
- a proposed set of social/scientific standards and how their service will develop relationships and foster beneficial social norms in line with these standards
- existing relationships applicants already have that would be beneficial in fostering access and standardising practices. As part of this, it is recommended that applicants consider how services will set expectations of data owners/custodians that data will be available for certain mechanisms, rather than to certain people or groups
- how applicants will provide a diversity of resources to 'bake-in' flexible design
- examples of the different skills and support the service will offer (e.g., discoverability, coding). A big question: how will services ensure that they cater for both researchers at the cutting edge of data research and those who may be new to DFD research?

A key priority emerging across the workshop on Research Infrastructure was on **building relationships with data owners, to foster access and standardise processes and legalities** across different data services. Participants were unanimous in their assessment that there is no 'universally agreed approach' to DFD processing, which means that data licensing conditions often become the 'de-facto' governing

factor amongst data sets, and thus the primary driver in how researchers currently share data. It is crucial, therefore, that data owners and researchers can work towards a set of agreed standards, to build towards beneficial social norms, and thus set rigorous examples for future DF infrastructure.

Recommendations from participants **emphasised adherence to existing infrastructure in the DFD space**, to avoid creating unnecessary duplication (or abstraction) of existing technical and social practices. As a current challenge is a lack of standardized infrastructure, it was argued that developing new social standards or licensing rules could be counterproductive, when what the DFD landscape really needs is more synergy between centres/platforms. Many participants recommended following the example of existing TREs, stressing that an existing set of FAIR principles already exist around scientific data management in these environments. That data should be **findable, accessible, interoperable, and reproducible** (FAIR) are core tenets of good science. It was strongly emphasised that these principles should continue to be operationalised for the Digital Footprints programme.

In line with discussions around a lack of a 'universal' infrastructure for DFD, it was reiterated by participants that there is no 'one-size-fits-all' approach for data services, particularly in aspects like data formatting and coding languages used. The need for **flexibility** in design was thus emphasised, not only to accommodate differences that exist between data sets, but also in relation to the continuing evolution of the DFD landscape, and the changes that will need to be made to adapt to a shifting terrain. It was agreed that consideration of different potential mechanisms for data generation was important, and that infrastructure should be made adaptable to these differences.

Participants emphasised a need for **skills**, both in terms of recruitment for those running the services and the researchers using them. Points raised include upskilling of researchers (particularly those working at different levels of expertise), the difference in methodological approaches (for example, those working from a qualitative perspective, or quantitatively approach) and how the data services centres will be able to meet a variety of needs and demands.

Also established across all workshops was the importance of **resource allocation** for maintenance and sustainability, with participants flagging that the rapid pace of change would require funding and investment capable of keeping up with changes in the digital footprints landscape.

# SUMMARY: LEARNING FROM INTERNATIONAL BEST PRACTICE WORKSHOP

Importance of:

- finding balance between secure research environments and overly restrictive/risk-averse approaches that then become a barrier to access
- legal documentation (e.g., data agreements) to take into account the complexities of international research
- long-term perspective in terms of both technological infrastructure (power, maintenance, sustainability) and human infrastructure (building and developing relationships between data providers and data services, training researchers to use data sets effectively and securely)
- accreditation as a way of indicating trustworthiness

## Overview of event

The 'Learning from International Best Practice' workshop took place on 18th May 2023. 60 people registered for the event, with around 30 in attendance on the day. The three-hour workshop was facilitated by colleagues from The Collective and participants were from a range of academic disciplines. The workshop hosted short presentations by Margaret Levenstein, Director of the US-based Inter-University Consortium for Political and Social Research (ICPSR) and Deborah Wiltshire, leader of the Secure Data Centre at GESIS Institute for the Social Sciences in Germany. Speakers were asked to provide a 3–5-minute talk that would focus on their experience of running a data centre and working in this area. Speakers were given several prompts: In your experience, what are the key technical, organisational, and legislative aspects that make a centre successful? What are your key takeaways or 'top tips'? What are some of the things that work well, what are the central challenges? What would you change if you were designing a centre afresh?

