University of Dundee

Doing ESRC Data Better

Allanson, Paul; Daly, Angela; Elahi, Awais; Geddes, Alistair; Malone, Maeve; Nic Daeid, Niamh

[Link to publication in Discovery Research Portal](#)

University of Dundee
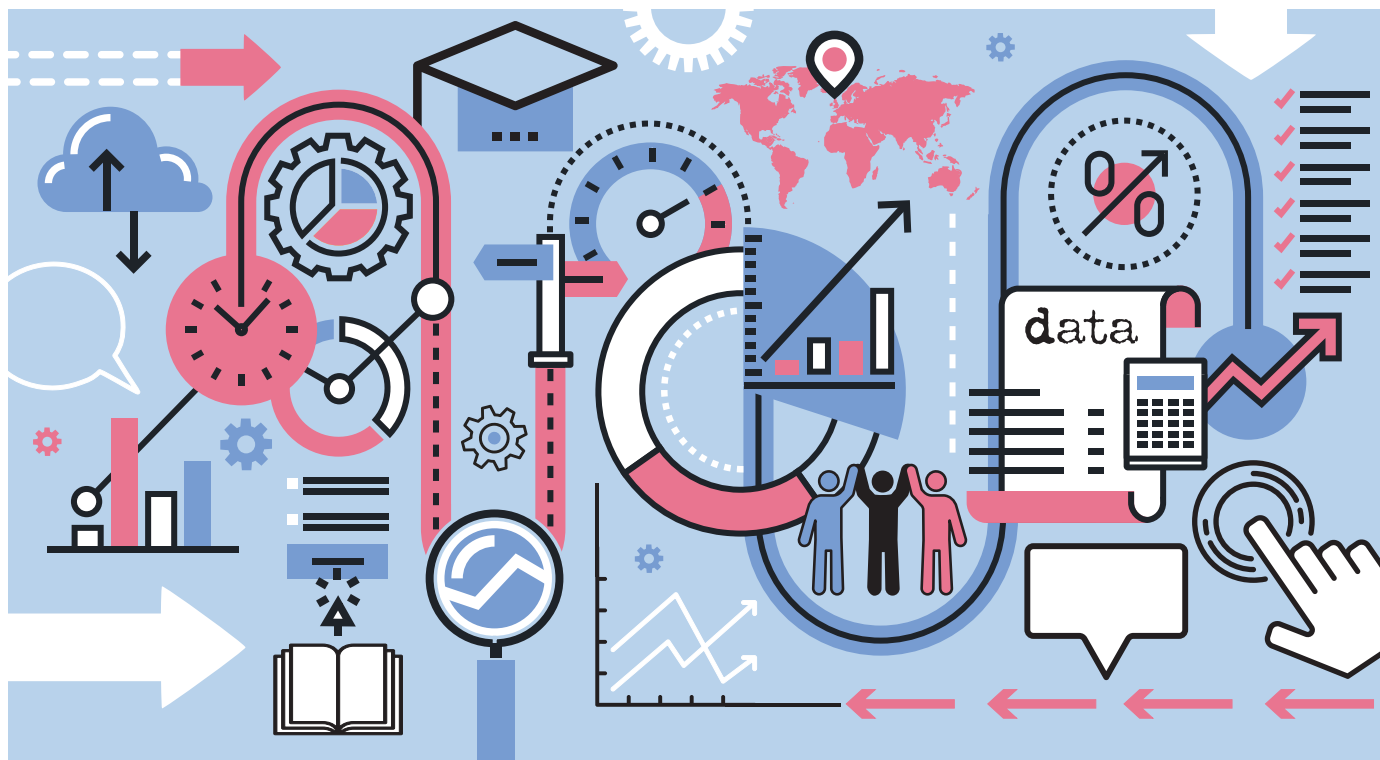
Doing ESRC Data Better

A study for the Economic and Social Research Council (ESRC)

Leverhulme Research Centre for Forensic Science – Science, Policy and Law series #2

Authors

Paul Allanson, Angela Daly, Awais Elahi, Alistair Geddes, Maeve Malone, Niamh Nic Daeid and Lucille Tetley-Brown – University of Dundee
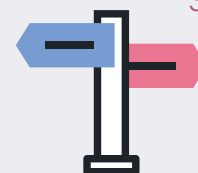
Acknowledgements

Contents

Acronyms

ADR UK	Administrative Data Research UK
AHRC	Arts and Humanities Research Council
AI	Artificial Intelligence
CARE	Collective Benefit, Authority to Control, Responsibility and Ethics
CDRC	Consumer Data Research Centre
Co-I	Co-Investigator
DMP	Data Management Plan
DOI	Digital Object Identifier
EDI	Equality, Diversity and Inclusion
EPSRC	Engineering & Physical Sciences Research Council
ESRC	Economic and Social Research Council
FAIR	Findable, Accessible, Interoperable and Reusable
FG	Focus Group
GDPR	General Data Protection Regulation
HDR UK	Health Data Research UK
ICO	Information Commissioner's Office
IP	Intellectual Property

ML	Machine Learning
MRC	Medical Research Council
NERC	Natural Environment Research Council
OECD	The Organisation for Economic Co-operation and Development
ORCID	Open Research and Contributor ID
PI	Principal Investigator
PID	Persistent Identifier
RCUK	Research Councils UK
RDA	Research Data Alliance
RDCS	Research Data Alliance DMP Common Standard
REF	Research Excellence Framework
TRE	Trusted Research Environment
TRUST	Transparency, Responsibility, User focus, Sustainability and Technology
UKDS	UK Data Service
UKRI	UK Research and Innovation
UoD	University of Dundee
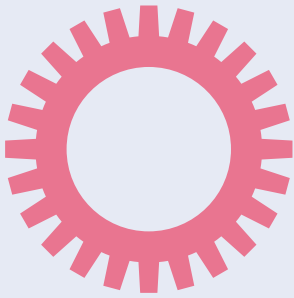
Executive Summary

Ensuring open data and open science in changing times remains a priority for publicly funded social science research in the UK. Understanding the changing landscape is key to achieving this and implementing it through appropriate policies and practices.

...the ESRC should aim to maximise the use of research data that it funds for public benefit, and that it exercises a key role in building public trust in research data use, by promoting those benefits

To this end, the ESRC commissioned a team from the University of Dundee to conduct an independent review of its research data policy and provide recommendations and insights on what changes may be needed. This report is the main and final output of that review. This work took place during 2023–24 and was led by Professor Angela Daly.

The ESRC research data policy, whose most recent version dates from 2018, is "intended to support ESRC grant holders who collect, produce and re-use data by defining researchers' roles and responsibilities, as well as the roles and responsibilities of the ESRC and its data service providers". A key element of the policy since the 1990s has been the default requirement for new data created through ESRC-funded research to be deposited with UK Data Service (UKDS) or another suitable depository once the ESRC grant ends (Van den Eynden & Corti, 2017). This requirement, in our view, serves a dual purpose of recognising open social science data as a public good funded by the public purse, and of facilitating the reproducibility and replicability of research. However, if the policy is to continue to meet these twin goals, then it needs to be updated to reflect significant changes to the kinds of research data produced and used by social science researchers, as well as broader changes to the legal, policy and technical landscapes both in the UK and internationally.

To understand these changes, we first conducted a scoping review and then elicited the views of different stakeholders involved with ESRC-funded research data via an online survey, which we explored further in a series of focus groups. Respondents were asked their views about social science research data and the ESRC research data policy, including the need for change. An analysis was also conducted of Data Management Plans (DMPs) which are a required part of the application process for ESRC funding and sit alongside ethical approval and other mandatory processes for certain types of research involving data.

The report makes a series of recommendations for updating the design and implementation of the ESRC research data policy, based on the research findings from the project and also drawing on the team's background knowledge and wider engagement with the social science research data community. These recommendations stem from the view that the ESRC should aim to maximise the use of research data that it funds for public benefit, and that it exercises a key role in building public trust in research data use, by promoting those benefits. Detailed recommendations cover changes to policy content, including alignment with other relevant policies, and policy implementation and promotion. Additionally, a case is made for regular policy review, to ensure that it remains fit for purpose, and for considering further research towards strengthening longer-term policy development and updates.

