Future data services

Elizabeth Green

Future Data Services Strategic Fellow

University of the West of England Bristol

31st May 2024

Contents

Introduction

The aim of Future Data Services is to enable ESRC to make a long-term strategic decision about how to invest in data services infrastructure from 2024 onwards.

Our vision for the supply of data services post-2024 is one in which the benefits of technology, new research practices, and improved legal gateways all come together to improve research productivity and deliver greater public good. This stream of work focuses on the Services, People and Skills strand.

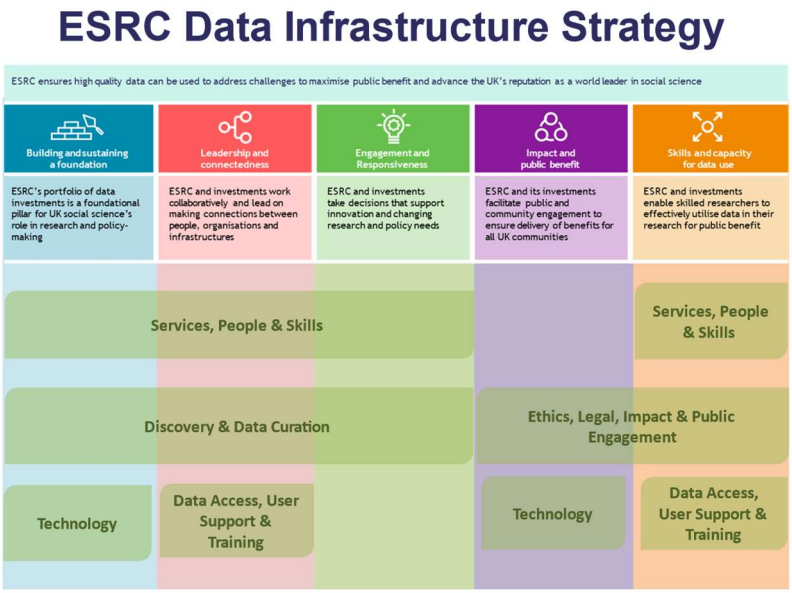

Figure 1 Adapted from ESRC Data Infrastructure Strategy (ESRC, 2022. P5).

The demand for data services has surged, partly due to government expectations for data accessibility and the need to demonstrate evidence-based policy and decision-making (Goldacre Review 2022, Digital Economies Act 2017). Consequently, there has been a significant increase in the number of applications and usage of these services, with the UK Data Archive (UKDA) reporting a 22% increase in secure lab users and an 80% increase in data collections from 2020-2022. The rise of data science has also led to innovative methodologies and approaches to data analysis. Frontline data service staff now often review outputs generated through new techniques like AI and machine learning.

Data service is a highly skilled and dynamic field, requiring adaptability to meet the demands of researchers, government agencies, data owners, and the public. Networks such as the Secure Access Data Professionals (SDAP) provide peer support and foster connections and development. However, there is a clear need for formal career pathways. Unlike traditional job roles with defined training requirements and progression routes, the data services sector is relatively new and constantly evolving, staff often find themselves in roles through drift or opportunity, the career of being data services professional is not widely known amongst graduates and the role itself is often poorly defined. This has resulted in few defined pathways or skill sets. This report found that staff retention and motivation are challenging, with high turnover, low diversity, and a lack of career goals and development opportunities.

The FDS analysis revealed that data services occupy a grey area within academia. Often treated as administrative staff, they do not fit into the typical progression structures that emphasize

managerial growth. In contrast, the FDS analysis of job specifications and recruitment showed that 'research' was the most requested skill, with 'academia' ranking fourth, suggesting a research-focused career structure. Figure 2 compares the current frameworks for data professionals with recruitment analysis findings, highlighting a disconnect between identified core skills and those prioritized by data professionals.

The Safe Data Access Professionals (SDAP) Competency Framework	Government Statistician Group (GSG) GSG Competency Framework	Government Digital and Data Profession Capability Framework	SFIA - a framework for cyber security skills	Skills identified from recruitment analysis (thematic areas)
Understanding analysis and analysts	Collect and prepare data	Asset and configuration management	Technology service management	Research
Understanding data legislation and licencing	Data quality	Continual service improvement	IT infrastructure	Data skills
Understanding data	Select and perform analysis	Customer service management	Network support	Publication
Understanding data setting	Analysis quality	Ownership and initiative	Database adminstration	Academia skills
Data acquisition	Present and disseminate data	Service focus	Release and deployment	Professional development
Metadata and data documentation	Disseminating quality of data	Service reporting	Storage management	Leadership
Quality assurance		Technical understanding		Communication
Data management		User focus		
Data transformation				
Support function				
Understanding analysis findings and outputs				
Statistical disclosure control (SDC)				
Effective analyst management and customer service				
Advocating for data				
Records and auditing and data security				
Training and education				
Data re-use and retention				

Figure 2 A comparison of skill frameworks and recruitment analysis

Interviews with data professional staff indicated that core skills should include soft skills (such as communication and leadership) alongside specialist skills like coding in R. This finding is echoed in the OECD (2017) report, where 80% of repository managers indicated a desire to undertake more activities but were constrained by infrastructure and funding limitations. This has led to feelings of disempowerment and demotivation among personnel. This report provides an overview of findings from the FDS project alongside recommendations for both data services and funders.

Evidence Gathering

To ensure that opinions was effectively captured, a three-way primary data collection approach was used to collect experiences from both researchers and data professionals. The different activities included reverse science cafes for both parties, a researcher-targeted online survey and 1-2-1 interviews with data professionals this then inspired a secondary data analysis of job advertisements (see figure 3).

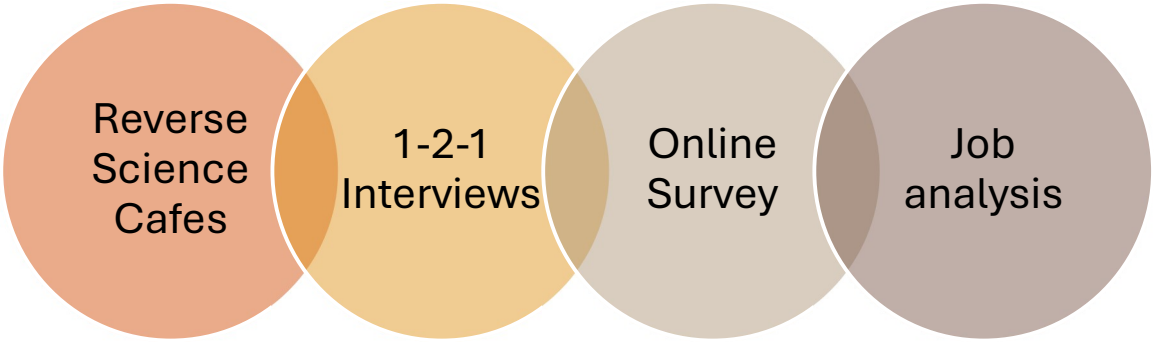

Figure 3 Evidence gathering activities

The structure broadly focused on the exploring researchers' thoughts, opinions and experiences on 6 key areas of FDS (see figure 4).

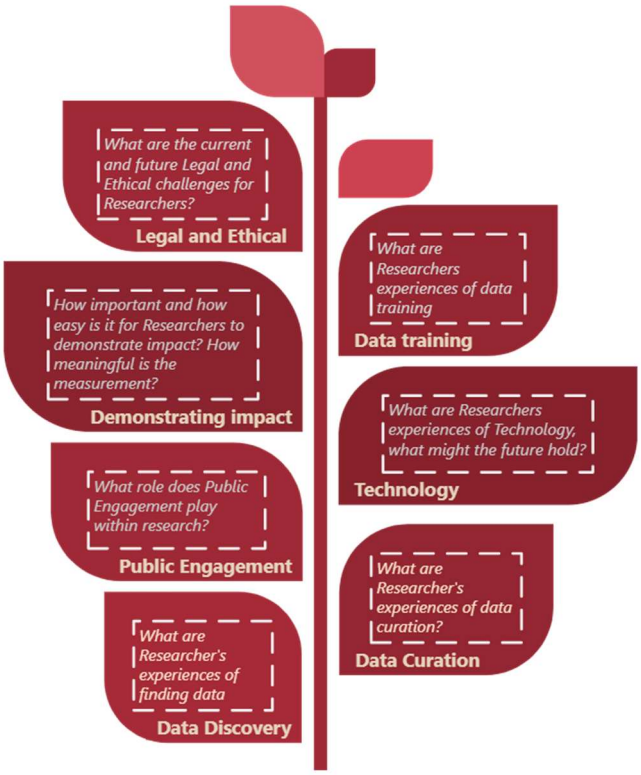

Figure 4 The 6 key areas of Future Data Services

Reverse Science Cafés

A reverse science café is a conversation related to a particular stream- we will ask a number of questions and then listen to answers from the audience to formulate their advice on how we can build and develop the work further[1]. Example of questions posed to participants in the 'Legal and ethics' reverse science café:

UWE Bristol University of the West of England

Questions

- What does Legal and ethical challenges look like for you?
- What are the challenges?
- Can you provide (good and bad) experiences of Legal and ethical training?
- Can you think of any quick wins or actions which would help the Legal and ethical challenges
- What do you think funders can do to help assist this?
- With the current challenges and pressures surrounding research how much significant would you place on the importance of Legal and ethical challenges?

Participants provided feedback during the event, and in follow-up emails. The reverse science cafes were open to all but aimed at particular researchers/teams/ workforces within the data and research community. An event page alongside an invitation to register was shared to the mailing list (see figure 5).

ESRC Future Data Services

The aim of the Future Data Services (FDS) programme is to establish the next generation of data services beyond 2024.

Go to this Sway

Figure 5 Example of sway invitation to reverse science café

[1] See Vlaams Instituut voor Biotechnologie guide for conducting a Reverse Science café [vib.be - Reversed science café]

Detailed notes were taken on the feedback raised during the reverse science café; this has been deidentified and shared in the next section of this report. Further detailed notes were taken on 1-2-1 sessions allowing the FDS team to reflect on points raised.

Reverse Science Café: Findings

Detailed notes were taken on the feedback raised during the reverse science café. This was then analysed and grouped into themes. The analysis was guided by the emergent themes rather than by any model in order to avoid imposing constraints. Thematic Analysis (TA) is the chosen method of qualitative analysis. This involved the researcher familiarizing herself with the data- noting ideas and relationships. The data was then mapped and organised into themes which are outlined below. This section outlines the emerged themes and the sub-themes within each. There were 5 core themes which arose from the different conversations

1. Legal and Ethical Challenges
2. Training and Terminology Challenges
3. Impact Measurement Challenges
4. Public Patient Involvement & Engagement (PPIE)
5. Data Curation Challenges

Legal and Ethical Challenges

On the theme of Legal and Ethical Challenges three master themes were raised: negotiating ethical boundaries; ethical concerns; and utility and technological concerns. Each master theme then possessed further sub-themes and mini- themes within these.

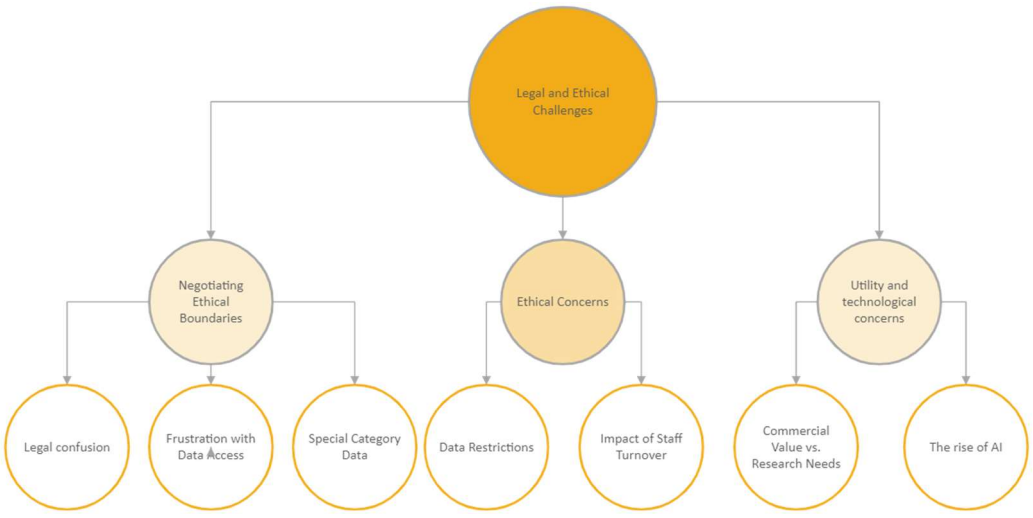

Figure 6 Legal and Ethical Challenges theme overview

Negotiating Ethical Boundaries

The crux of this theme was that researchers felt there needed to be more clarity and consistency surrounding access procedures, how the law was being interpreted and consideration for impact of not using particular datasets for analysis. One participant explained that they were researching a minority population and that trying to get detailed data was often fraught with challenges from data services due to concerns for disclosurey. Tthe participant reflected that there was little concern for ensuring that the minority population was being researched/ serviced.