**Deborah Wiltshire** provided a short, pre-recorded video and Margaret Levenstein presented live. The short presentations were each followed by breakout discussions between workshop participants in small groups, who were then asked to summarise in plenary roundups with the whole cohort. Presenters remained in the discussions after their talk to contribute and answer questions.

Key points from Deborah Wiltshire's presentation included thinking about how to strike a balance between the necessary restrictions on researchers and providing clear communication on why such restrictions are necessary, and the difficulty of scaling up services as demand grows. The latter, she suggested, requires keeping a close record of demand for data so that it is possible to anticipate busy periods, as well as adequate resourcing of support services.

**Margaret Levenstein's** presentation pointed out that their centre was designed to host and disseminate survey data rather than digital footprints data, but that their operations have expanded to include links to other kinds of data such as

administrative data and biological data, creating legal, social, cultural, and technological challenges. She emphasised the importance of the need to use what has already been created, as in the current climate different communities are tending to make their own standards (for metadata, for example) and this makes integration in ways that are scientifically valid difficult.

She also suggested that challenges for all those working in this area will increase as we use AI more, as this erases provenance so users will have even less information about where data comes from and how it was put together in its current form.

The presentations focused on how technological solutions are the beginning but not the end of the answer, pointing out that the challenges we care about in terms of data access are in questions of legal and social community norms. As potential solutions, Margaret suggested the need to build and reinforce social norms around good practice to develop an effective ecosystem in the digital footprints space, using legal and (self-)enforcing agreements to provide incentives to respect the protection and safety of sensitive data, with researchers losing future access if they do not comply.

In particular, she suggested that the difficulty of enforcing agreements across national boundaries requires self-enforcing ones created by community incentives and consequences. Such solutions, she pointed out, require that we train researchers in confidentiality protection and how to respect the privacy of those represented in the data used by researchers, and in how to share results in transparent ways. The ICPSR is developing a 'researcher passport' (as is GESIS), based on researchers' experience and responsibility in protecting sensitive data, so different organisations will be able to validate researchers and indicate trustworthiness.

Margaret also spoke about differences between the US context and elsewhere; how in the US there are some spheres with strict rules but unlike Europe there are few legal restrictions on, e.g., the use of social media data for research. However, there are increasing restrictions on researcher access from private companies. She suggested there is a case for resisting private restrictions that undermine scientific research, as researchers should not be bound by rules which do not protect confidentiality of subjects' data but only the property rights of companies over content.

## Emerging Themes

The following themes were prominent in the discussions with participants:

### Legal infrastructure

A key theme that emerged from participant discussions in our last report was ethics and legalities. Issues of legalities also formed a prominent theme in the Learning from International

Best Practice workshop, with a focus on the challenges caused by using and linking data sets both across and within national boundaries. Some key points raised in discussions:

- Challenge of variation across nations – US does not have a consumer privacy law, e.g.
- Lots of research is multi-national, with the need to be aware of the impact of local legislation on access provision agreements
- No centralised unit for data provision – different laws mean access can take a long time depending on sensitivity of data and who is providing access
- Need to align sensibly to national/international legislation but also not simply accept standards of other communities if these do not fit ethically or are not robust
- Need to align within nations as well – ADRUK, e.g.

These discussions raise points that will be crucial to consider in light of the first report's recommendation to develop standardised documentation (e.g., data agreements) to be used across data services and beyond. Such documents would need the flexibility and complexity to take into account different legislative requirements when working internationally, without becoming so complex as to be unworkable.