1 | Introduction

The context in which the ESRC research data policy exists has changed in various ways since its most recent revision in 2018, with significant changes to the kind of research data produced and used by social science researchers, along with several other changes to the legal, policy and technical landscapes both in the UK and internationally.

This is the final report from a University of Dundee team led by Professor Angela Daly, which conducted an independent review of the ESRC research data policy between August 2023 and April 2024. The team undertook a scoping review involving identification of the broader context of research data practices, policies and legislation. This set the scene for our engagement with the ESRC research data community, which in turn led to novel empirical research consisting of two elements to elicit their views on how well the current policy is working and how it could change: an online survey and a series of discursive focus groups conducted in late 2023. Additionally, we evaluated a sample of Data Management Plans (DMPs), given the planned change to UKRI grant application procedures following the Tickell Review. The findings of the team were also informed by a series of consultation meetings with key figures in the ESRC research data community that were organised on our behalf by the ESRC under the Chatham House rule. This final report was written in early 2024.

The team comprised a range of expertise across both ESRC and non-ESRC disciplines, and included both research data creators and research data users. Led by Angela Daly, a socio-legal scholar of the regulation of data and digital technologies, the team comprised: Paul Allanson (economics); Alistair Geddes (geography);

Maeve Malone (law); Niamh Nic Daeid (forensic science); Lucille Tetley-Brown (sociology). The team was supported by an administrator, Awais Elahi.

Our work focussed on social science research in the UK funded by the ESRC. This has led to certain limitations of scope. We have not explicitly considered social science data funded by other UKRI councils or other funders in the UK, or not funded at all. Furthermore, cross-border research and data sharing – including through UKRI/ESRC joint programmes of funding with counterparts in other countries and the UK's recent association with the Horizon Europe scheme – is beyond the scope of our work, but our analysis is informed by events including the UK leaving the European Union (EU) and the possibility for divergence in data protection standards that may ensue. The implications for such cross-border research and data sharing need further work and consideration. Another issue beyond our scope is the economic value attributed to data and datasets. Ownership of data is an associated issue also beyond our scope, and one which encompasses a range of complexities (Boyd, 2017).

The next section of the report presents our scoping work, giving some context and background to ESRC-funded social science data in the UK. Then Section 3 contains the results and findings from our three-pronged empirical research: DMP analysis; survey; and focus groups. Section 4 presents our recommendations for the ESRC to implement in its next iteration of the research data policy and in other relevant initiatives. Finally, we offer a conclusion with some suggestions for future work.
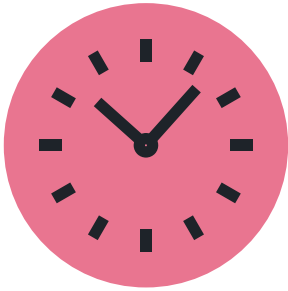
2 | A Changing Context

The context in which the ESRC research data policy exists has changed in various ways since its most recent revision in 2018. These include as regards data practices, law and policy.

The last ten years have seen major developments in the kinds of data, digital tools, methods and related outputs like code and software that social science researchers produce, share and use

2.1 Data Practices

The last ten years have seen major developments in the kinds of data, digital tools, methods and related outputs like code and software that social science researchers produce, share and use, as identified in another recent study commissioned by the ESRC from the Software Sustainability Institute (SSI) on software and data used in social sciences (Aragon et al., 2023). There have also been new ways of creating and sharing data and code, and new kinds of data such as synthetic data are increasingly prominent. The use of data to train machine learning (ML) models for artificial intelligence (AI) is another significant development.

Research data are defined in the current ESRC research data policy as "information relevant to, or of interest to researchers, either as inputs into or outputs from research". New and emerging forms of data leading to the generation of novel research-relevant digital objects may need to be deposited as part of the ESRC funding requirement. Accordingly, we adopted a similarly broad interpretation of 'data', encompassing not only recorded information but also metadata, algorithms, code, software and workflows.

Nevertheless, as Aragon et al. found (2023, p. 4), 'traditional' forms of data still dominate:

> *Surveys (22%) and interviews (19%) are the most dominant forms of data used in social science research, followed by a long tail of data sources that include APIs, behavioural data, social media, human participants, new data, and questionnaires.*

Data may be categorised in many ways. Within the ESRC disciplines, there is a broad tripartite conceptualisation of research data into: data from traditional social science research such as surveys; data from the public sector arising from our interactions with public services ('administrative data'); and more recently, 'smart data' arising from our interactions with digital systems. Yet not all social science research data fits easily within this tripartite conceptualisation. In any event, different disciplinary and methodological approaches, norms and practices give rise to different issues for data and data sharing (e.g., from an anthropological perspective see Zeitlyn, 2021). Different human participant groups may give rise to different issues for data and data sharing (see, e.g., Thomson et al., 2024 on seldom heard families).

Different disciplinary and methodological approaches, norms and practices give rise to different issues for data and data sharing
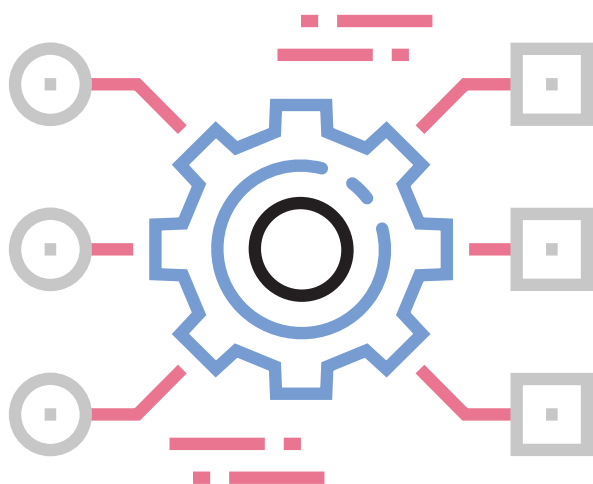
The development of synthetic data (i.e., information that is not real-world data but generated artificially using a mathematical model or algorithm) is another important recent development, which brings various opportunities and challenges for research (Heyburn et al., 2018; Dankar & Ibrahim, 2021; see also Gal & Lynskey, 2023). Data can be used in new ways, including to train AI – in particular ML models (Jefferson et al., 2022). There are many opportunities for new and innovative research, but also privacy and security concerns, among others, which emanate from these developments (see, e.g., Kerasidou et al., 2023).

Data access and sharing arrangements can present significant challenges to researchers. In particular, gaining access to administrative data can involve long and cumbersome processes. The recent independent Lievesley Review of the UK Statistics Authority (Lievesley, 2024) found barriers to data sharing by government departments (including from a lack of resource prioritisation, compared to other departmental functions, to make data research ready), and recommended that government departments "prioritise data sharing for statistics and research purposes" (p.10) and that the UK Statistics Authority improve its partnerships with universities, think tanks, business and the private sector (p.13). Accessing smart data or data from digital infrastructures, which are often commercial in nature, can also present challenges for researchers due to issues such as costs and licensing, which is often bilateral between the data holder and one research organisation, thereby limiting or preventing researchers in other organisations from accessing and using that data without the negotiation of another bilateral licence. This leads to issues down the line for the sharing of new data generated by research using such datasets. While the UK Intellectual Property Office has devised the Lambert Toolkit to assist with university-business collaborations, issues still exist, involving negotiating and obtaining licences to use such data, restrictions on publishing and sharing, and obstacles relating to data protection and commercial confidentiality. These issues are not confined to the UK, being experienced by researchers in other countries too (e.g., the Netherlands – see Institute for Information Law, 2023).

There are also new modes and cultures of sharing associated with some novel types of data and outputs, compared to the ESRC convention of data deposit with UK Data Service (UKDS). For instance, ML models are often deposited in open commercial repositories such as GitHub. This raises issues for identifying the original datasets on which, e.g., synthetic data and ML models are based, and how best to ensure reusability, replicability and findability of data and its derivatives in ways which preserve public access to publicly funded data.