Legal confusion: Researchers find it more challenging to navigate ethical boundaries, and gauge what requirements might be especially with newer regulations such as GDPR. Often within the same data service researchers had experienced differing opinions and differing advice.

Frustration with Data Access: Accessing data has become cumbersome and time-consuming. A researcher cited a delay of nearly two years to access data due to concerns over legal risks and staff turnover.

Special Category Data: Data related to sexuality, while classified as "special category," was paradoxically easier to trace through unrestricted data. Restrictions on such data raise ethical questions about the potential harms of not conducting research in this area and further driving data inequality.

Ethical Concerns
The sub-theme of ethical concerns relates to participants raising broader concerns with whether data services were creating unnecessary delays to data use. This was then further explored by the question which ethical process is deemed worthy of the application and why.

Data Restrictions: Researchers questioned whether it is ethical to restrict data collected for public benefit. For instance, when access to a particular dataset is required through a physical presence in a secure pod, this slowed the process due to travel needs and over subscriptions of the pods.

Impact of Staff Turnover: High staff turnover at data services caused significant delays and, it was felt, contributed to mixed advice and guidance. One researcher reflected that they had experienced mixed advice, and felt that the data service staff were well meaning but over-stretched.

Utility and technological concerns
This theme pertains to (1) concerns surrounding technological advancements and how data services (and indeed more broadly research) will keep pace, and (2) the creep of commercial use of data and whether this may shape data acquisition and availability in the future.

Commercial Value vs. Research Needs: The commercial value of data often conflicts with research community needs. Data collected for commercial use is often designed differently and holds different information and characteristics which focuses on the utility of the data, in contrast to research driven which focuses on the quality of the data. There was concern that data services in some places support commercial access and whether going forward the focus for the services will be on supporting commercial insights as opposed to research insights. It was emphasised that not all data services provide access or support to the private sector but there were concerns surrounding an eventual creep.

The rise of AI: The rise of AI is seen as a legal and ethical threat, particularly for data providers lacking technical backgrounds, fostering greater risk aversion. It was noted that often data services lacked expertise in this area and held a defensive stance.

Training and Terminology Challenges

This theme describes the training and terminology challenges participants described within the cafes. It consists of three subthemes; insufficient training, challenges in data applications and training gaps.

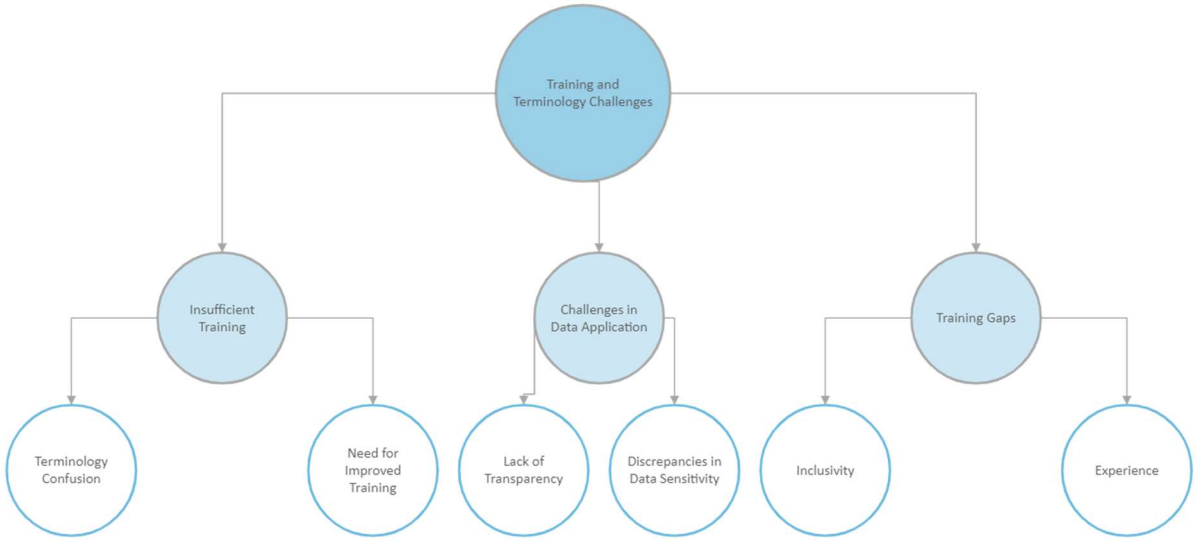

Figure 7 Training and Terminology Challenges theme overview

Insufficient Training

This theme highlights the lack of consistency and application of particular terminologies, and considers the wider need for training at an earlier stage.

Terminology Confusion: There is widespread confusion over terms like “data controllers” vs. “owners” and “anonymity” vs. “de-identified”, and how this is related back to the researcher and how they use the environment/ data. A common reference was what the NHS classically consider ‘anonymised’ would be considered ‘de-identified’ in the secure data communities.

Need for Improved Training: Researchers suggested better training at universities across disciplines embedded at an early stage, particularly for early career researchers (ECRs) and less experienced users. It was noted that often research methods and analysis training at university focused on ethical applications and informed consent, there was no training in secure routes to data access (through data services) and no training in disclosure control for outputs.

Challenges in Data Application

This theme considers the process of data application and how a lack of consistency across data services can create delays- this was attributed to high-staff turnover in particular services.

Lack of Transparency: The application process for data access is opaque, particularly concerning who is authorised to approve/ verify ethical issues that can cause delays.

Discrepancies in Data Sensitivity: Sensitive information like location data was found to be inconsistently restricted and metadata was not always clear.

Training Gaps

This theme covers the training gaps, both for data service staff and for researchers. The gaps predominately covered (1) lessons learnt/ avoiding common pitfalls in applying for data access and (2) ensuring data services were trained to consider wider ethical considerations of ensuring data equality and representation for minority groups.

Experience: Learning common pitfalls made when trying to access data seemed to be only obtained through the experience of undertaking a data application. A PhD student changed their research topic after failing to access necessary data over five years. Early communication about potential access issues could prevent such situations.

Inclusivity: Questions were raised about the inclusivity and awareness among data services about ensuring data equality especially for minority groups. It was felt that data services and data owners needed to have training and a commitment to ensuring data equitability and representative for minority groups.

Impact Measurement Challenges

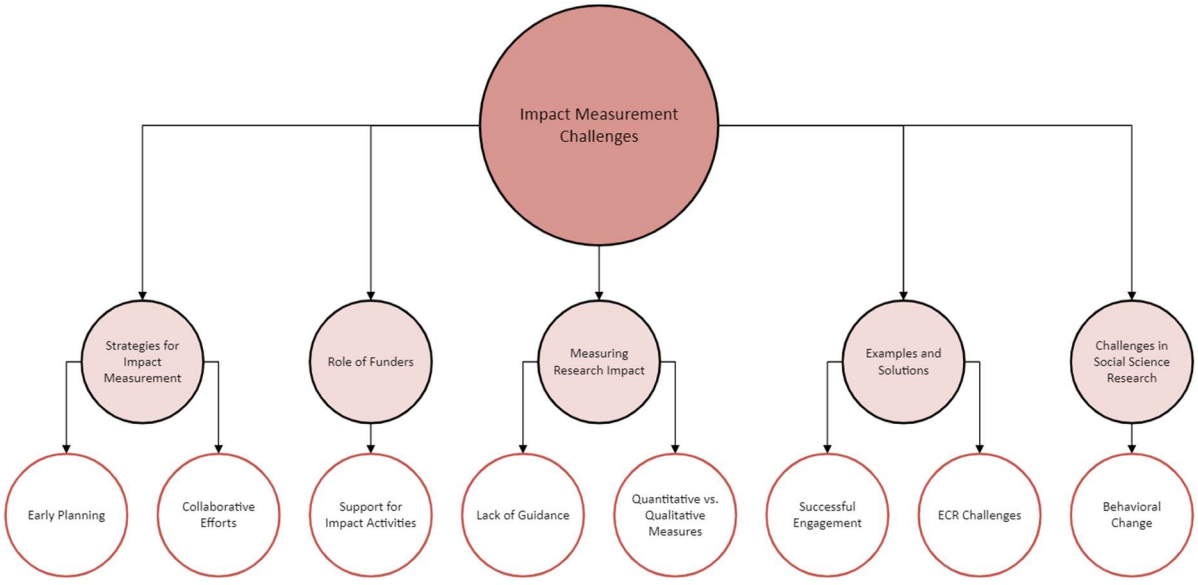

Figure 8 Impact measurement challenges theme overview

Measuring Research Impact

Lack of Guidance: Researchers often lack training on how to measure research impact, confusing engagement with impact.

Quantitative vs. Qualitative Measures: Measuring impact quantitatively (e.g., policy changes) can be impractical. One data service explores relational values, but obtaining feedback on research influence is difficult.

Strategies for Impact Measurement
Collaborative Efforts: Universities are developing databases to track references to hosted research and policy changes. Collaborative projects with external organizations are beneficial for measuring impact.
Early Planning: Early planning of impact measurement is recommended but can be hampered by a small pool of advisory resources.

Challenges in Social Science Research
Behavioural Change: Impact in social sciences often involves long-term behavioural change, making it hard to measure. Continuous relationship-building is crucial for sustainable impact.

Examples and Solutions
Successful Engagement: Data services and wider community provided examples of having hosted events which was successful in engaging policy makers and facilitating impactful networking across the data landscape.
ECR Challenges: Early career researchers struggle to build contacts, and are often dependent on senior academics. Mentorship and permanent job positions could help.

Role of Funders
Support for Impact Activities: Funders should include time and resources for impact measurement in proposals. Encouraging good leadership and providing internships for ECRs could broaden their experience and network.

Public Patient Involvement & Engagement (PPIE)

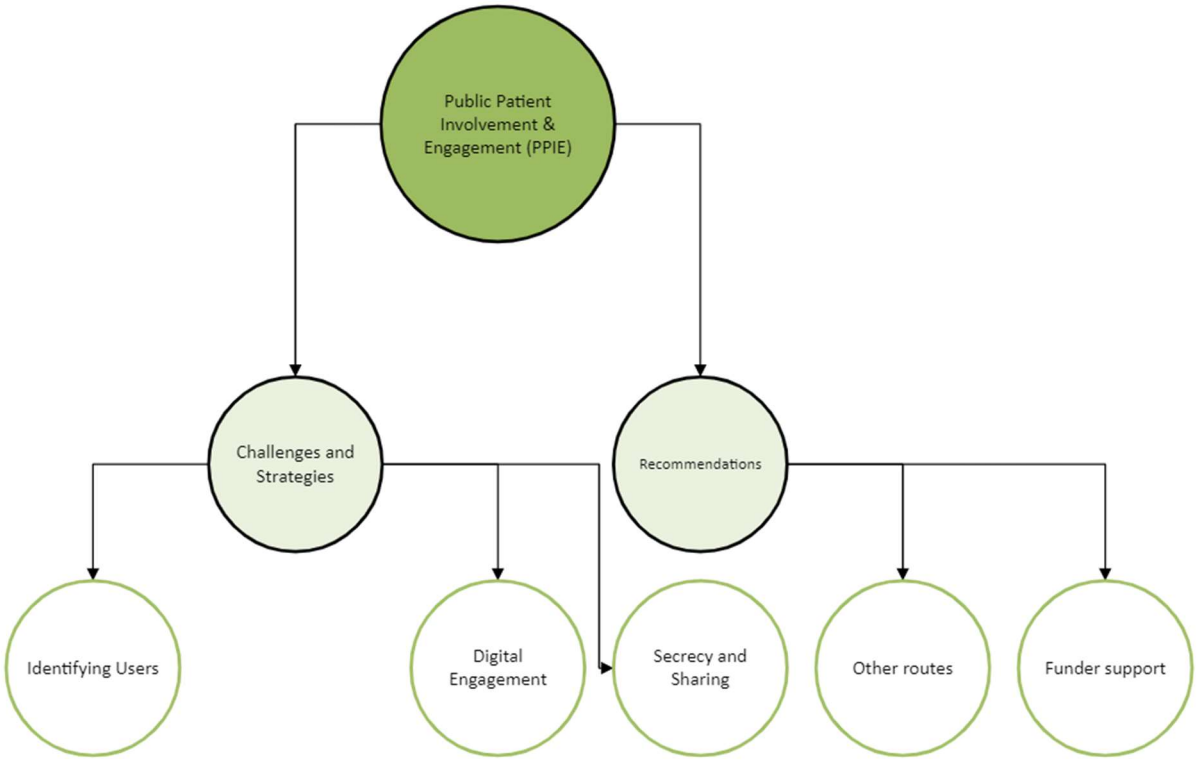

Figure 9 Public Patient Involvement & Engagement (PPIE) theme overview

Challenges and Strategies

Identifying Users: Data services needs to better identify and reach potential users without technical backgrounds. Partnering with organizations for targeted workshops could help.

Digital Engagement: Online training and tools like Mentimeter, Twitter, and YouTube have been effective for public engagement.