## Technical and human infrastructure

A common theme emerging from the discussions was the importance of human infrastructure, alongside technical infrastructure, and the need for a long-term perspective. Relationships between data owners, data services and researchers need careful structuring, maintenance, and time to develop. The large amounts of data involved will require resources not just for access but for maintenance and to allow the services to develop and evolve over time. Some key points raised in discussions:

- Challenge of mediation between data providers and data services; diversity of data providers: some are keen to provide metadata and discuss data generation processes, some not so much
- Issues with sustainability of large amounts of data – can the dataset survive beyond the period of initial funding for the project?
- Importance of adopting a long-term perspective, not only how services will provide data over the next few years, but how infrastructure will evolve over time
- Need to make researchers aware of methods required for employing datasets
- Communication strategies of data services – e.g., too many inconsistent acronyms and jargon
- Issues of responsibility – whose responsibility is it that data is fit for purpose for researchers? – policymakers, data curators, or researchers themselves?
- Move towards citizen science – need to make sure any data governance structures are not exclusive to certain groups

## Access

Another key theme that has remained prominent in both this and our last report is barriers to access to digital footprints data. In this workshop, access was still a key theme and participants did raise issues of barriers, but the focus was also on how to ensure that access is provided in a secure way which fosters trust without creating unnecessary challenges. Many participants spoke about the need for robust standards

to be in place and the importance of data services being seen to adhere to those standards.

- Balance between openness and privacy – security without creating unnecessary barriers to using data, which slows down workflows
- Robust rights management needed, with accuracy prioritised over speed – tendency of researchers to try to Hoover up as much data as possible, but then possibly not be able to use it
- Exemptions to GDPR for certain uses of certain data
- Drawing lines between data coming from social media and the commercial sector
- Scale/volume of data can be problematic – it can be so huge it's difficult to process or analyse
- Who owns the data can be a barrier
- Institutions err on side of being risk-averse which dampens research

## Quality and trust

Following from above discussions about the need for balance between secure, trusted environments and the need for open access, another prominent theme emerging from the workshop was how to ensure the trustworthiness of data services in terms of their reputation and relationships with data providers and with the wider public.

- How to tell if a data service is trustworthy – accreditation
- How to assess a data source for trustworthiness/quality – do data services work together to point people in the right direction?
- How do researchers decide where to deposit data?
- DFD can be commercially sensitive/valuable
- Issues of public trust
- Researchers asking about access to data can draw attention to collection of data that owners do not want attention on

## Best practice examples

The workshop provided the opportunity to hear examples of best practice, not just from the invited speakers, but from participants as well. Some key practical examples raised:

- Researcher 'passport' being developed in US and Germany
- UK reasonably open when compared to other nations in terms of negotiating access if researchers can show their work is in the public interest (e.g., access to banking transactions)
- GESIS:
  - Helps researchers find data who are unfamiliar with the process
  - Simple and clear communication
  - Understands what researchers need – power, software, ease of use, tech support

## Outstanding issues and open questions

### Standardisation

A key issue that has been raised in previous workshops as well as here is that of standardisation of documentation. This has particular implications when considering the international



context where there must be consideration of differing legal implications. The design of documents will need the flexibility and complexity to take into account different legislative frameworks when working internationally, without becoming so complex as to be unworkable.

### Linkage and interoperability

As linkage of data is a priority that has been raised in all workshops and groups, there is a need to think about not just the legalities of working across international boundaries, but what the mechanics of linkage will be across different data sets in different geographical locations.

# SUMMARY: RESEARCH INFRASTRUCTURE WORKSHOP

## Executive highlights

A key focus emerging from the workshop on Research Infrastructure was on standardising processes and legalities across different data centres. This aligned with discussion in prior workshops and reiterated a need to 'bake-in' flexibility and in infrastructural planning and design.