2.2 Law

One prominent distinction in law is between personal data and non-personal data, because the former triggers the application of data protection law. The largest legal change over the last ten years of the ESRC research data policy's lifetime has been the update to data protection law, governing personal data, in the form of the EU's General Data Protection Regulation (GDPR). The GDPR was adopted before the UK formally left the EU ('Brexit') and remains part of UK law at the time of writing despite Brexit ('UK GDPR'), along with the Data Protection Act 2018 (henceforth, 'UK data protection law').

Data protection law is highly relevant to research data in the social sciences as a considerable proportion will involve data from living human beings classified as personal data. In such cases, UK data protection law will apply. If there is no personal data in the research data (e.g., data from deceased people), then the law will not apply. If the data are truly anonymous, then UK data protection law also does not apply. UK data protection law contains principles, rights, and risk-based rules around the processing of personal data, which can include activities such as data generation, use and sharing in research contexts. The Information

Commissioner's Office (ICO), the UK's data protection authority, issues **Guidance** on the research provisions of data protection law. The ICO also released **draft updated Guidance** for consultation in 2022, which remains current at the time of writing. Furthermore, the ICO has published **detailed Guidance** on determining what 'personal data' is, which can assist researchers in determining whether they are handling personal data or not.

Some research data which comes within the ESRC remit is administrative data from the UK Government and devolved nation governments. Access to administrative data involves plotting a course through multiple pieces of legislation and policies depending on data type and jurisdiction across the UK and devolved nations (see ADR UK, 2023). The Digital Economy Act 2017 includes provisions on research use of data to facilitate **"the linking and sharing of de-identified public sector data to support valuable new research insights about UK society and the economy"** which involves an approvals process to ensure projects are for the public good (see also, Goldstein, 2017). When approval is obtained, data are de-identified and made available to researchers in a secure environment subject to disclosure control.

There are also other pieces of legislation which are relevant to research data. The Human Rights Act 1998 and Equality Act 2010 implement rights to ensure that data is handled in non-discriminatory and non-infringing ways. Public sector bodies also have positive duties under the latter Act to advance equality under the Public Sector Equality Duty.

Another key area of law is intellectual property (IP) law. Data, depending on the circumstances, may be protected using trade secrets/confidentiality, copyright, and the sui generis database right in UK law. Typically, researchers and/or their institutions hold such IP rights over data they generate, and in accordance with ESRC funding rules, such data is deposited under licence in UKDS or other appropriate repository, unless an exemption has been granted. Confidentiality, especially in situations involving administrative data or commercial data for research, may preclude data deposit. As mentioned in the previous section, IP, as well as data protection, can inhibit access to third party data in the first place, posing issues for research involving text and data-mining (despite relevant yet potentially outdated exceptions to infringement in jurisdictions such as the UK – see Fiil-Flynn et al., 2022), and for access to social media data, as recognised in several recent international studies (Aufderheide et al., 2024; Morten et al., 2024). Further issues are now arising, both for data protection and IP, around the use of data for training (in particular, commercial) ML models especially when this is not explicitly included in a licence over the data.

There are several legislative changes currently under discussion in the UK which are relevant to research data. Among the most prominent is the Data Protection and Digital Information Bill, which if passed unchanged into law in the version current at the time of writing, would involve changes to definitions of terms relating to the processing of personal data, new definitions for 'research and statistical purposes', and changes to data subject rights and purpose limitation provisions. There is also a Private Member's Bill on AI, the Artificial Intelligence (Regulation) Bill, which if passed would impact on AI research, through the implementation of a set of regulatory principles (including the FAIR principles, see below) and the introduction of regulatory sandboxes.

There are a number of developments in the EU in data law and policy (see Institute for Information Law, 2023). Among them is the new Digital Services Act, which aims to open up data from certain very large online platforms and search engines to vetted researchers by providing legal mechanisms for its access (Albert, 2022) – although as some such as Iramina et al. (2023) argue, this does not go far enough to achieve its aims. Since these developments post-date Brexit, UK laws will not be aligned with them, although UK-based researchers may be able to become vetted researchers. The UK research community will nonetheless need to be aware of these developments and divergences, especially in international collaborations with EU-based researchers, and in engagement with Horizon Europe.
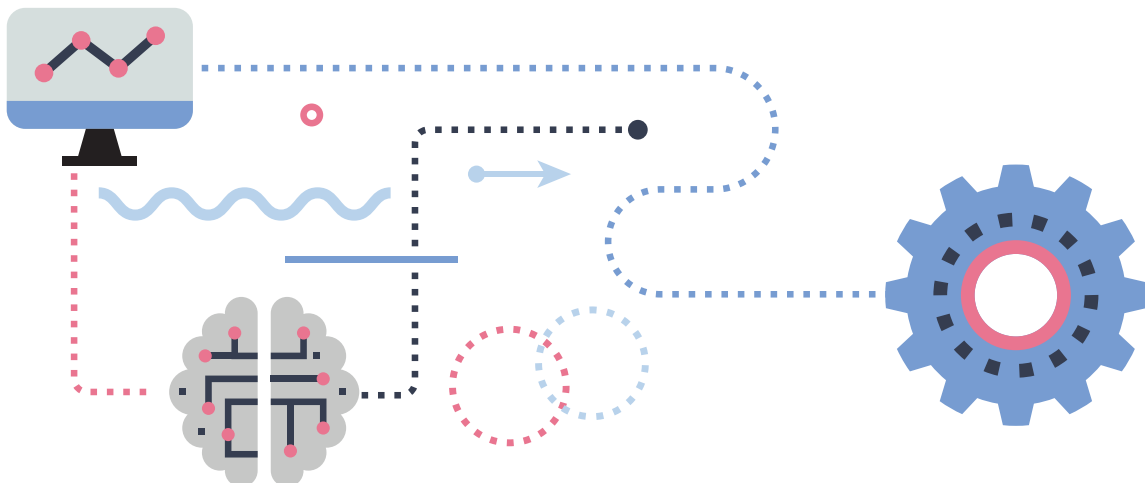
2.3 Policies

Among the UKRI research councils and associated organisations and initiatives, there is a proliferation of, and some variability among, research data policies.

Some UKRI research councils, notably the EPSRC and MRC, have recently updated their own equivalents of the ESRC research data policy, the **Principles of EPSRC research data policy framework** (2022) and the **MRC data sharing policy** (2023) respectively. Other councils, notably the AHRC, do not have a standalone research data policy. In any case, alignment across these UKRI policies as much as possible is a desirable goal. A lack of alignment or ambiguity as to which policy does or should apply can impact upon activities jointly funded by the ESRC and other UKRI bodies. An example includes the new **Population Research UK** co-ordination hub, which is jointly funded by the ESRC and MRC, and has been instructed to follow both the ESRC research data policy and the MRC data sharing policy. While there is alignment between these policies, there are differences – for instance, the MRC does not have the deposit of data as a default requirement unlike the current ESRC research data policy.

Under the auspices of the ESRC, there are also various 'sub' policies, which will apply depending on the kind of research being conducted, the sources of the data and the kind of data at stake. Among ESRC investments these include the **CDRC Data Service User Guide** and the **UK LLC Reproducible and Reusable Research Policy** and **related LLC Data Access and Acceptable Use Policy**. For Trusted Research Environments (TREs)/Safe Havens, the 'Five Safes' apply (Desai, Ritchie & Welpton, 2016).

The ESRC research data policy is part of a suite of ESRC policies and initiatives around data. Another key ESRC policy is its **data infrastructure strategy 2022-2027**, which sets out the ESRC's plans and intentions for its continued investment in "sustaining a world-leading data infrastructure" (ESRC, 2022, p. 4). Another key initiative is the **Future data services** programme whose aim is to "establish what is needed for the next generation of data services beyond 2024" (ESRC, 2023).