Secrecy and Sharing: A culture of secrecy hinders data sharing. Tailored communication and early education on data access could address this issue.

Recommendations

Other routes: Researchers need access to data to develop proposals. Synthetic data could serve as a less risky alternative for preliminary access.

Funder Support: Encouraging code sharing and synthetic data use, along with a dedicated outreach role, could improve engagement.

Data Curation Challenges

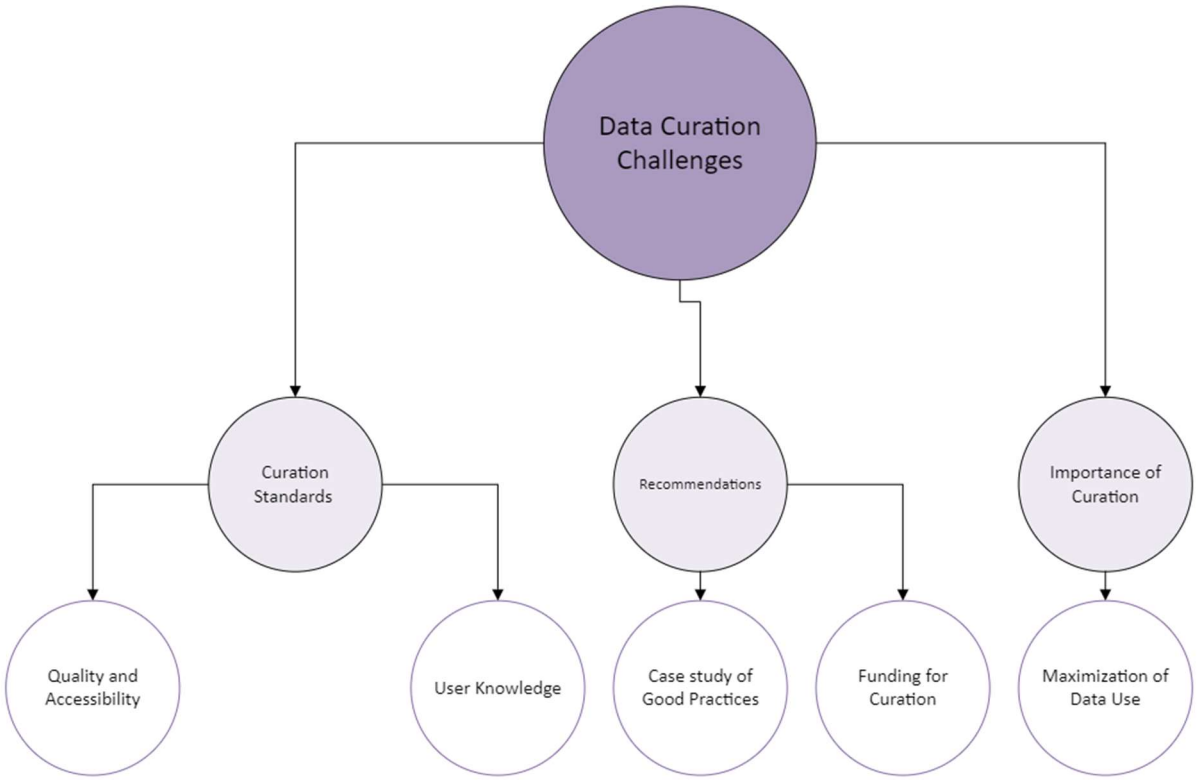

Figure 10 Data Curation Challenges theme overview

Curation Standards

Quality and Accessibility: Well-curated data should be clean, well-documented, and accessible in a safe environment.

User Knowledge: Different levels of user knowledge require tailored training and support.

Recommendations

Good Practices: Understanding Society data is well-documented, whereas some datasets have issues, like unrealistic default values skewing results.

Funding for Curation: Funders should allocate resources for data curation and access post-project to ensure data is well-utilized.

Importance of Curation

Maximization of Data Use: Proper data curation ensures broad access and reproducibility, crucial for validating research and minimizing redundant efforts.

Researcher survey

To support the Economic and Social Research Council's (ESRC) strategic review of data services infrastructure – the Future Data Services (FDS) – a survey was conducted to review experiences with accessing economic and social data in the UK. This survey was developed by ESRC FDS Team alongside the Strategic. It was part of a broader engagement initiative, its primary objective was to inform ESRC's funding strategies for data services infrastructure beyond the year 2024.

At its core, this survey was designed to explore the firsthand experiences of individuals in accessing economic and social data within the UK. Of particular interest was the experiences of users regarding microdata.

Question summary

The questions asked during this survey fall into the following categories.

Demographic questions

The first set of questions covers the participants basic demographics, information about the participants organisation and their specific role.

Questions were as follows:

- Age category
- Gender
- Ethnic group
- Organisation region
- Organisation sector
- Role within organisation
- Managerial role?
- Discipline of work

Data use questions

The next section covered the types of data that participants used and the availability of data and services within their area and organisation.

Questions were as follows:

- Type of data set worked with
- Level of data anonymisation
- Frequency of data activities (data searching, applying for access, acquiring data, processing data and analysing data)
- Size of data held and worked with
- Where data is acquired from
- Data type (quantitative, qualitative, or mixed)
- Where is data deposited
- Data services used

- Ease of use of services
- Frequency of use of services

Data discovery and training questions

The following set of questions covered the training provided for participants and how they discover data information.

Questions were as follows:

- How do you find information on data?
- What do you do if you struggle to find data?
- Is your training adequate?
- Training you would like but do not have provided

Applying for data questions

This set of questions discussed experiences when applying for data and the application process of obtaining data. Participants were asked to draw from their most recent application.

Questions were as follows:

- Most recent data application
- Where was ethics application made
- Was university ethic accepted by data service
- (If applicable) How easy was it to understand the application for Accredited Researcher Status
- Ease of booking a Safe Researcher Training course
- (If applicable) Did UKDS inform you when application was sent to ONS
- If data custodian had any queries, how were you informed? How did you address them?
- (If applicable) How long did it take for Accredited Researcher Status to be approved
- How long did it take for application to use data to be approved
- How often were application status updates given
- Any other comments on application process

Self-assessment tool questions

This section covers experiences using self-assessment tools and how they were discovered by participants.

Questions were as follows:

- Ease of use for UK Statistics Authority self-assessment tool
- Was advise given on accessing and completing the Ethics Self-Assessment Toolkit?
- How did you find the Ethics Self-Assessment Toolkit?
- Ease of use for the Ethics Self-Assessment Toolkit
- Who was the Ethics Self-Assessment Toolkit sent to
- Additional comments

Data experiences and barriers questions

This final block of research questions assessed barriers experienced by participants when accessing data and how they felt that these barriers impacted their most recent project.

Questions were as follows:

- Which types of barriers have you encountered accessing data?
- Description of barriers experienced
- Ranking issues by level of impact to project
- Have delays to data acquisition prevented a project from proceeding
- If data approval is needed, when was the approval sought out
- Best practices regarding data services you have experienced
- Any additional comments

Final questions

Finally, participants were asked if they wished to be involved in this work further, either through future interviews, workshops, both or neither and asked to (if they did wish to continue their involvement) provide a contact email address to facilitate this. This came with an option to opt out of joining a mailing list for this programme of work.

Survey Results

In total, 98 participants responded to the survey. Not all participants replied to all questions.

Demographics

Age

Medium age of participants was 35-44 years old. The data showed a slight skew towards the older age categories.

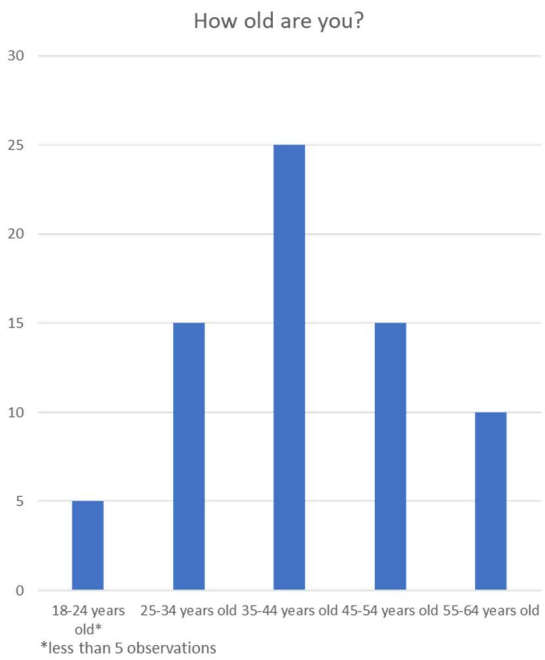

Figure 11 Survey Demographics: Age

Gender

Significantly more women completed the survey than other genders with 59% of respondents being female, 29% being male and 12% other.

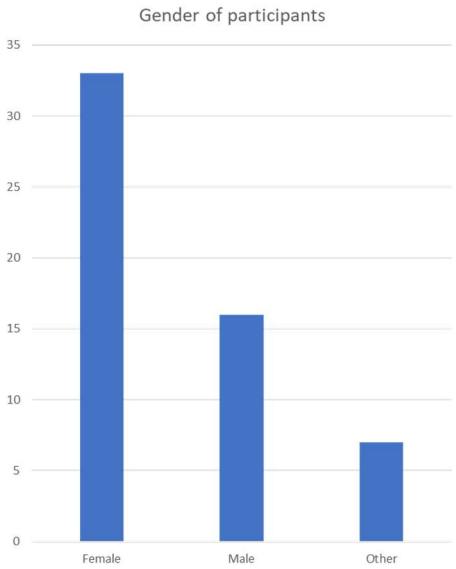

Figure 13 Survey Demographics Gender

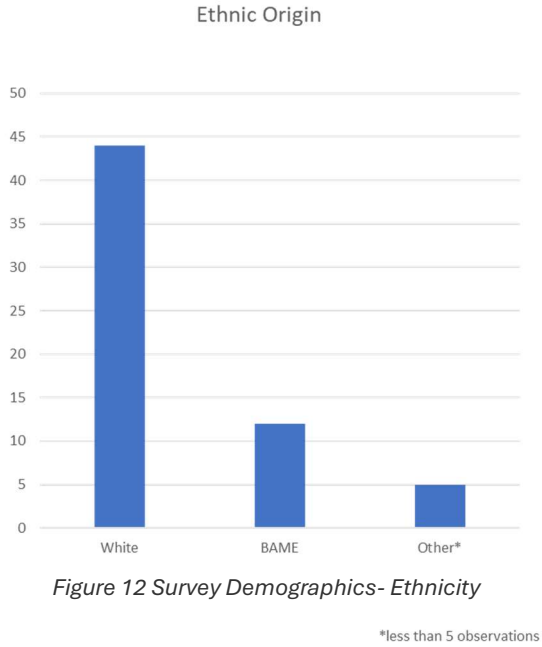

Figure 12 Survey Demographics- Ethnicity

Ethnicity

The majority of the respondents were white, with 70% falling into that category. 12% fell under Black, Asian and Minority Ethnicities (BAME) and less than 5 respondents selected other or chose not to disclose. The actual number in this category is obscured to prevent identification.

Disability

1 in 5 of the participants reported having a long-term physical or mental health condition or illness.

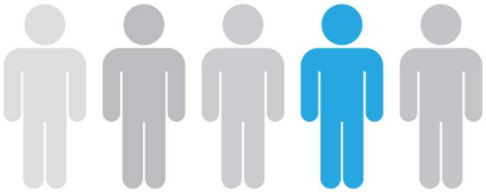

Figure 14 Survey Demographics- Disability

Location of respondents

Responses were received from participants across England, Wales, and Scotland, providing a broad spectrum of perspectives from different regions. The regions with the highest representation, in descending order, were London, Northwest England, Scotland, and the East of England. Additionally, smaller numbers of respondents contributed from the Southeast, Southwest, East Midlands, Yorkshire and Humberside, and Wales, adding diversity to the dataset with insights from various areas.