Needs identified include:

- To develop a shared set of standards to connect disparate data practices and policies
- To ensure transparency and open access across data services
- For ESRC to leverage 'soft power' to mitigate legalities which are hindering linkage
- To consider (and account for) differences in upskilling between researchers and those running data services

## Overview of event

The Research Infrastructure workshop took place on the 7th of June 2023, between 1300–1500pm. 69 people registered for the event, with around 40 attending on the day of the workshop. As with prior workshops, this two-hour session was facilitated by colleagues from The Collective. Participants were from a range of academic disciplines and included repeat participants from previous workshops. The workshop included a series of short provocations from several speakers, both presented in real-time and via video recording.

Speakers included:

### Fergus McDonald and Rob Baxter from DARE UK

Key takeaways:

- DARE's technical infrastructure and how it protects sensitive data (or data that then becomes sensitive through linkage)
- Stressed DARE's aims to develop a network of TREs which are interoperable and afford cross-domain linkage

### Kieran Jarret (ESRC) on Computing for Social Science

Key takeaways:

- Discussed using agent-based modelling to simulate large-scale social systems
- Identified challenges with data access and the incompatibility of workflows between services

### Sebastian Bacon from OpenSAFELY

Key takeaways:

- How OpenSAFELY uses synthetic data, open-source code, and auditable public logs to encourage transparency and ethical usage
- Identifying the platforms core principles as: Privacy; Efficiency; Transparency; Reproducibility

### Nick Bailey from UBDC

Key takeaways:

- The focus of UBDC being to use DFD to promote social, economic and environmental well-being in cities
- The importance of 'discoverability' in enabling data policy and practice

### James Cheshire from CDRC

Key takeaways:

- Consideration that an overlooked part of infrastructure is the people employed, those who have the knowledge to provide 'research ready' data
- How digital footprints data can be leveraged as an impactor for policy-making

## Emerging Themes

### Flexible design

In line with repeated themes surrounding the lack of a 'universal' infrastructure for DFD, it was reiterated during discussions between speakers that there is no consistency across existing data centres in fundamental aspects of processing, such as data formatting and coding languages used. The need for flexibility in design was thus emphasised, not only to accommodate the differences that exist between services, but also in acknowledgement of the continuing evolution of the DFD landscape, and the changes that will need to be made to adapt to a shifting terrain. Participants cautioned against building in obsolescence to DF infrastructure, by ensuring that the commissioning of DF data services be construed as an iterative process that can better account for changes in the data landscape. It was agreed that consideration of different potential mechanisms for data generation was important, and that infrastructure should be made adaptable to these differences.

### Interoperability and linkage

As in prior workshops, interoperability and linkage were still prevalent as key themes. The topic was approached that although there is a lot of activity currently occurring in the arena of digital footprint data infrastructure, it is all rather disconnected, or fragmented. Participants identified features of infrastructural design which could mitigate disconnection between services, such as improved searchability amongst datasets and enhanced version control relating to features such as metadata. It was discussed how front-end design would be beneficial, as this would ensure infrastructure is more responsive to the specificities of user needs and ensures that data access is directed more towards query and linkage, rather than individual expertise/authority. It was also debated, as in previous workshops, how ESRC could leverage its power as a national organisation to manoeuvre around legalities

that are currently hindering linkage, with the goal being not to centralize data, but to ease accessibility for otherwise disconnected services.

## Outstanding Issues

### Skill Gaps

In terms of challenges, a consistent issue across the workshop related to skills and skill gaps – a challenge that extends to those running the services, as well as the researchers using them. As discussed previously, it was highlighted how different researchers bring different skillsets to data services, and that this range of abilities should be both accommodated and bolstered. The question emerged: how do research services ensure that they can cater both for researchers at the cutting edge of data research, and those who may be new to DFD research? For those running the services, or those with prior knowledge and experience, their skill gap may exist more in the areas of data curation or harmonization – it was debated often in the discussions, for example, whether it is more productive for data services to offer access to raw data, or curated (e.g., 'pre-processed') data, with one speaker evidencing how synthetic data could be adopted for the sake of best practice in data processing. This evolved into a discussion of who ultimately 'does the work' in data curation: the data owners, the processors, or the researchers? This presents, once again, an issue wherein standardisation of practices and standards across data services would offer necessary regulation and mitigation of skill gaps.