The current ESRC research data policy is based on the **RCUK's (now UKRI's) Common Principles**. The Common Principles are themselves based on the OECD's **Principles and Guidance for Access to Research Data from Public Funding**, which have recently been revised (OECD, 2006; 2021). This revision, in the form of **Council Recommendation on Access to Research Data from Public Funding** (OECD, 2006; updated 2021) addresses new technologies and policy developments. The OECD Recommendation covers not only research data narrowly defined, but also bespoke algorithms, workflows, models, and software (including code), which are essential for the interpretation of research outputs, and is accompanied by an analytical report including the context of the growing importance of AI and the Internet of Things. The updated Recommendation provides an overarching set of principles and policy guidance to help governments reconcile potential risks and benefits, whilst seeking to cultivate greater re-use of all types of data across and within sectors, jurisdictions, organisations, and communities. In 2022, OECD countries including the UK signed the **Declaration on Access to Research Data from Public Funding** (OECD, 2022) which further formally recognised the significance of research data accessibility for additional uses and invites the OECD to support the development of guidelines based on commonly agreed principles to facilitate optimal cost-effective access to digital research data from public funding.

A significant catalyst for the updating of the 2006 OECD Recommendation was the COVID-19 pandemic which demonstrated how a lack of harmonisation and standardisation can impede research data flows in the scientific community while also highlighting the need to respect privacy and ensure the ethical management of sensitive health data. The updated OECD Recommendation mandates, incentivises and supports actions to harmonise and streamline the responsible and timely provision and exchange of data at various levels. 'Responsible' provision of data necessitates recognition of the need for (certain) data to be secure, highlighting nuances in data sharing and a shift from the previous position of open data.

Other UK-level policies include the **UK Concordat on Open Research Data**, initiated by the UK Open Research Data Forum. The Concordat built on the legacy of the initial 2006 OECD Recommendation, the RCUK Common Principles and the *Science as an open enterprise* Royal Society (2012) report, as well as on EU policies such as the **Commission Recommendation on access to and preservation of scientific information.**

Internationally, the FAIR principles – that research should be Findable, Accessible, Interoperable and Reusable – have been developed by the global research community and stakeholders (Wilkinson et al., 2016). These principles are widely applied in research data policies and are included in the ESRC research data policy, which stipulates that they should be followed, and the Concordat on Open Research Data. The FAIR principles are accompanied by practical guidance and other resources issued by the **Go Fair Initiative**.
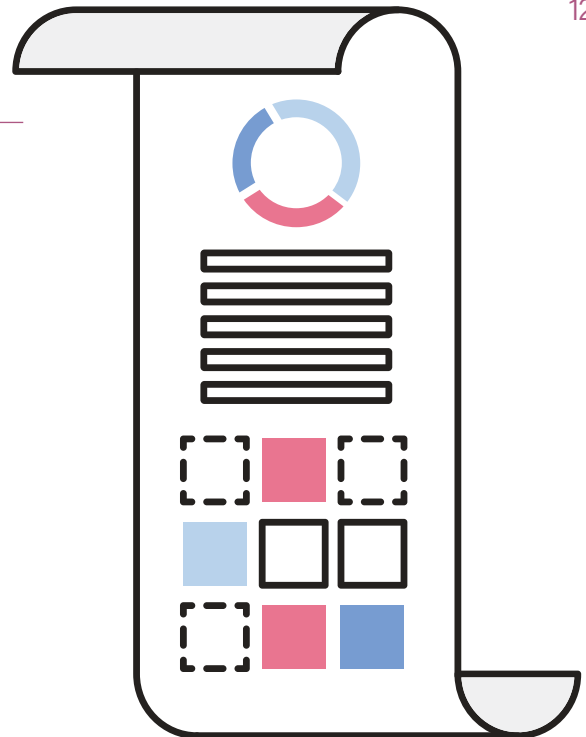
Additional principles for research data are being developed by different stakeholder groups. One prominent example is the **CARE Principles for Indigenous Data Governance**, which were created to address the gaps and tensions that prior open data and open science frameworks engendered by not fully engaging with Indigenous Peoples' rights and interests. The CARE principles comprise: Collective Benefit, Authority to Control, Responsibility and Ethics (see Jennings et al., 2023). Researchers generating data with Indigenous communities are invited to #BeFAIRandCARE.

Another set of relevant principles are the TRUST Principles, which inform the development of trustworthy data repositories, supporting efforts to make data FAIR. Encapsulating 'Transparency, Responsibility, User focus, Sustainability and Technology', the TRUST Principles provide a common framework to facilitate decision-making and implementation of best practice in digital preservation across the data sharing ecosystem, building on international data community discussion and consensus via the Research Data Alliance (Lin et al., 2020).
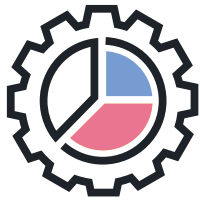
3 | Project Research and Findings

Against this backdrop of the external environment relevant to the ESRC research data policy, we conducted original empirical research between September and December 2023 on the ESRC research data policy. We did this in three parts, comprising Data Management Plan (DMP) analysis, a survey, and focus groups. This section presents each of these strands of research along with a summary of our findings.

3.1 Data Management Plan (DMP) Analysis

DMPs have been a required part of the application process for ESRC funding since 2011 and sit alongside other processes such as institutional ethics processes which may be required for certain types of data gathering. The DMP requirement is subject to change, as the ESRC and other UKRI bodies transition from the Je-S system for grant applications to the UKRI Funding Service. For DMPs, this involves a change from an uploaded document to the utilisation of a 500-word text box. Our analysis thus comes at an opportune time to reflect on how the current system of DMPs has worked, and whether the reform will preserve the benefits of current practices and address its challenges.

To do this, the project team first conducted a light-touch review of a select number of openly available DMP templates. We subsequently analysed a sample of completed DMPs which had been submitted as part of funding applications to the ESRC: 10 completed DMPs were provided to the team by the ESRC, and 9 were provided by our own institution, the University of Dundee (UoD), on a confidential and de-identified basis. This sample comprised a range of examples as regards data deposit: for some the data had been deposited, for others the data had not yet been deposited, and for others still full or partial waivers had been sought as regards data deposit. The UoD DMPs had all been created using the **Digital Curation Centre DMPonline tool**. We mapped each component of the completed DMPs onto the principles in the existing ESRC research data policy.

Overall, the sample of completed DMPs demonstrated the variety and diversity of social science disciplines and data comprised by ESRC applications and funded projects. In terms of length, a common and adequate length was 2 to 3 pages. DMPs in excess of this were often repetitive and/or included superfluous material (as elaborated below). DMPs shorter than this did not contain sufficient material to fully understand and assess them. This is significant given the proposed move to a 500-word text box for DMPs envisaged in the Funding Service, a wordcount in which the applicant is required to both demonstrate compliance with the ESRC research data policy and ESRC framework for research ethics as well as addressing other issues including legal and ethical considerations, challenges to data sharing and planning for the data lifecycle. A 2–3-page plan is a more appropriate length for the kind of material being sought, and should be formed through applicants being prompted by questions which would autofill part of the template.

Despite similar formats and prompts being used across the sample of DMPs, we found variance among these documents as to the information contained within them, in terms of structure, content and detail. Some of the variation in DMPs' content reflected variation in data types to be collected and analysed, which in turn related to different disciplinary norms and methods used for research within ESRC disciplines and projects. In one DMP analysed the authors elucidated the norms around sharing data and code for their discipline in order to help or inform a DMP reviewer from another discipline.

However, some variance in DMPs may be attributable to differences in DMP authors' knowledge and understanding of certain issues relevant to DMP construction, which may reflect different levels of internal resourcing, knowledge, and expertise on data management among researchers and support staff, and in research organisations as regards DMP writing.

There are varying levels of alignment between DMP content and specific ESRC research data policy principles. For instance, principles concerning research recognition and attribution, and 'value for money' in using public funds appropriately and efficiently for research data are rarely and barely addressed in the sample of DMPs. By contrast, all engaged with certain other principles, including those around re-use of existing data before creating new data, and making data created in the grant reusable as much as possible.

In terms of differences in how DMP authors have substantively addressed certain principles, some of this naturally reflects different disciplinary norms and practices around data, but of concern would be any variance and deficit in terms of completing DMPs which stem from a lack of knowledge or resources. Some information submitted in some analysed DMPs was patchy and incomplete. The most significant gaps in knowledge in terms of inaccuracy or incompleteness concerned consent and anonymisation when dealing with human participant data, and issues around IP including co-ownership of copyright. Another issue identified in the analysis of some DMPs involving

international collaboration was a lack of knowledge and information about what would happen to data generated by the international partners – whether It would also be deposited in UKDS or other ESRC-approved repository along with data from UK parts of the study, and/or be deposited elsewhere.