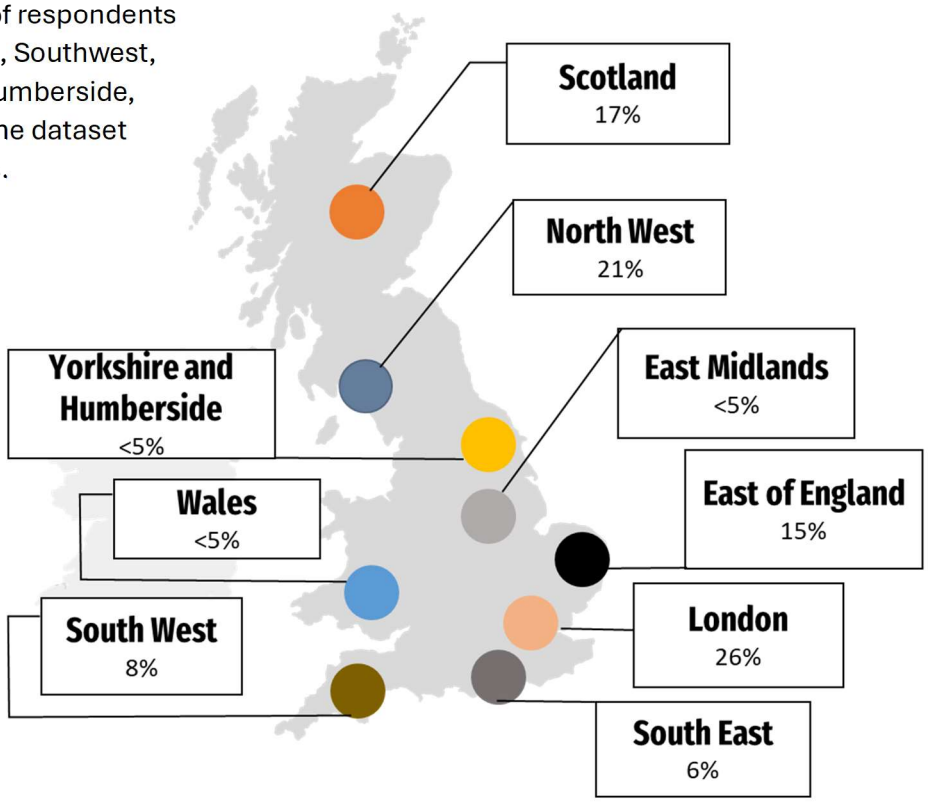

Figure 15 Survey Respondents

Organisation of respondents

90% of respondents worked in Higher Education institutions, with the other 10% working in Central government, the Charitable sector or other. Therefore, these results are most likely to be applicable to the experiences of researchers working in the Higher education sector.

Role of respondents

The survey revealed a diverse range of participants career stages. A significant portion, constituting 40% of respondents, identified themselves as mid-career or senior researchers, indicative of seasoned professionals contributing their perspectives. Following behind, 20% represented post-doc or early career researchers, highlighting the presence of emerging perspectives within the professional sphere. Students, including undergraduates, postgraduates, and PhD candidates, comprised approximately 15% of the respondent pool, with another 20% identifying as data managers or other data professionals.

Furthermore, a notable gender disparity in roles surfaced, with 46% of female respondents occupying junior positions, contrasting with only 13% of their male counterparts. This is especially interesting considering the large presence of female respondents.

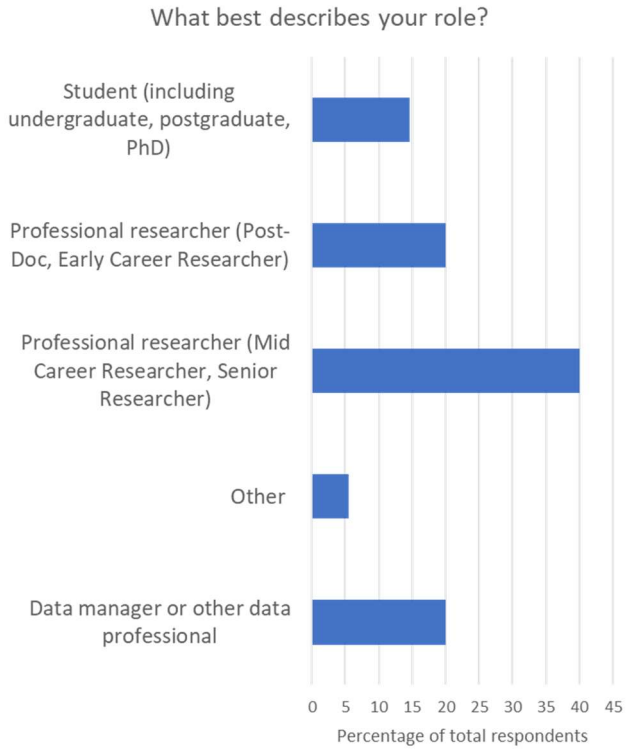

Figure 16 Survey Respondents Job Role

Research discipline

The survey respondents came from a wide range of research disciplines, with a couple being more prevalent than others. In descending order of representation, sociology was the most prevalent research discipline, with 20% of respondents listing it as their area of focus. Following closely behind, economics accounted for 15% of respondents' research interests, while health constituted 9%. Education and social statistics were each listed by 7% of participants. Demography and psychology were each represented by 6% of respondents. Additionally, environmental planning, management and business studies, politics, science and technology, and human geography each composed of 4% of participants. The other 13% listed their discipline as other. By encompassing such a range of fields, the findings of the survey incorporate a wide array of experiences and perspectives from across these domains. This inclusivity ensures that the insights derived from the survey are reflective of the varied landscape of research endeavours.

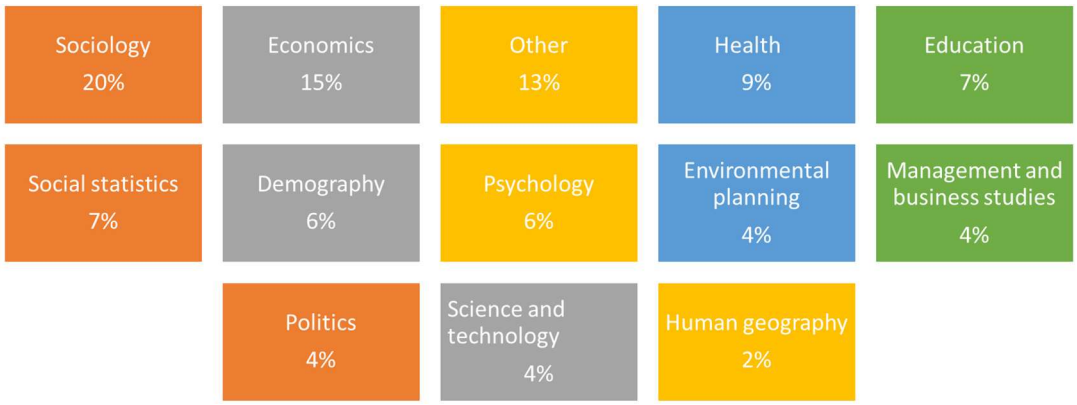

Figure 17 Survey Participants Research Discipline

Data types

The survey participants operate with a diverse range of data types and levels of control. A significant portion, comprising 40% of respondents, utilizes data sourced from external research facilities like the UK Data Archive (UKDA). Around 17% of participants acquire data that hasn't been specifically prepared for research, encompassing various sources such as administrative data, mobile phone data, and traffic information. Furthermore, 26% of respondents undertake the collection of their own primary data. In terms of data handling practices, approximately 70% of participants reported working with open data, while around 30% indicated working with closed data. Roughly half of the respondents, accounting for 50%, described their work involving controlled data. Moreover, over 70% of participants consider themselves working with safeguarded data. A small fraction, constituting 3%, either utilize data of other types or were unsure. Safeguarded and Open data types were by a significant margin the most commonly analysed data types. Participants also worked with quantitative, and qualitative data as well as mixed types.

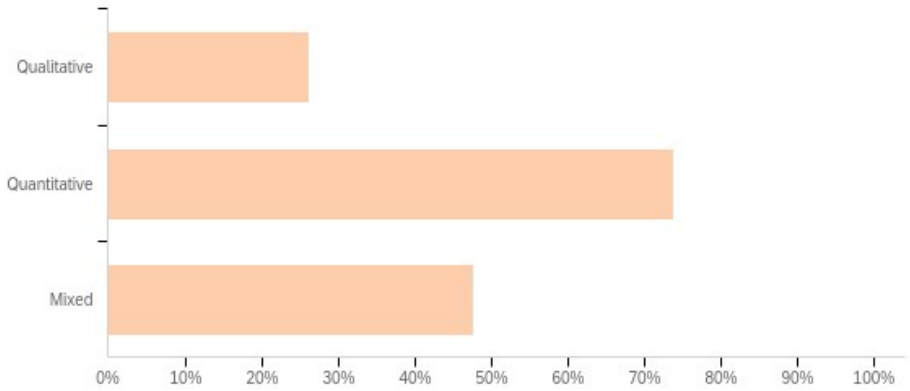

Figure 18 Survey Respondents - type of data participants were working with

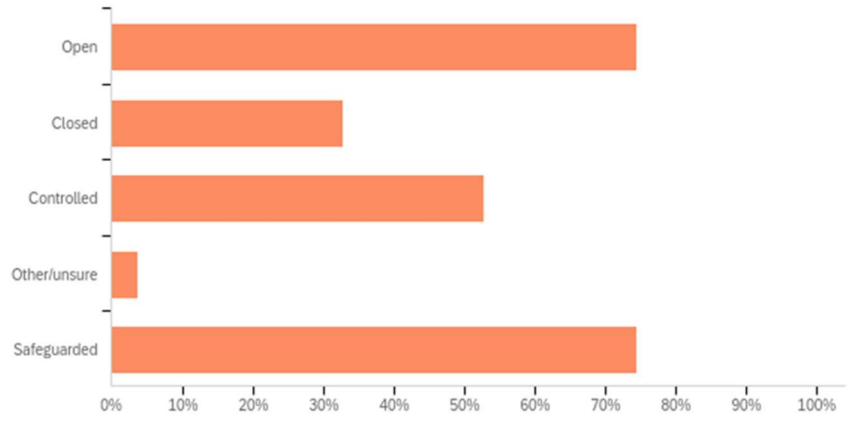

Figure 19 Security of data respondents accessing

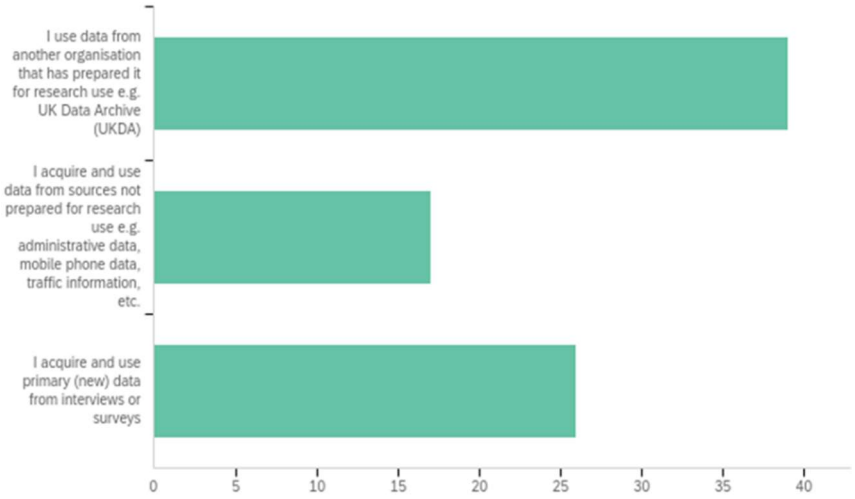

Figure 20 Methods of getting data- How survey participants gained the data they needed

Data Use Results

Data identifiability levels

A notable trend emerged regarding the types of data utilized by participants, with the vast majority, accounting for 89%, reporting experience working with anonymized data. Additionally, a significant portion, comprising 60% of respondents, had experience handling pseudonymized data, whereas a smaller proportion, only 25%, had worked with identified data. Across different disciplines, sociology, health, and economics emerged as the most prevalent users of anonymized data. Meanwhile, sociology, economics, and other fields were prominent among those working with pseudonymized data. Finally, sociology, economics, and management stood out as the most prevalent users of identified data, showcasing variation in data usage patterns across disciplines.

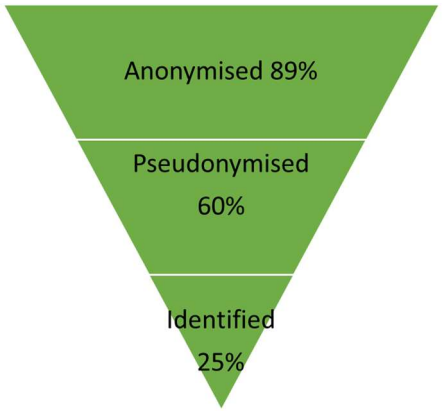

Figure 22 Level of identification in data used by survey respondents

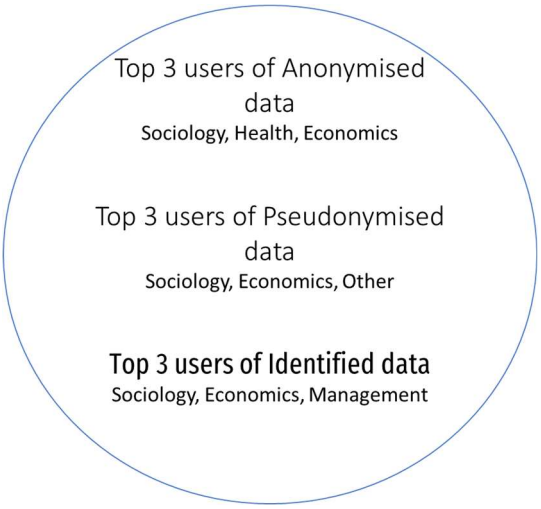

Figure 21 Which disciplines are the top users of identified data?