### Inclusivity and accessibility

As part of the discussion on skill gaps, it was established that there is not as much focus on researchers (end users) who are unused to working with DFD – those with no prior skills but 'lots of questions'. These researchers will need support with basics of coding, searching, and linking data sets. This creates demands for support that can differ drastically between one person to another, and ultimately trickles down into questions of accessibility and inclusivity: without a regulatory environment to mediate access, services (or the researchers themselves) would effectively be 'picking and choosing' who gets access to DF data, which can write exclusion into the very infrastructure of DF data processing.

As in previous workshops, participants stressed the need for services to set down expectations for data owners/custodians that data will be available relating to certain mechanisms, rather than to certain people or groups. Though this poses difficulties – namely when coming up against organisational risk aversion – it was stressed that inclusivity in infrastructure must be considered an integral part of best practice. Participants agreed that there was a need to push for resources that will lower as many barriers as possible. Overall, it was highlighted that a range of tools must be available to ensure that a diversity of expertise's and skills can be catered for. This, in line with previous issues, would ensure democratisation of access, and promote DFD literacy, whilst still ensuring that resources are available for those at the cutting edge of this type of data processing and research.

# SUMMARY: ADVISORY GROUP MEETING

## Executive highlights

- Need to use, align with, and adhere to what is already in operation (infrastructure, legislation, ethics).
- Existing set of 'FAIR principles' for scientific data management (findable, accessible, integrate-able) – could be operationalised for the DF programme
- Need to create a regulatory environment in which data services cannot pick and choose who gets access (or allow researchers to pick and choose)
- Setting up expectations of data owners/custodians that data will be available for certain mechanisms rather than for certain people/groups

## Overview of event

The Advisory Group (AG) and Strategic Advice Team (SAT) for the Digital Footprints programme met on the 5th of June 2023. There were 8 members of the AG present, and they were asked for their input on research infrastructure and how they would envision managing data services in terms of how to bring together data from providers, curate this data and make it available to researchers.

Discussion questions posed included: What do researchers need (both technological and otherwise) and how could data services provide it? How might issues of integration and linkage be addressed? What is expected from data providers?

## Emerging Themes

### Technological infrastructure

Relating to resources, an issue that was raised amongst Advisory Group members was the need for services to possess substantial computing power. It was noted that even existing TREs are not fit for large complex data sets, and that these models will need more complex tools that require both system stability and significant processing power (likely cloud-based) to ensure both ease of access and the ability to handle multiple users. It would be important, also, that in democratising the infrastructure of a DFD centre, centre security is not compromised, meaning that different levels of security would have to be built into the platform to ensure users do not end up in secure areas by 'default'. Again, this brings up outstanding issues relating to funding and resource allocation.

One user of an existing data service told us of the issues they faced; in theory this service allows requests for linking data but in practice this proves extremely hard to do as they are often understaffed, searches are not intuitive, and the data recommendations can be poor. They suggested the need for better searchability and metadata, as well as proper version control on data sets.

Some participants spoke about the fact that some TREs still provide researchers with secure access only in specific geographical locations, requiring travel which has implications for project costs and convenience. While providers do seem to be generally moving away from this practice, it was raised as something to be avoided when considering the setup of data services, instead offering access via secure cloud environments.

### Problems may not be easily solved through technological solutions

A point raised by AG members is that people working with Digital Footprints data often want to bring together different data sets, but in practice these sets must often be kept separate due to confidentiality or legal requirements. A participant pointed out that often services will try to solve such an impasse with new technology, but the issue is more one of regulation/social norms/terms of use from data custodians, and so solutions must be sought which are more holistic.

One participant's contribution took the form of several questions, asking how we might solve the problem of comingling data, how we can jointly access and analyse data, and how we could access that data 'where it is' rather than in the custody of a provider. They suggested that this, again, was not wholly a technological problem, as it requires an approach based on relationship and trust building in order to create an environment in which such data sharing would be achievable.