It may well be that the DMPs, submitted along with an application, occur at too early a stage for a comprehensive consideration of such issues. Indeed, there is overlap between DMPs and other processes such as internal ethics processes and information/data governance, which usually follow a grant being awarded. Yet any lack of knowledge of issues such as IP and data protection should be addressed by more supportive and structural guidance about how to complete a DMP and how to handle these kinds of issues.

Furthermore, the ESRC may want to introduce a 'risk of harm' principle or point to be considered in DMPs. Risks and harms were mentioned and addressed in some of the sampled DMPs even though this was unprompted. Riskier and more harmful data sharing may, broadly speaking, be subject to more restrictions on access and re/use (e.g., in the event of the future re-use and sharing of the data for a different purpose(s) when compared to the original purpose(s)), and these concepts may aid researchers in writing their DMPs and planning and managing their research data throughout its lifecycle. Recognition of the data lifecycle, and DMPs as part of that, could lead to a 'living DMP' process, which we discuss in more detail in our recommendations.
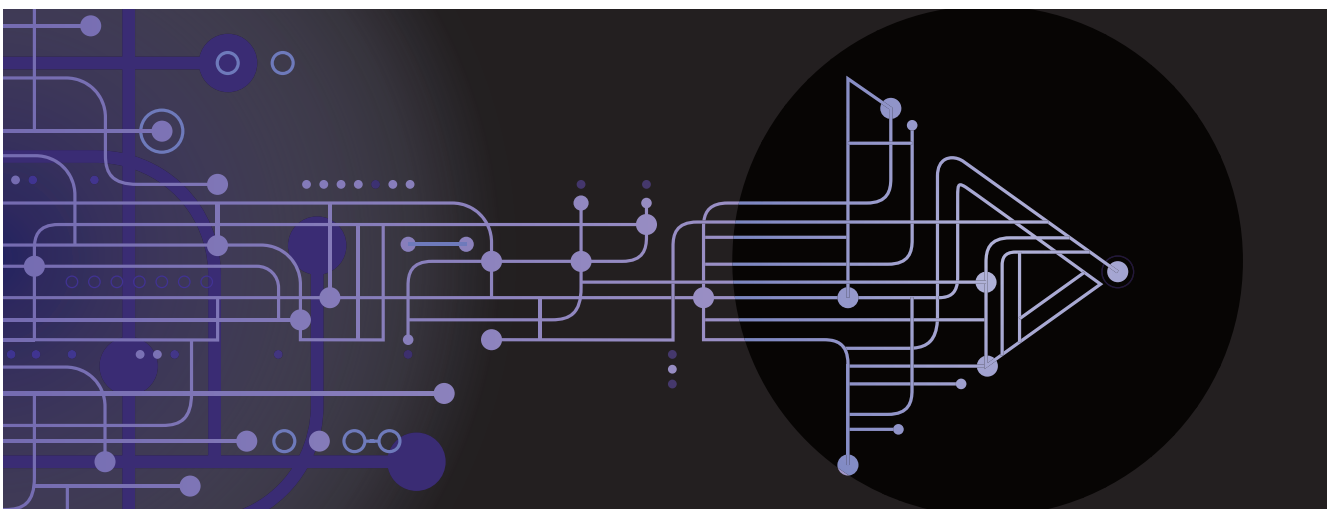
3.2 Survey

The survey was used to ask UK-based stakeholders for their views on the current design and implementation of the research data policy and how the ESRC should respond to changes in the landscape amid the growing importance of new and emerging forms of data.

The survey was approved by the University of Dundee Research Ethics Committee and administered using the JISC survey platform, with invitations distributed through a range of ESRC-provided, academic and research team contacts and networks. Following an initial pilot to refine the questionnaire, the survey was then opened for a four-week period in late 2023.

There were 135 completed responses in total, with 72 identifying as female, 113 giving their ethnicity as white, and a good spread across all age groups from 25–34-year-olds to over 65s. Approximately one-fifth (25) indicated they would be willing to be contacted again for an interview or to take part in a focus group.

The structure of the survey included a set of policy-related questions presented to all respondents plus four more specific sets of questions customised to each of the following stakeholder categories:

individual respondents based on their own experience as either research data depositors, users, or management staff, and respondents on behalf of some group or organisation. Each question set consisted of several statements inviting responses on a five-point scale ('strongly disagree' to 'strongly agree', plus 'don't know/not applicable'), followed by two further open-ended questions allowing respondents to elaborate answers and add other relevant points. Table 1 shows that there were 119 individual respondents, including 43 whom we were unable to assign to the data depositor, user or management staff categories and were therefore only asked the policy-related questions. Of these individual respondents, 110 worked in the higher education sector, with good coverage across career levels and social science disciplines. Of 84 individual respondents identifying as researchers, two thirds (56) had been awarded an ESRC grant as PI or Co-I. Most of the representatives of organisations were acting on behalf of academic groups or organisations rather than third sector or private sector ones.

Table 1. Survey completions by stakeholder category

Stakeholder Category	Description	Number of responses
Data depositors	Researchers with experience creating and depositing data associated with ESRC-funded projects	17
Data users	Other researchers with experience using ESRC-funded data and/or data services – of whom 11 had worked with controlled data in a Trusted Research Environment	32
Data management staff	Research data management and support staff	27
Representatives	Respondents on behalf of a group/organisation with a stake or interest in ESRC-funded research data issues	16
Other respondents	Respondents not elsewhere classified	43
		Total 135

There was also strong support for the proposition that the ESRC should consult regularly with the research community regarding the policy and its implementation

Main findings from the policy set of questions are as follows. Across all respondents, only a minority agreed that knowledge of the ESRC research data policy was widespread, but there was strong agreement that the policy should include the rights and responsibilities of data users, which are not separately addressed in the current version.

There was also strong support for the proposition that the ESRC should consult regularly with the research community regarding the policy and its implementation. Most respondents agreed that the aim of the ESRC should be to maximise the use of research data for public benefit, that it should improve accessibility of its data and related resources, and increase training and support to data users. Respondents also agreed that user needs should drive ESRC research data activities and investments, with strong support for prioritising investments in new and emerging forms of data and simplifying access to linked data sets, but a greater spread of views on making the provision of datasets for AI/ML training a priority. Regarding research data use, most agreed that the ESRC should play a prominent role in building public trust and promoting public benefits. Using Fisher's exact test (StataCorp, 2023), no significant differences were found between the response patterns to any of the policy-related statements across the full set of respondent categories.

Most in the data depositor category had deposited data from an ESRC grant in the UKDS-curated data depository, followed by institutional repositories, UKDS self-deposit ReShare service and then other locations. Most datasets deposited were quantitative rather than qualitative in nature – with only one depositor having deposited code – and were designated as open access, followed by safeguarded, and then controlled/secure. Most of the depositors were aware of and used the ESRC research data policy, and most agreed strongly that grant holders should be required to deposit data, with this obligation not being seen as a constraint on the

types of research undertaken. However, there was less agreement on whether grant holders should be required to deposit their software code for others to use. Most strongly agreed that they were well informed about data management issues and found DMPs to be a useful tool. While most also agreed that there was sufficient guidance on how to deposit data, fewer agreed that information on how to document data was adequate, and the majority thought that there was insufficient funding to cover documentation and deposit costs. There were mixed responses to statements about whether the period of privileged data use was adequate and that citation practices provided due recognition to depositors. Finally, most agreed that datasets on human subjects should normally include demographic/diversity characteristics.