Data deposit locations

Among the surveyed participants, the UK Data Service emerged as the most common location for depositing data, indicating a prevalent preference for utilizing this resource. Following closely behind were repositories housed within the participants' institutions, slightly edging out dedicated data services within these same institutions. Notably, only a small minority, comprising four respondents, did not deposit their data in any archive or repository, suggesting a widespread adherence to best practices regarding data preservation and sharing within the academic community.

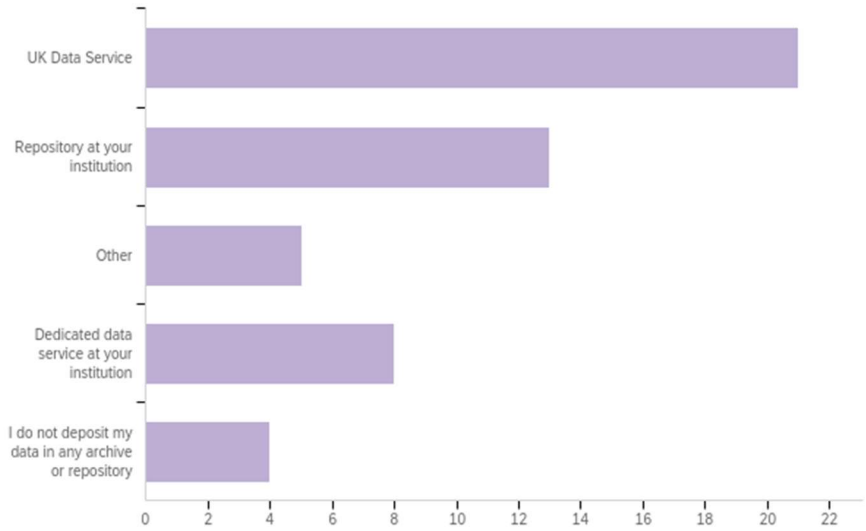

Figure 23 Data deposit locations

Data Service usage

The UK Data Service (UKDS) was the data service most frequently cited by participants, with 28% indicating its use. Following was the Office for National Statistics (ONS), utilized by 19% of participants, and CLOSER Discovery, accessed by 11%. Notably, the Office for National Statistics (ONS) Secure Research Service garnered usage by 9% of respondents.

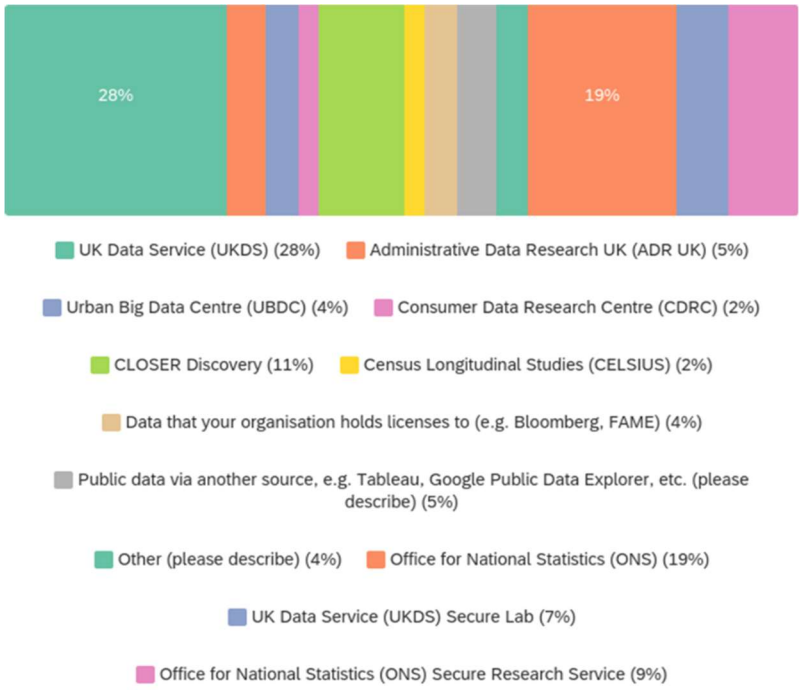

Figure 24 Data services used by participants

Data searching

Participants searched for data in a variety of ways, with google and other web searches and ESRC-funded data services each being used by over 70% of participants to locate needed data. Following this, most respondents (almost 60%) reported asking their colleagues for data recommendations. Interestingly, significantly fewer respondents reported asking support staff e.g. librarians for help locating data.

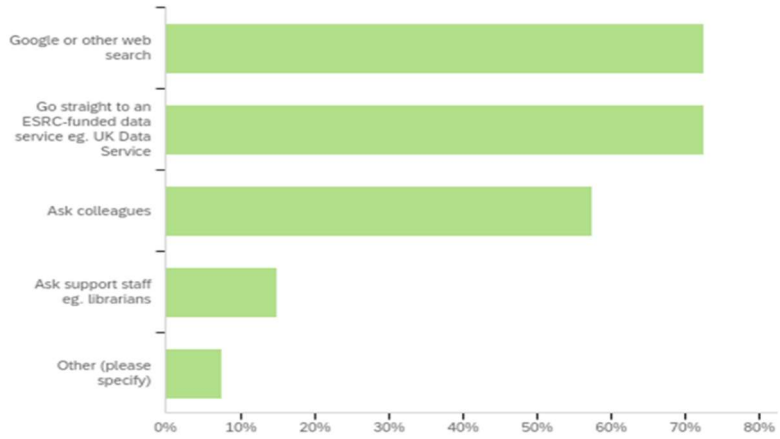

Figure 25 How participants searched for data

Approximately 70% of participants expressed confidence in their training to discover, access, and utilize data effectively. Moreover, when encountering difficulties in finding specific data, the majority demonstrated adaptability, with only 15% indicating they would 'give up' if they struggled to find needed data. Instead, around 60% indicated a willingness to modify their research proposal or question to align with available data or seek assistance from support staff or colleagues. Nearly 40% expressed an intention to develop a research proposal aimed at addressing the identified data gap, illustrating a proactive approach to overcoming research challenges.

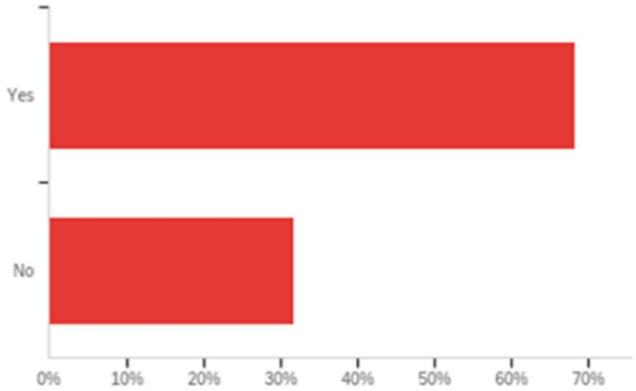

Figure 26 Participants confidence in their training to: discover, access, and utilize data

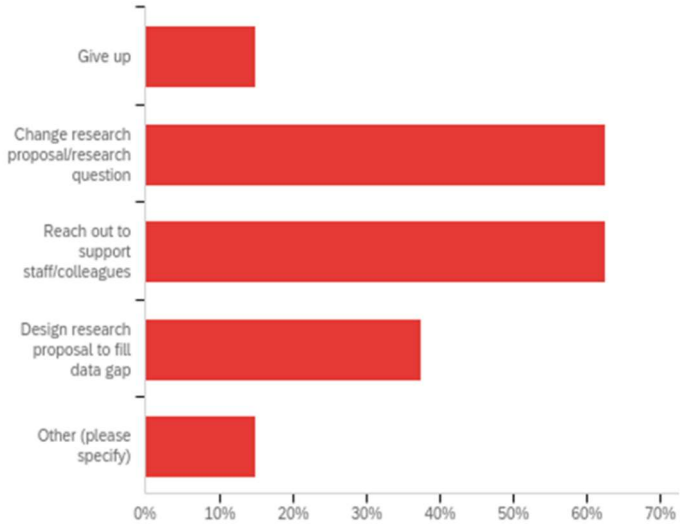

Figure 27 Responses to what do you do when you encounter difficulties in finding data

Application process

The following section was only completed by those who had made an application to access data within the last two years. 34% of participants were not eligible to complete these questions as either they had not made an application within that time frame (in the case of 29% of participants) or, they were not sure where the application was made to or it was taken care of on their behalf by the data service (5%). Of the 66% of respondents that had a recent data use application, the most common service to apply to was the Research Accreditation Service (RAS), a service that processes applications to access controlled Office of National Statistics (ONS) data. These applications were to access data through the UK Data Service Secure Lab or through the ONS Secure Research Service (SRS) at a similar rate (21% and 29% of total respondents respectively). Following this, around 10% of participants had made applications to the ESRC Accredited Research Scheme with a minority of participants (2% and 5%) applying to the Consumer Data Research Centre /Urban Big Data Centre or to other services.

Answer	%
Research Accreditation Service (RAS) for applications to access Controlled ONS data through the UK Data Service Secure Lab	21%
Other	5%
I haven't made any applications in the past 2 years	29%
ESRC Accredited Research Scheme (for applications to access Controlled non-ONS data through the UK Data Service Secure Lab)	10%
I'm not sure, it was taken care of by the data service for me.	5%
Research Accreditation Service (RAS) for applications to access Controlled ONS data through the ONS Secure Research Service (SRS)	29%
Consumer Data Research Centre or Urban Big Data Centre	2%

Figure 28 What application process have you used within the past 2 years?

Barriers and Tools

A significant portion of the participants who utilized the UK Statistics Authority self-assessment tool reported finding it somewhat easy, according to survey results. Approximately 20% of responses indicated that the tool was rated as either somewhat or extremely difficult to use, suggesting that the tool was comprehensible to a majority of users.

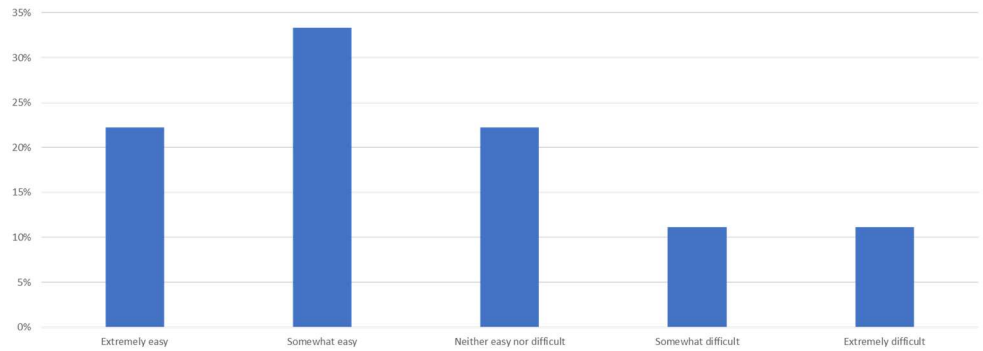

Figure 29 Participants responses to ease of use of UKSA self-assessment tool

For those who applied for Accredited Researcher status, the same trend emerged with most participants stating that they found the application somewhat easy to understand, however the distribution was closer to neutral with around 30% finding the application somewhat difficult.

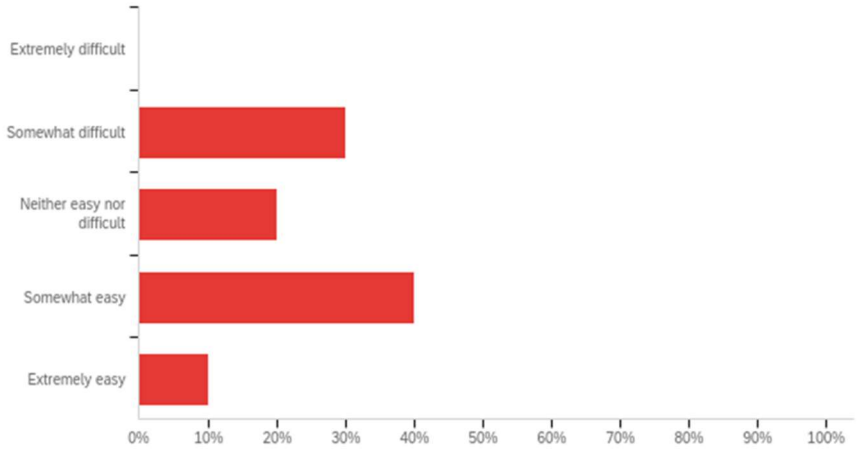

Figure 30 Participants experience of applying for accredited researcher status

When participants were asked to list the barriers they encountered in accessing data, the most frequently mentioned obstacle was the time taken to access the data, which was cited by a majority of respondents. Following this, nearly half of the participants identified obtaining permission or authorization to access data as a hindrance. Barriers originating from within participants' own institutions were reported more frequently than those from external sources. Approximately 25% of respondents indicated that they had not encountered any significant barriers, suggesting that encountering challenges when accessing data is commonplace. The word cloud below summarises the most used terms to describe the barriers to access.