### Legalities

As raised in earlier workshops and groups, the issue of reconciling different legal and ethical requirements when working across national boundaries was raised, and one participant also pointed out that different types of data sets will have different legal 'wrappings'. A suggestion in line with standardising processes and legalities was that ESRC should leverage their supervisory power in negotiating access to data and relationships between stakeholders, applying 'soft pressure' to ease issues around licensing and proprietary data.

A fundamental problem is that there is not a universal model or source of data that can be streamlined on a UK service. This is a regulatory issue as data licensing conditions that govern proprietary data sets will structure the way researchers are able to share data. Participants pointed out that there is not yet a universally agreed approach in government either, heightening worries that as government departments become more aware of the value of the data they hold and seek to control its use, we will end up with each department having its own siloed secure environment, therefore making data sharing even more difficult.

## Qualitative research

One participant spoke at length about the challenges faced by qualitative researchers in the Digital Footprints arena. Doing interpretive work is a large area that will need a lot of support as these methodological approaches are often complex and can be unwieldy: for example, combining discourse and textual analysis with a scholarly corpus and a data set. Guidance on how to bring together these kinds of data are not clearly established or commonly defined, which poses another challenge in terms of integrating this sort of research in the work of the data services.

Similarly, this participant raised several questions about how to curate heterogeneous data in a way that would be useful for qualitative research methods, asking what would make data queryable and what interfaces would be needed to create data for interpretive research. These are issues that need to be considered in the front-end design of services rather than later in the process, as the interface used, and initial findability, will be key for this kind of work. They also pointed out that qualitative researchers trying to integrate data may 'sign off' from data infrastructure issues because they do not consider this as 'their kind of data'. Data services will need to take into consideration that those doing interpretative work on digital heterogeneous data sets will therefore need a lot of support. Skills training could be part of a solution to this issue, and participants also pointed out the need for a range of tools to be available so researchers can use the tools they are most familiar with/have skills with (e.g., not just Python but a range of languages).

## Best practice exemplars

One participant gave the example of the UK Data Service as good framework for infrastructure, stating its searchability, the resources offered, the good quality of documentation on the site and the attention to metadata, make it a useful 'roadmap' for what could be done for DFD infrastructure. It was acknowledged, however, that licensing issues were still 'tricky', as is often the case with data centres/resources (see next section). The same participant then offered examples of their own work-in-progress model as adopting a 'use it and lose it' approach to data access: researchers can apply to use data from different institutions and departments, and the data is linked for the specific purpose of the research, which (it was emphasised) was one of the few consistent requests made for data access, given the differing nature of data sets. This system therefore works on access relating to research questions, which promotes better linkage between data sets.

Another participant spoke about what they are doing that works well, for example working with researchers to find out what they need in order to access the centre's datasets; having data readily available; having systems in place to assess whether data is of good enough quality; having a simple interface which allows researchers to scope what is available without having to make official requests; having as much information as possible that is easy to find and accessible.

Many participants across workshops and groups have emphasised the importance of 'not reinventing the wheel' and using existing infrastructure, both literally in collaboration, and as exemplars to follow. For example, one participant suggested using UKDS as an exemplar for Digital Footprints data infrastructure as it is searchable and has good documentation and metadata.

## Outstanding issues and open questions

### Resource allocation

There is an emerging theme which presents the need to consider both technological and people-based solutions to the issues raised. Moving forward it will be important to consider where resources are best allocated in this regard.

### Regulations and governance

One key outstanding issue is the lack of consistency and/or standardisation in legal/ethical/governance/licensing requirements, which has been an ongoing discussion throughout all workshops and groups. Approaches to governance will need to consider how to create standardised documentation that aligns with what already exists and is flexible enough to allow for complexity without becoming overly complex.

# 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