Most in the data user category had used quantitative rather than qualitative ESRC datasets. Open-access datasets were used most frequently, but only slightly more than safeguarded datasets, with fewer using datasets designated as controlled/secure. Among those providing further dataset characteristics, survey data was cited most frequently. The majority of users agreed that relevant datasets were easy to locate, that adequate metadata and documentation were provided, and that software code used in producing research data should also be available. There were more mixed views on the adequacy of support and guidance available from ESRC-funded data depositories and providers, and the utility of online data access and analysis tools. Most agreed that conditions of data use were clear and appropriate but a specific statement that access procedures were appropriate to the sensitivity of the data elicited more mixed views. Most also agreed on the need for consistency of datasets over time and/or space, including between UK and non-UK datasets, but there was an even spread of responses on whether the comparability of data across datasets was easy to assess. Like depositors, most agreed that datasets on human subjects should normally include demographic/diversity characteristics.
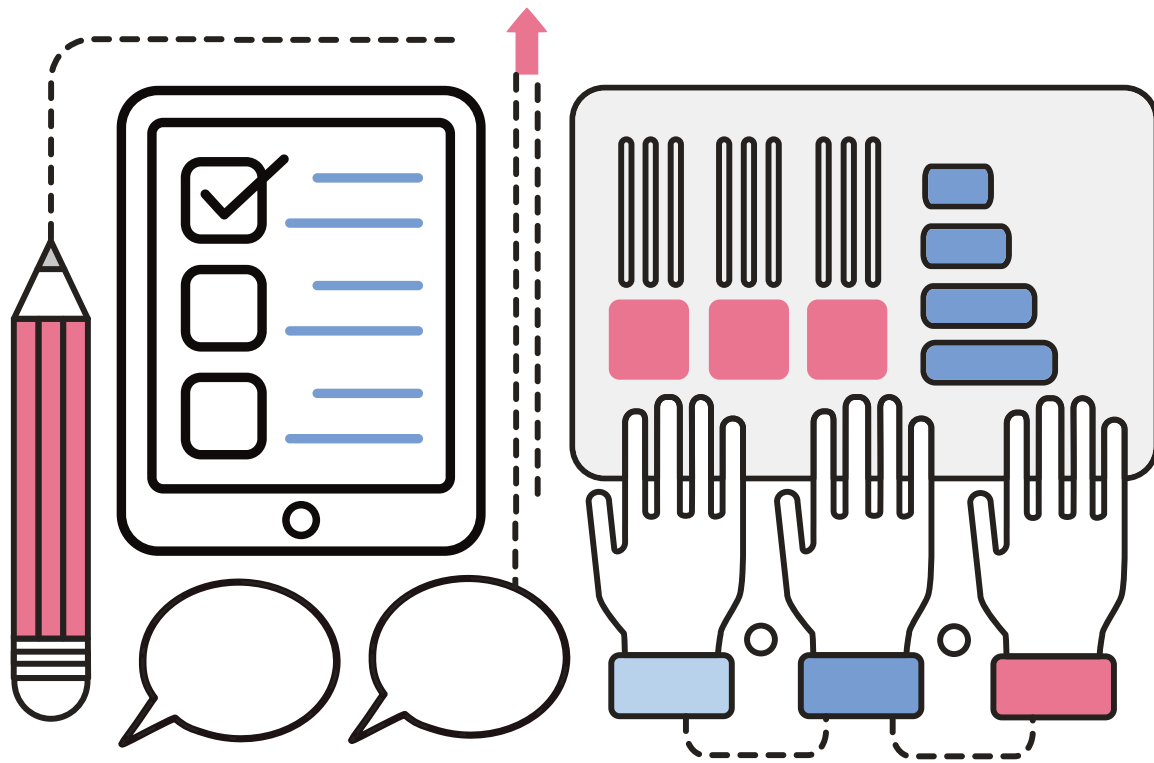
Users who had worked with controlled data in a TRE were asked an additional set of questions based on this experience. Virtually all in this subset agreed that controlled data users should be required to undergo special training with most also agreeing that access to controlled data should be subject to a public benefit test, but with responses divided as to whether the public benefit criteria were clearly defined and appropriate. Most strongly agreed that TREs prevented the disclosure of personal data. However, virtually all disagreed that working with controlled data was as straightforward as possible, with split views on the appropriateness of access approval processes, effectiveness of data import and linkage mechanisms, and the appropriateness of procedures for exporting results. Most agreed that the use of controlled data would be facilitated by the provision of synthetic data and that anonymised versions of datasets should be provided for use outside of TREs.

Most respondents in the data management staff category were research support staff, research data depository managers, managing TREs/secure environments, or engaged in research data policy information governance. Most agreed that the ESRC research data policy was easy to interpret and use, and that it was well integrated with legal and other data management requirements. Additionally, most agreed that the ESRC had appropriate requirements for research data management, including clear conditions for data deposit and use. Nevertheless, a subsequent statement on whether monitoring and compliance procedures were fit for purpose elicited a spread of responses. There were also mixed responses on whether adequate support was available for research data management staff, with a majority disagreeing that funding for research data management was adequate. Nevertheless, most agreed that research data management staff were able to provide adequate guidance to researchers, with an even spread of responses on whether researchers are well informed about data management issues.

The statements put to representatives of groups and organisations were largely drawn from those addressed to data depositors, users and management staff. The majority agreed that the research data policy was easy to locate but disagreed that researchers are well informed about research data policy issues. Virtually all agreed that ESRC grant holders should be required to deposit both data and software code, but they gave more mixed responses compared to data depositors on statements on whether this might act as a constraint on certain types of research and the adequacy of funding to cover deposition costs. In contrast, their responses were largely similar to those of data users regarding the adequacy of support and guidance available to research data users, but with more agreement as to the appropriateness of access conditions. Responses were also similar to those of data management staff regarding ease of interpretation and use of the ESRC research data policy and its integration with legal and other data management requirements, and on data deposit and use conditions. However, a larger proportion disagreed about the appropriateness of research data management requirements, the adequacy of research data management funding and about support available. Finally, most agreed that the ESRC research data policy should take account of equality, diversity and inclusion (EDI) issues.

The open-ended questions were answered by varying proportions of respondents in each stakeholder category: full response was not anticipated given the greater effort required to formulate and write in answers. Several free text responses in the policy-related question set recognised that what the ESRC could do was necessarily limited by budget constraints. Comments from data depositors noted the 'burden' of deposit (including due to insufficient time, funding, procedures or support), questioned the necessity of the deposit requirement and raised concerns over how data may be used subsequently. The subset of controlled data users gave proportionately the greatest number of free-text responses, often expanding on the difficulties with accessing, using and working with data in TREs, and offering some suggestions on how to remedy them. Similar data deposit issues were also raised by data management staff, with other comments from these respondents echoing the perceived need for more support and the difficulties with the findability of datasets within the increasingly diverse landscape of ESRC data investments. Comments from both representatives and other respondents generally dealt with similar themes to those from other stakeholder categories, including data deposit requirements and procedures, data access, and support to researchers. Other respondents also made comments about how to maximise returns to the UK from the investments being made in ESRC research data and the need for UKRI to uphold academic freedom.
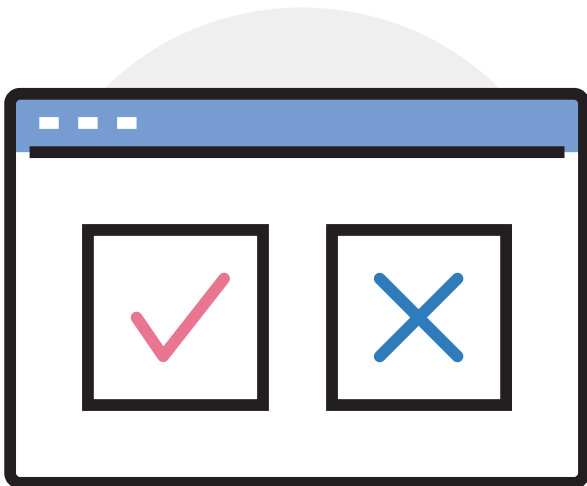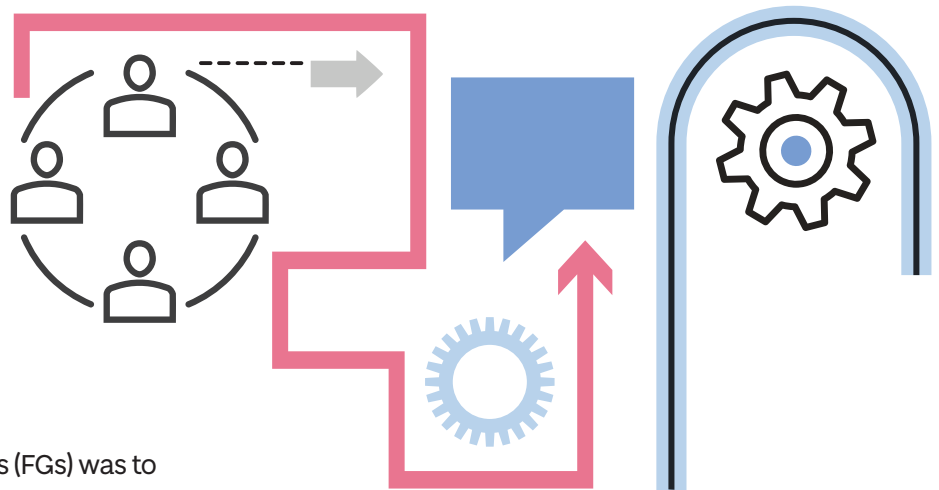
3.3 Focus Groups