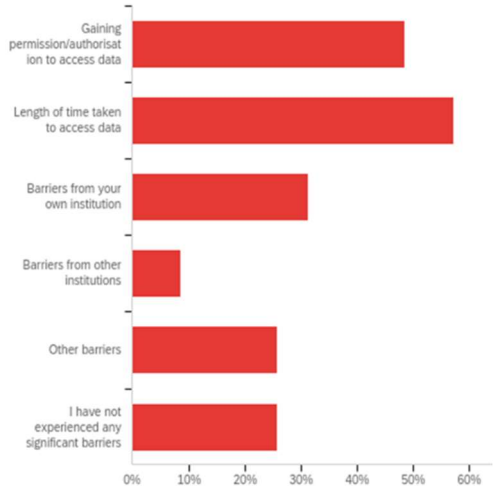

Figure 32 Participants experiences of barriers

Figure 31 Word Cloud describing barriers to access

When discussing the barriers contributing to significant project delays, the most commonly cited challenges included obtaining data within necessary timelines and securing permission to access data. Additionally, hurdles related to funding applications and acquiring staffing and resources were identified as potential sources of notable delays. In contrast, for less significant delays, acquiring ethical approval emerged as the most frequent issue, followed by challenges with funding applications, while other barriers played a comparatively smaller role.

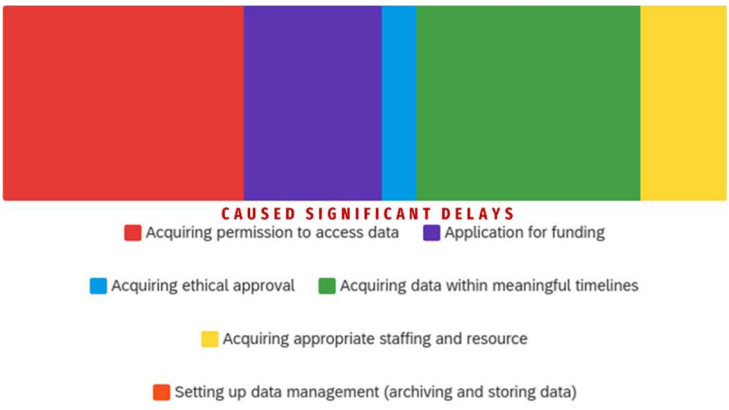

Figure 33 Barriers that caused significant delays

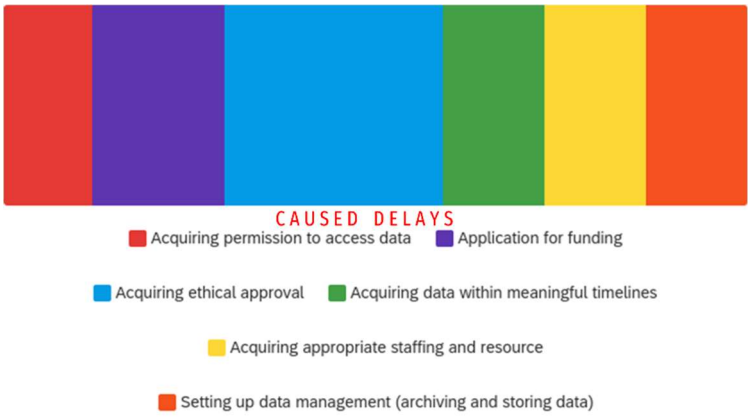

Figure 34 Barriers that caused delays

Interestingly, a significant majority of participants, approximately 70%, reported experiencing project delays leading to project failure. Among them, 50% indicated encountering this situation on multiple occasions. This demonstrates that the above delays are having a serious negative impact on the progression and success of research utilising secure data.

Examining a particular barrier, namely the time required to obtain approval to use data, participants were asked about their timing for submitting applications to utilize data for funded projects. Approximately 50% indicated that they submit the data usage application after the grant application has been approved, while between 20% and 30% submit it either before or concurrently with the grant application.

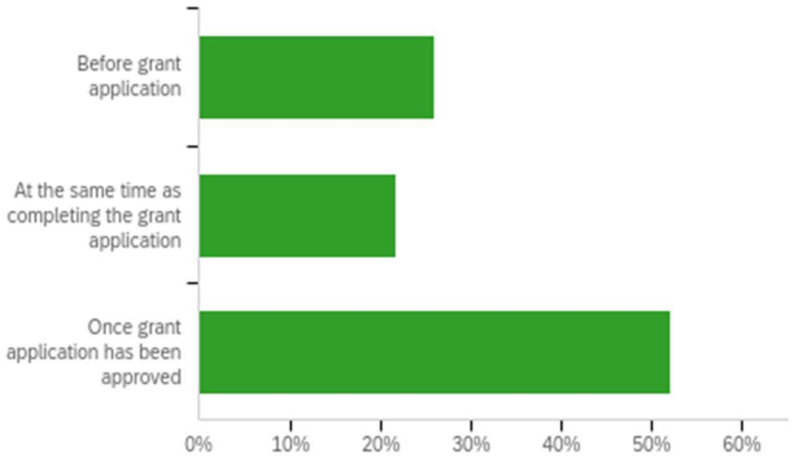

Figure 35 At what point do participants apply for data access

Lastly, participants were invited to offer their perspectives on best practices concerning data services. Their responses were analysed, and the most frequently used terms were visualized in the subsequent word cloud.

Figure 36 Word Cloud words associated with best practices concerning data services

Interviews with data professionals

40 people across UK data services were interviewed in a semi-structured conversation format. A wide range of different job roles and positions within the organisations participated, from senior leadership roles to administrative roles, technical roles and frontline roles. Access to an interview sample was facilitated by the ESRC and the individual data services themselves. The researchers developed a short questionnaire on Qualtrics, and the link was circulated by the data organisations to members. The Qualtrics questionnaire asked members who were interested in participating in the research to complete a short set of questions. Opportunistic sampling was used as, 'it entails us recruiting participants in any way possible (or convenient)'[2]. Being that the research target population is staff, and it is important that they feel under no obligation or coercion to participate, it seemed fitting to use opportunistic sampling. Staff were self- selecting; they all received the same invitation to be interviewed, which may result in a biased sample (i.e. individuals with extreme experiences of working in the services). However, as is the nature of working with qualitative data, it is befitting to explore and understand these experiences and how this can occur within the current structures of the data services.

The interviews took place online using Microsoft Teams and were recorded and transcribed using Teams functions. See the appendix for the questions that were used in the interview. The interviews were thematically organised, and the transcript was anonymised, coded, and classified using NVivo. The average interview lasted 44mins and the interview was structured loosely addressing the following topics.

- Entry and onboarding experience. Many people entering the field are new to working in data services how did they find the transition to working in this type of organisation (what helped, what might help future people coming into these roles).
- Job role. What are the key skills the staff feel is needed to do their job (i.e. good customer support and communication) and what skills do they feel is missing at the moment? Are there any skills they would like to develop in the future and how they feel these would support their role? What processes do they feel work well, and if they could make any suggestions for change what would they be?
- Future roles. What do they think the future of data services might look like and how do they think it will affect their department/ job? What areas do they think might evolve? For example, more complex outputs (brain scans, qualitative data, video data), more supporting researchers to develop and curate linked/ unique datasets, or (if admin/ finance worker) do they envisage say an increase of freelancers/ experts brought in to help with unique one-off tasks such as assistance with setting up a new process, or an expert providing disclosure control support for a large complex machine learning model.

[2] Langdridge, D., & Hagger-Johnson, G. (2009). Introduction to research methods and data analysis in psychology. Pearson Education.p85

- Career ambitions .What future ambitions do the staff have for their career- does it involve the skills they have acquired in their current job?

Analysis was guided by the emergent interview themes rather than by any model in order to avoid imposing constraints on the analysis. Thematic Analysis (TA) is the chosen method of qualitative analysis. This involves the researcher familiarizing herself with the data- noting ideas and relationships. The data was then mapped and organised into themes, which are outlined below. Data analysis took place alongside data collection, to ensure saturation of the data. We also endeavoured to involve as many different people as possible across different organisations (Glaser and Strauss, 1967).

Interview Findings

Entry and Onboarding Experiences

The experiences shared by the individuals regarding their onboarding to data services reflect a mix of initial uncertainty, adaptation challenges, and the importance of skill alignment.

One individual found the guidance from their manager crucial in understanding the necessary experience and knowledge for the role, which helped alleviate their doubts about fitting into the job.

"I'm not sure if I'm fitted for the job considering my experience. So, it was very helpful because he (manager) went through the experience and knowledge needed (to demonstrate) for the job."

Another person highlighted the lack of awareness about data services as a potential career pathway, suggesting a need for better promotion and information about these roles.

"It's not kind of clear that they exist, so I'm trying to think, do people actually know that this is a potential career pathway?"

The transition from an academic environment posed significant challenges for one individual, who took some time to adjust despite eventually understanding the job requirements.

"It took a while for me to find my feet, not because I didn't understand what was going on, and although that was part of it initially, just that academic world"

Lastly, someone felt confident in their choice of job due to its alignment with their skill set, indicating that proper matching of skills to job roles can enhance the onboarding experience.

"I went for this job because it suited my skill set more"

Takeaway Comments:

Clear Guidance: Effective onboarding should include detailed guidance from managers about the skills and knowledge required, helping new hires feel more confident and well-prepared.

Career Path Awareness: There is a need to raise awareness about data services as a career pathway, ensuring that potential candidates understand the opportunities available.

Support for Transitions: Onboarding programs should provide additional support for those transitioning from different backgrounds, such as academia, to help them adapt more smoothly.

Skill Alignment: Ensuring that job roles align well with an individual's skill set can lead to a more positive and efficient onboarding experience.

Job role and core skills

The key skills identified by staff for effectively performing their jobs in data services include clear communication, customer service, a willingness to learn, project management, and positive interaction with colleagues.

Simplifying objectives and tasks is crucial, as one individual highlighted the need to understand and meet objectives and tasks clearly.

"You know, just simplistically been held to say like what is an objective? What is a set of tasks like..Are we meeting or not? How?"

Others emphasized the importance of understanding other perspectives and maintaining excellent customer service.

"...trying to understand better"

"It is all about customer service"

A willingness to learn and good project management skills were also noted as essential.

"Willingness to learn."

"Project management"

Positive interaction with colleagues and a positive attitude were highlighted as important for a productive work environment.

"Interaction with your colleagues"

"Having a positive attitude"

Takeaway Comments:

Clear Communication: Staff need the ability to simplify and clearly communicate objectives and tasks to ensure everyone is aligned and understands their responsibilities.

Customer Service: Excellent customer service skills are vital, indicating that staff must be adept at understanding and meeting customer needs.

Willingness to Learn: Continuous learning is essential, as the field of data services is constantly evolving, requiring staff to stay updated with new knowledge and skills.

Project Management: Effective project management skills are necessary to handle tasks efficiently and ensure projects are completed on time and within scope.

Colleague Interaction: Positive and productive interaction with colleagues is important for teamwork and collaboration, fostering a supportive work environment.

Positive Attitude: Maintaining a positive attitude contributes to a healthy workplace culture and enhances overall job performance.

This word cloud summarises the most common words in response to questions regarding skills needed to work in data services:

Figure 37 Word Cloud- words rising from question surrounding skills needed to work in data services

Career ambitions

The future ambitions of the staff for their careers reflect a mix of desires for stability, growth, and further education.

One individual noted that staying with their previous organisation might have offered better pay, a more supportive framework, and greater opportunities.

"If I'd stayed in [organisation], I would have been paid probably more than I am now... I probably would have had more framework around me to. And in practical terms, more opportunity"

Another person expressed uncertainty about their current direction due to a lack of practical research experience in their day-to-day job, which contrasts with the research outputs they focus on.

"It's kind of bit difficult to see where I'm at now, which sort of direction to go in. Because we don't do a lot of ...because we look at a lot of research as outputs, and I don't really have much practise on the day-to-day job of doing that ourselves."

There is also an interest in potentially splitting time between work and further studies to enhance career prospects.

"I've thought about, like maybe splitting or going part time and then sort of pursuing further study alongside this job now."

Additionally, the importance of an organisation's culture in career development was highlighted as a key factor.

"The culture of an organisation is actually key in career development"

Takeaway Comments:

Stability and Opportunities: Some staff members feel that their previous roles might have offered more financial stability and career opportunities, suggesting that organisations should provide clear pathways for growth and development.

Practical Experience: There is a need for more hands-on, practical experience in day-to-day tasks to help staff feel more confident and clearer about their career direction and skills.

Further Education: Combining work with further education is a considered option for some, indicating that organisations could support staff pursuing additional qualifications to enhance their skills and career prospects.

Organisational Culture: A positive and supportive organisational culture is crucial for career development, emphasizing the need for a nurturing environment that fosters growth and advancement.

Future of data services

The staff's views on the future of data services highlight a dynamic and evolving landscape that will significantly impact their development and job roles.

The advent of synthetic and automatically generated data for real-time analysis may lead to a shift towards more dynamic data handling and perpetual updates.

"...synthetic data or automatically generated on the fly for new emerging or streaming data so that it's live real time, you know analysis is kind of like perpetually updating."