The primary aim of the focus groups (FGs) was to explore awareness of the ESRC research data policy and hear respondents' views on the policy as a means of facilitating data sharing in the context of social science research data practices, in ways which delved deeper than the survey. Four FGs were conducted in late 2023, with each FG session lasting 90 minutes and designed to enable a facilitated discussion on a specific theme – data management; sharing data; open social science, replicability and reproducibility; and data as a public good – but with scope to address other relevant topics. While the team had prepared facilitative questions beforehand, the format was discursive in practice. All the FGs were conducted online using the Microsoft Teams platform.

The team recruited UK-based attendees from key social science research data stakeholder groups, with the majority of participants being research data support staff from within research organisations, including higher education institutions, the ESRC investments and other ESRC-aligned entities (e.g., ADR UK, HDR UK). There were 29 participants in total across the four FGs, of whom 55% were research support staff; 18% researchers and analysts; and the remainder were employed in various strategic data roles, such as related to institutional governance, engagement and open science culture. There was attendance from all four nations of the United Kingdom.

The data from each FG were analysed using reflexive thematic analysis techniques from Braun and Clarke (2006; see also Clarke & Braun, 2017). This data comprised preparatory documentation (protocol and topic guide), transcripts, original audio recordings and notes taken by the team, and messages from participants in the Microsoft Team chat function. Transcripts from all four FGs were analysed altogether as a whole, while ensuring sensitivity to each theme.

We identified several overarching themes that are discussed separately below: data management planning; researcher roles and academic culture; and data's openness and closedness. Finally, we discuss several other issues that emerged during the FGs.

3.3.1 Data management planning

There was a strong sentiment across participants in all FGs as regards the foundational requirements of data management, both in general and specifically in relation to the DMPs required of ESRC grant applicants. There was a prevailing view that a lot of work had gone into supporting researchers on the need for good data management, as a funder necessity but also as a core component of good data practice to support the re-use of research data consistent with the principles of open science. There was also discernible concern that the proposed reduction in the length of DMPs would undo progress that has been made in raising the importance of good data handling procedures, where researchers reflect on their data needs and intentions in some depth at the outset of new projects. Participants saw a risk in diminishing the significance of data planning for researchers applying for or beginning an ESRC-funded project.

Comments were made about perceived unsatisfactory levels of engagement with the data management processes at the early or pre-project phases by researchers who prioritise getting new projects underway. The prevalent culture was perceived to be one involving little exploration of existing data sources during preparatory work to evaluate what data might already be available. The time periods for obtaining access to existing data sources that may be applicable for a new project were cited as a major barrier against greater data re-use on the demand side of data sharing. There was a sense among some FG participants that the constraints of funding calls and grant periods, including the lead-in time, often precluded scope to explore wider data sources (repositories, catalogues, etc.) for data that could be drawn in to new projects. Frequently time pressures resulted in an emphasis on putting together a grant application with justification for novel data collection.

Participants called for a marketing campaign to promote the secondary use of datasets, which could be aligned to the ESRC's secondary data analysis initiative. However, it was acknowledged that there is still an issue about the status of data re-use projects, i.e., where 'old' data is used: the landscape is shifting but frameworks and awareness of data resources and apposite methodologies have still to catch up.

Participants also noted that when new data is created, there are minimal time and resources allocated to preparing new datasets for sharing, to enable further (onward, new) usage. Weaknesses in data management planning were therefore seen to exist across both pre- and post-project phases, perhaps reflecting a lack of both capacity and capability. However, several participants stated that the need for timely planning to ensure good management of research data is starting to be more widely accepted across the social science research community. There was a strong view that the decision to reduce the length of DMPs may result in a return to perfunctory data statements, more akin to box-ticking, and thus unhelpful for good research data practices and sharing.

It is particularly noteworthy that in every FG, the topic of DMP changes arose organically without a direct question or any prompt on this issue, which suggests the proposed changes are causing consternation, at least among research support staff who constituted the majority of participants.

3.3.2 Researcher roles and academic culture

Another pertinent issue that arose across all FGs was the extent of the 'ask' of researchers given limitations on both their capacity and capabilities. This point was highlighted particularly in relation to additional outputs from research, such as those which may help meet wider societal demands. The increasing requirement to translate research findings into layperson terms was a key theme, with a sense among FG participants that responsibility for this task should not automatically fall on the researcher alone but be shared among a wider team. In relation to research data, the point was made that there are wider data sharing ecosystems with many roles, responsibilities, and interests inclusive of, and beyond, research and researchers, such as work relating to data management and public engagement. There was a call for more clarity about the demarcation of roles and greater joined-up working.

Core incentives and rewards systems came up across all FGs. Participants discussed the fundamental culture and structure of an academic career, whereby novel papers provide the highest measure of career success/accomplishment. The significance and profile of verification or confirmatory research is viewed as of lower value and interest (e.g., by publishers and the academic community) than new findings, which aligns to wider issues of research replicability and reproducibility. Participants repeatedly made the point about the status and gravitas (or lack thereof) of data re-use and data use adjacent work, linking to themes of career trajectories and what is rewarded and incentivised. This point is echoed internationally, and among non-social science disciplines too, in the form of one of the 'key takeaways' the 2023 State of Open Data report (Digital Science et al., 2023, p. 5), namely "the perception that [researchers] don't receive sufficient recognition for openly sharing their data". Furthermore, Pownall et al. (2021, p. 530) identify how not only is there insufficient recognition, but that labour-intensive work on data is "largely unrewarded and often falls to ECRs [early career researchers] and other minoritized groups".

Another topic that arose in all FGs was that of the quality of data and its adequacy for varied contexts of use. This issue linked to intended uses of data, and in particular there was discussion about data underuse. The challenges of catalysing and encouraging increased re-use of data, i.e., looking for and developing projects that draw in secondary data, were a major topic of discussion, as noted above. A reason provided as to why this is less popular was, for the FG participants, the culture of academia itself where there is more 'credence' given to primary and new/novel data and associated findings based on that data.

Moreover, additional data-related outputs are not given due recognition: publishing one's data is still not usual let alone standard practice, and nor is the application of persistent identifiers (PIDs) to datasets and citation of these by subsequent users.

Nevertheless, some respondents noted that usage of PIDs, such as digital object identifiers (DOIs) and ORCIDs, was becoming more common in accordance with FAIR principles. Also highlighted were the risks from word sense disambiguation (contextual meaning of terminology and dataset attributes, via, e.g., application of natural language programming) and 'link rot' during digital data preservation – a problem also recently highlighted by Eve (2024).