Some feel challenges may emerge in discerning the appropriate times and contexts to integrate new data effectively.

"The challenge is going to be knowing when it's right to absorb. The right bits at the right time, isn't it?"

Ensuring reproducibility within secure environments is another key concern, alongside the increasing importance of diverse data types, such as qualitative data and transcripts.

"Reproducibility and how that works with the secured environment"

"Different types of data like qualitative data, for example, or transcripts"

The evolving nature of roles in response to these changes is seen as a positive aspect, offering continuous learning and adaptation opportunities.

"Because of the landscape, there are changes in roles which is fantastic, the roles are never static"

Staff recognize that the rapid pace of change in the field requires them to stay agile and continuously update their knowledge and skills.

"It's a changing landscape, no matter how well you know a subject, it will change sometimes you blink. And something new appeared."

Takeaway Comments:

Real-Time Data Handling: The future of data services involves more dynamic, real-time data analysis, requiring staff to adapt to new technologies and methods for handling continuously updating data.

Data Integration: Knowing when and how to absorb and integrate new data is crucial, emphasizing the need for strategic decision-making skills in data management.

Reproducibility and Security: Maintaining reproducibility within secure environments remains a priority, suggesting that staff will need to focus on developing robust and secure data practices.

Diverse Data Types: The increasing importance of handling various data types, such as qualitative data, indicates that staff will need to broaden their skill sets to manage and analyse diverse data sources effectively.

Role Evolution: The continuously changing roles in data services offer opportunities for growth and development, encouraging staff to embrace flexibility and lifelong learning.

Agility and Learning: The rapid changes in the field underscore the necessity for staff to stay agile and continuously update their knowledge to keep pace with new developments.

Room to change

The staff responses reflect significant frustrations among staff regarding the limited room for change and development in data services.

There is a sense of inequity and reliance on informal, background communications, which hampers progress.

"...it's not very equitable"

"Kind of always fall on these soft communications, which are just happening in the background,"

Individuals often have to invest their own time and effort into career development, leading to dissatisfaction.

"I have to put my own time and effort into that. That's what my career's built on."

Outdated practices and a hierarchical organizational structure further restrict innovation and growth.

"We are behind the curve on all the sort of management thinking and all that sort of stuff, like clunky old practises."

Although there is recognition of the importance of skills like R and Python, the lack of structured opportunities to learn and apply these skills within work hours is demotivating and prevents staff from developing their skill base without investing significant portions of their own time.

"[Discussing skills] more users now who are using R and Python, so having knowledge of that would be a good thing. I've spoken to my manager about this and there's not really any sort of like use case or anything to actually learn these skills. So, it's really hard to learn it unless you're sort of learning your own time."

Overall, there is a pervasive feeling of organisational resistance to change and a lack of focus on personal development, contributing to employee frustration and stagnation.

"source of my frustrations is that nobody seems to want to do anything about it"

"We've always done it this. For someone new, it's just soul destroying."

"They don't focus on people at all."

Takeaway Comments:

Inequity and Informal Communication: The perceived inequity and reliance on informal communication channels limit effective collaboration and progress, suggesting a need for more structured and equitable communication practices.

Self-Investment in Development: Staff often have to invest their own time in learning and development, highlighting the need for organizations to provide more formal opportunities and support for skill enhancement during work hours.

Outdated Practices: The persistence of outdated management practices stifles innovation, indicating a need for modernization and adoption of contemporary management strategies.

Hierarchical Structure: The hierarchical "upstairs-downstairs" dynamic creates barriers to change and development, suggesting a need for a more inclusive and collaborative organizational culture.

Skill Development Barriers: While skills like R and Python are recognized as valuable, the lack of practical opportunities to learn and apply these skills within the workplace is a significant barrier, necessitating better integration of skill development into job roles.

Organisational Resistance to Change: The resistance to change and lack of focus on people and their development within organisations lead to frustration and disengagement, indicating a pressing need for a shift towards a more people-centric and flexible approach to management and development.

Job advertisement analysis

Following feedback from the interviews surrounding the complexity and ever-changing landscape of working in data services. Previous research in this area includes work conducted by Thielen & Neeser (2020) exploring hiring of data professionals to offer research data services in American University Libraries. These roles often require different experience than traditional librarian positions, increasing interest in hiring outside the typical library and information science (LIS) pipeline. The study included an analysis of 177 job advertisements from 2013 to 2018 found that over one-third (35%) did not use "librarian" in the title. While 88% required a Master's in LIS, 67% accepted an equivalent degree. The findings indicate academic libraries are recruiting from outside the traditional LIS pipeline, emphasizing both research data and traditional LIS activities.

For the FDS project we have conduced an analysis of the UK job advertisements for data professionals working in research and/or government infrastructures/ civil services. A total of 315 data professionals job adverts were used; the person specifications was also downloaded to allow a full exploration.

Analysis was conducted exploring the:

- Job requirements
- Skills advertised in the job advert
- Skills in the person specification

The aim of this analysis was to explore whether there was a potential disconnect between the job advertised versus the actualities of working in that job.

Findings

Using NVivo software the adverts were analysed. First, exploration surrounding the word ‘requirements’ demonstrated the most common requirement listed was, ‘nationality requirements’ with 190 of the jobs listing this. Due to recent legislative changes resulting in visa for graduates minimum salary for a sponsored skilled worker has risen from £26,200 to £38,700[3] which is higher than many entry roles into data services.

[3] Changes to legal migration rules for family and work visas in 2024 - House of Commons Library (parliament.uk)

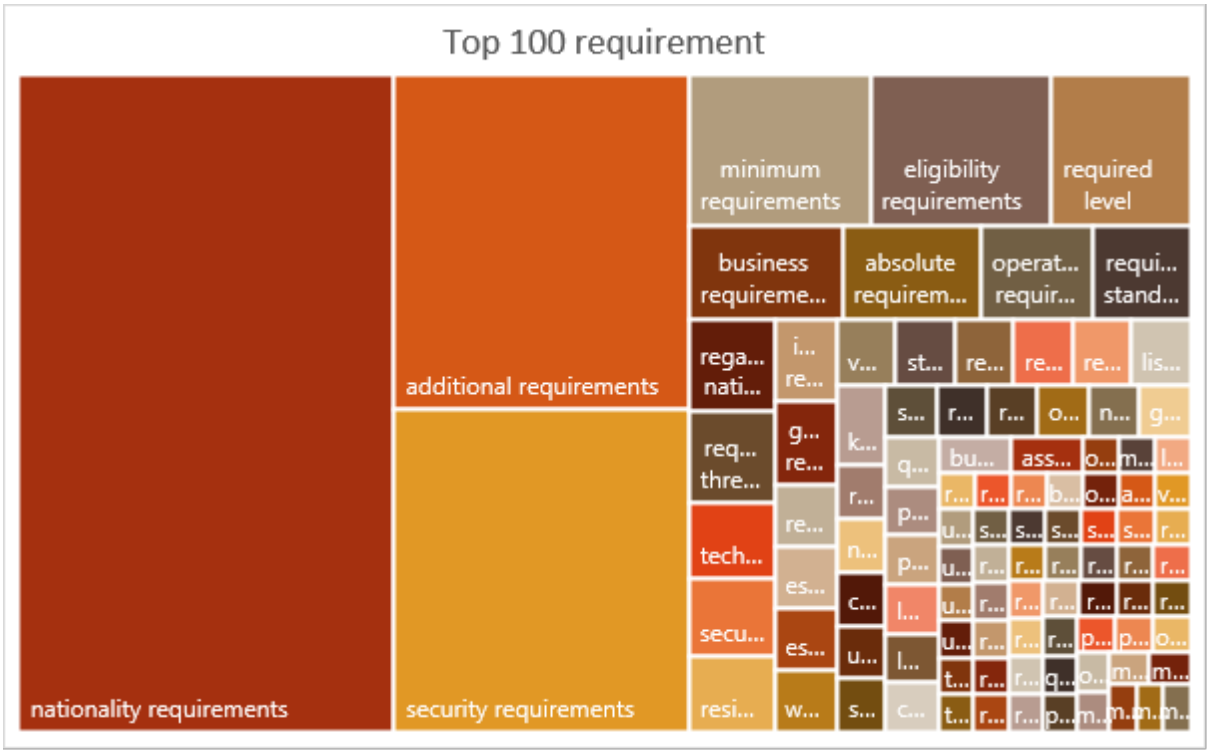

Figure 38 Top 100 requirements from the job advert analysis

When analysing the person specification documents the following traits/ key areas where identified with research, data, skills and academia being listed at the top. This is in conflict with skills identified within the interviews which focused on the more soft customer service components. On further inspection it was found that many job and person specifications appear to be derived from boilerplate templates: where data services form part of an academic institution, templates for professors and researchers have been reused causing a mismatch been job advertisement and job actuality.

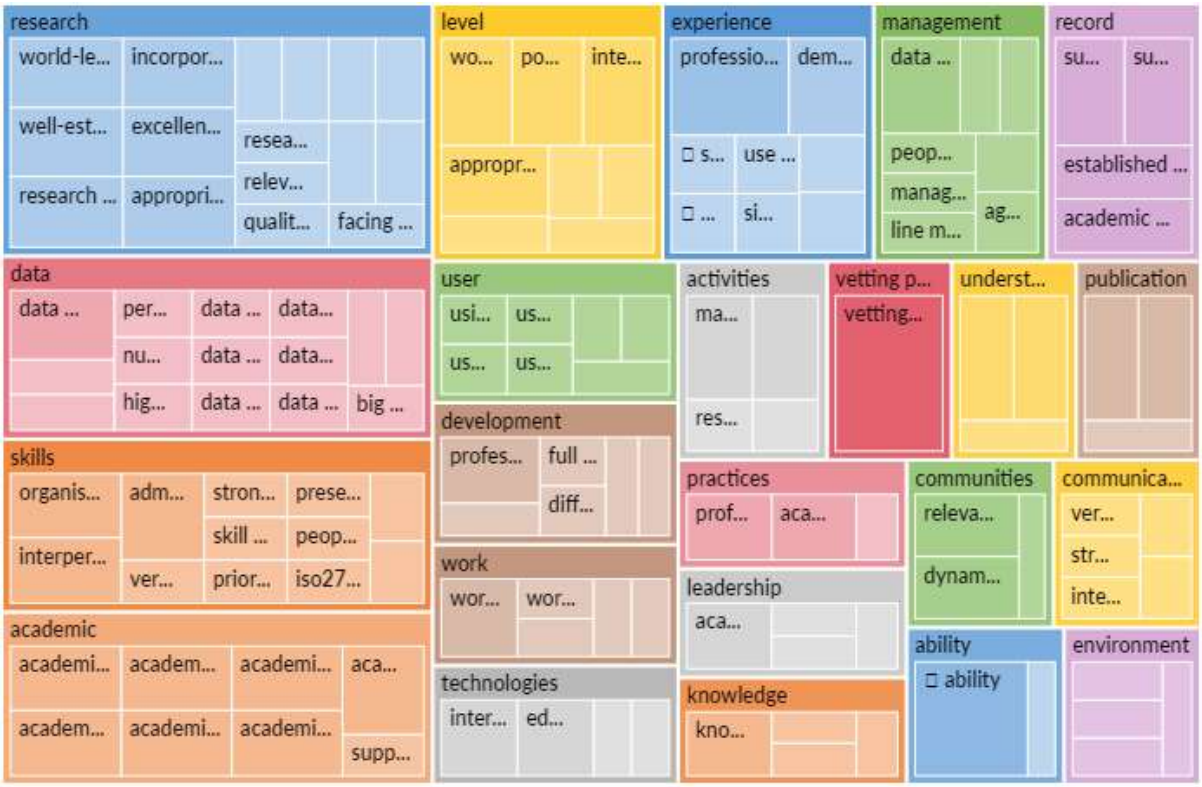

Figure 39 Key themes from job advert analysis

Figure 40 Detail of the Research theme from job analysis

Figure 41 Detail of data theme from job analysis

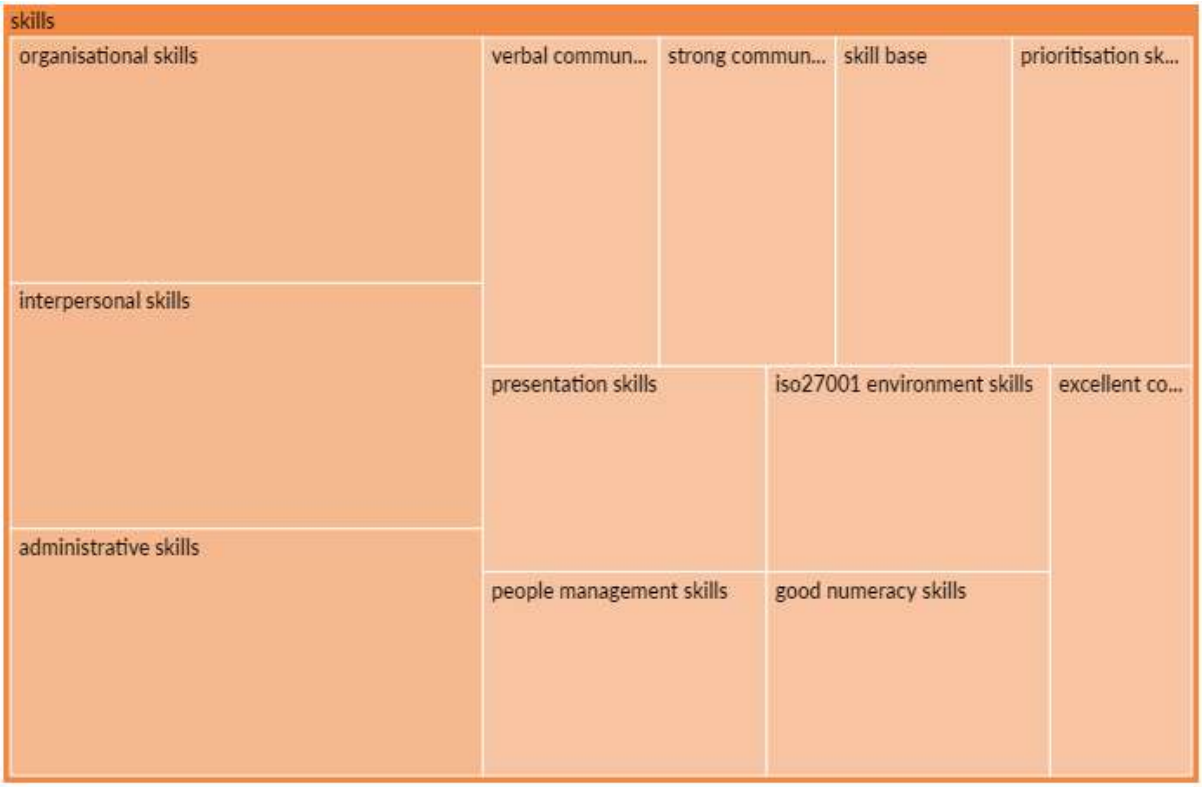

Figure 42 Detail of skills theme from job analysis

When exploring the top referenced skills the first most referenced skill was holding technical skills, in contrast in interviews this skill was found to be of a lesser importance.

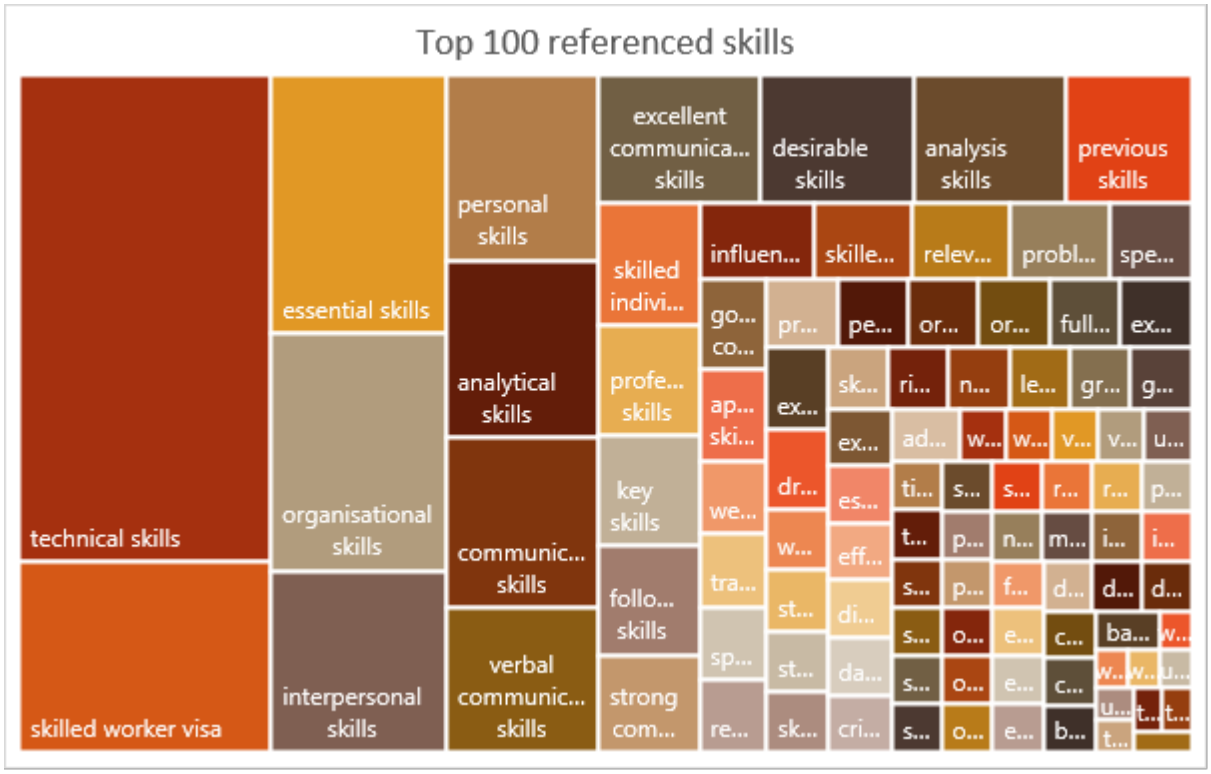

Figure 43 Top 100 referenced skills from Job analysis

Output checking retreat

In March 2024, the WUE FDS Strategic Fellows organised a three-day residential retreat on output checking, part of a TRE''s operating processes. The retreat was attended by 20 staff from 15 data services in health and social science. The retreat had two objectives.

The first, not relevant to this report, was to identify ''best practice" output checking processes and help participant organisations review and improve their operations against the standards set by others. The retreat resulted in a "Guide to output checking processes" currently undergoing peer review by TREs. The second objective, relevant to this report, was to explore whether some sort of ''data academy' could help data services share information and learn from collective engagement. Continual improvement will be one the key FDS recommendations, but the FDS team recognise that this requires some support to enable this to happen. The output-checking retreat was a test of the proof of concept.

On this measure the retreat proved a substantial success. Participants felt that the chance to discuss their day-to-day activities with people carrying out similar activities in other data services was the most important outcome of the event. A follow-up survey led to comments such as "It was an event I didn't know I needed!!! I'm in a small team and it was great to meet others doing similar jobs. There's no real community in SDC and this event provided community, support and interaction with others facing similar issues. It made me aware of the control we have as SDC checkers over the process; something I was taught was a process that was done this way like it had always been done. But now i see room for change, development and improvement."

"As the first of its kind, this workshop was fantastic! Not only did it provide an opportunity for experts from different TREs across the UK to share best practices and support each other in a rapidly specialized area, but the sense of community it fostered ensures a lasting impact among members."

"There are plenty of opportunity to talk to all of the participants, the atmosphere is friendly and supportive."

"Having an in-person event worked great - it's easier to engage with the sessions and meet people [...] in an in-person format. I think I got so much from the sessions being broken up by informal breaks and opportunities to honestly chat with other people in similar roles. It allowed time to reflect and un-pack the content of the sessions further."

Two follow-up hybrid and virtual workshops re-iterated the value of coming together with others doing similar jobs in different data services – for some of the smaller data services, there is only one person in the organisation doing a particular job and so cross-data service events can provide a substantial amount of support not available in-house.

The retreat and follow-up workshops all recommended future 2 similar events. While the time commitment was recognised as a potential burden, it was felt that this was outweighed by the deeper quality of interaction in the extended-workshop format of the retreat. Indeed, one participant suggested that the most productive part of the workshop had been the conversations held over dinner.

Participants also requested more attendees. The workshop had been limited to one participant per organisation in most cases due to limited funds, but there was strong enthusiasm for more

attendance, particular from more junior staff (invited staff were practised professionals in the area). It was recognised that a larger event might lost some of the intimacy of connections in the small group and suggested that perhaps more frequent events would be a way to square the circle. Finally, it was proposed that some sort of “data services conference” would be warmly welcomed as a way to build a community feeling amongst data service staff.

In summary, a regular (but not necessarily frequent) in-person residential gathering was identified as a strongly likely to build cohesion within and across data services staff, to help identify gaps and good practice in staff training and development and develop an appropriate sense of the uniqueness of data service skills.

Recommendations

We present two types of recommendations. The ‘recommendations for data services’ are intended to help data services improve their internal processes and/or support external users and stakeholders more effectively.

These are suggestions to data services. However, UKRI can play a significant role in helping data services implement change. These are the second set of recommendations, reflecting UKRI''s triple role:

1. As a funder who can directly influence how data services operate
2. As a research funder who can help build the evidence base for change
3. As an investor in national infrastructure, standards and training who can support activities that are best developed across data services

Recommendations for data services

All the recommendations in this section are concerned with changing existing practices to reflect unusual (in some cases unique) attributes of data service roles.

1.1 Develop a service level strategy for an inclusive workforce which reflects the particular skills sets needed for the data service

1.2 Define career pathways with training and skills matrix which encompasses actual job activities, particularly with reference to soft skills

1.3 If research is to be part of the job specification, ensure that sufficient time is allowed to enable independent research to be carried out

1.4 Capture meaningful KPIs, and benchmarks for accountability, that reflect data service job roles, not academic or administrative ones

1.5 Define recruitment targets and strategy for an inclusive workforce

1.6 Reflect on and refresh the phrasing used in recruitment documents/ adverts to properly reflect job roles in practice

1.7 Define staff on-and off-boarding processes to reflect the particular needs of data services staff, rather than using generic materials

1.8 Encourage staff participation in cross-data service activities to build a wider sense of community

1.9 Develop senior management/ leadership vision statements that show awareness of the particular characteristics of data service roles

Recommendations for UKRI

As a funder

2.1 When commissioning data services across multi-organisations ensure consistency in staffing contracts

2.2 Integrate a staff report with key metrics such as turnover, diversity etc. as part of funders report.

2.3 Where funded roles are claimed to have a research component, ensure that this is genuine and that funding is available for it

As a promoter of research:

2.4 Commission research into generating a diverse data professionals workforce

As a national body supporting data services

2.5 Support the development of national competency framework and best practice guidelines for role identification, recruitment, induction, onboarding and progression.

2.6 Initiate conversations with University and HE federations to examine staffing differences between academic posts and data professional posts, in particular focusing on the research component of such work.

2.7 Support the development of a range of residential activities designed to both showcase new developments in data services and allow data service staff to build an organic community.

References

Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. Qualitative research in psychology, 3(2), pp.77-101.

Economic and Social Research Council (ESRC) 2022, ESRC data infrastructure strategy, UK Research and Innovation, viewed May 2024, https://www.ukri.org/publications/esrc-data-infrastructure-strategy/.

Glaser, B. G. and Strauss, A. L. (1967) The discovery of grounded theory: strategies for qualitative research. Chicago: Aldine.

Goldacre, B., Morley, J., & Hamilton, N. (2022). Better, broader, safer: using health data for research and analysis. Department of Health and Social Care.

Government Digital and Data Profession Capability Framework. (2023). [Online]. Last Updated: 2023. Available at: https://ddat-capability-framework.service.gov.uk/role/service-desk-manager [Accessed 19 February 2024].

Government Statistical Service (GSS). (2021). The Government Statistical Service (GSS) Competency Framework. [Online]. Government Analysis Function. Last Updated: 2021. Available at: https://analysisfunction.civilservice.gov.uk/policy-store/competency-framework-for-the-government-st [Accessed 19 February 2024].

House of Commons Library. (2024) 'Changes to legal migration rules for family and work visas in 2024', UK Parliament. Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-7274/ (Accessed: May 2024).

Langdridge, D., & Hagger-Johnson, G. (2009). Introduction to research methods and data analysis in psychology. Pearson Education.p85

Lazer, D. M., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., ... & Wagner, C. (2020). Computational social science: Obstacles and opportunities. Science, 369(6507), 1060-1062.

Organisation for Economic Co-operation and Development. (2017). Business Models for Sustainable Research Data Repositories. OECD Publishing.

Secure Data Access Professionals (SDAP). (n.d.) Competency Framework. Available at: https://securedatagroup.org/guides-and-resources/competency-framework/ (Accessed: May 2024) .GSG competency framework

Secure Data Access Professionals (SDAP). (n.d.) Secure Data Access Professionals (SDAP). Available at: https://securedatagroup.org/ (Accessed: May 2024).

SFIA. (2023). SFIA - a framework for cyber security skills. [Online]. SFIA- The global skills and competency framework for the digital world. Last Updated: 2023. Available at: https://sfia-online.org/en/tools-and-resources/cybersecurity-skills-framework [Accessed 19 February 2024].

Thielen, J., & Neeser, A. (2020). Making Job Postings More Equitable: Evidence Based Recommendations from an Analysis of Data Professionals Job Postings Between 2013-2018. Evidence Based Library and Information Practice, 15(3), 103–156. https://doi.org/10.18438/eblip29674

UK Parliament. (2017) Digital Economy Act 2017. Available at: https://bills.parliament.uk/bills/1859 (Accessed: May 2024).