3.3.3 Data's openness and closedness

An interesting sentiment across the FGs was that of the limitations of the perceived binary view of data (tangibly, as discrete and named datasets) as alternatively 'open' or 'closed'. Participants discussed the difficulties of navigating the more complex reality and lamented what was seen as a prevailing view of data needing to be categorised as either one or the other, with the suggestion of the need for a scale to accommodate different 'tiers of sensitivity'. There was a sense that within data sharing efforts, being pushed towards the open-closed dichotomy overlooked the nuances of data sharing and unavoidable limitations. Comments were made that the lack of nuance, including around accommodating disciplinary norms, could lead to frustration and disconnection, or even disillusionment around data sharing work. In fact, data openness does exist in a continuum, but this was not always apparent in the FG discussions, pointing to confusion on how this is currently done.

Relatedly, participants in all FGs commented on the tensions that stem from the two directions (open it up versus keep it safe and secure). The agency and empowerment (or rather lack thereof) tied to accounts of uncertainty and confusion, where respondents reported feeling pulled in opposite directions by different needs (open or closed). These tensions point to the need for more clarity on the spectrum of data sharing and where opportunities and burdens fall for greater data re-use.

An approach to improvement needs sensitivity to the nuances and requires provision of bespoke support to fully realise the opportunities

3.3.4 Other issues

Having an overall aim of increasing social science research data sharing to accomplish greater data use was supported by FG participants, however the breadth of data work means there is an appetite for greater guidance and clarity. There was a prevailing view that whilst some standardisation is appropriate, taking a discipline-agnostic blanket approach would not necessarily bring greater clarity. By keeping terminology broad and given participants' lack of awareness of provided examples or best-practice cases, the current perceived high level one-size-fits-all research data policy is insufficient by itself and requires further interpretation and application via a suite of guidance materials across the expansive range of data types, sources, and usage purposes in ESRC disciplines.

More guidance is directly sought from the ESRC, ideally via real examples and case studies specific to particular disciplines and data work areas (quantitative, qualitative, archival, personal data, dataset linkage, etc.). Furthermore, there was a view of the need for both clear top-down guidance and 'global' best-practice expectations, alongside discipline-sensitive operationalisation matching data handling needs across different social science research areas and types.

Among FG participants, there was admission of confusion and uncertainty about what is expected and how to conform to the requirements of the ESRC research data policy especially the deposit requirement. This is a suitable area for further research, for instance, following up if, when, and how data does not end up getting deposited after the closure of a research project, and reasoning for non-applicability of the deposit requirement. Otherwise, there is risk of undoing or undermining work and progress to date that has sought to encourage research data sharing and expand the scope for re-use. Mandates from ESRC to grant recipients to deposit all admissible research data without a strong enabling environment as a foundation may reduce likelihood of achieving aims of greater data sharing and increased use. Participants in the FGs referred to unilateral decisions by researchers about their project data (or parts thereof) being inadmissible for sharing, and a lack of oversight to 'audit' or check data handling and management during and after projects. The entities (such as research organisations, other sectors, support networks, and global communities like the Research Data Alliance) that are part of providing the enabling environment and foundation, constituting the broad ecosystem of support, could be more clearly articulated and better resourced.

Overall, there is as one FG participant put it, "a cocktail of challenges" to enable more effective and thriving social science research data use. There is clear positivity about the potential benefits, but an on-the-ground operational realism was also evident across the FG participants, who were all accustomed to data work, in a range of ways (as researchers, as support staff, as technical staff). A prevailing view at the heart of the discussion in the FGs was that there is not only one way of sharing data. Neither is there a singular data form or type within or across social science. An approach to improvement needs sensitivity to the nuances and requires provision of bespoke support to fully realise the opportunities.
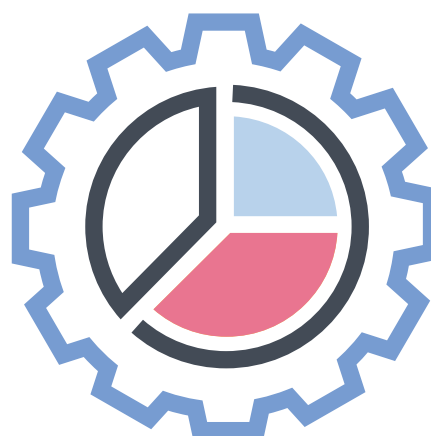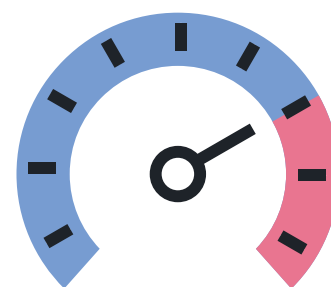
3.4 Summary

Across many of those we consulted, there is a view that the current ESRC research data policy has got various things right. However, there are certain ambiguities and tensions, which, as we move further forward into a more complex and dynamic data environment, are only going to be become more significant, unless measures are taken to update the policy with respect to clarifying roles, rights and responsibilities. An overarching theme emerges from the three strands of empirical work: confusion and conflict of purpose across the research community in serving the goals of open science and enabling onward uses of research data because messaging about roles, rights and associated responsibilities are unclear or opposed.

Whilst tiers of access and 'openness' of research data exist, the expectancy of open, viewed as the best outcome (i.e., data unencumbered for forward uses), is impacting researchers whose data cannot be made open, and who feel uncertain about the determination of appropriate onward, novel uses of their data.

Alongside this lack of clarity, the preparation of research data for sharing is under-supported in a financial way. Those with data necessarily needing to be held in a more closed way could develop data outputs pertaining to their core data, but the easier solution is to opt out and deem the deposit requirement in the ESRC research data policy as not applying in their case. It is, however, a necessary piece of good, credible research that data are deposited, even in fully closed institutional archives. The holding of research data is needed for replicability and reproducibility where feasible, as a disciplinary norm, and for the integrity of scientific results. The research community, including researchers and support staff, get pulled onto their next initiatives, and incentivisation of optimal preparation of research data from a concluded project is weak. Also, readying research data for onward use is more nuanced that simply depositing 'the data'; oftentimes projects do have data able to be shared, but equally there is an unknown scale of social science research data that remains unseen where it was not part of initial data management planning, since the parameters and specifics of a research project often evolve during delivery.

Good data management is the foundation for good data sharing

The themes congregate around the central point that data work is an interconnected 'ecosystem' of multiple actors, who have different roles (sometimes the same person will hold multiple roles), and that social science research data created, used, and shared is greatly varied in type. The variation in data type, including additional research related outputs – such as software and code, necessary to understand research findings and derive a research project's core insights from its data use – should be better recognised. The existence and updating of disciplinary norms, pertaining to good practices in specific academic fields, in a changing data landscape is a relevant component. Accommodating variance in data as a material resource, and fundamental (varied) onto-epistemologies of research itself such as is common in the social sciences, is vital to strengthen the data sharing landscape.

Many of the opportunities for the ESRC link back to data management, including the tool of DMPs. The maturity of the research community in relation to data, especially following digital data transformations over the past decade, which has seen the arrival of expanded data-linked outputs as well as new methodologies, necessitates more emphasis on the opportunities from good data management. Good data management is the foundation for good data sharing and future uses, which are appropriate and conforming to ethical and legal standards. DMPs can be an effective vehicle for increasing data management capability.

4 | Recommendations

The project team were tasked with making recommendations to the ESRC to support the development and implementation of an updated ESRC research data policy in alignment with other relevant ESRC and UKRI strategies and policies. The following recommendations for updating the ESRC research data policy draw on the review findings presented in the preceding pages and are also informed by consultation meetings with the ESRC itself, its investments and with other stakeholders.

Recommendations are grouped under the following four themes:

1   **Aims and guiding principles** for an updated ESRC research data policy, to enhance its coverage of roles, responsibilities and rights of all relevant parties, as well as its forward applicability, at least over near-term timescales.

2   **Means for implementing** policy developments, through the terms and conditions that the ESRC sets for funded researchers and research investments, and through guidance and advice it issues both to the aforementioned groups and to other relevant parties (e.g., research organisations and research support officers).

3   **Scope for aligning** the policy with other relevant ESRC and wider UKRI policies and developments, leveraging synergies and reducing gaps and differences where possible.

4   **Topics and issues** meriting additional consideration beyond the scope of the present study.

4.1 Guiding Principles

1 ESRC-funded research data for public benefit

We recommend that, in updating its research data policy, the ESRC is guided by a vision of maximising the use of research data that it funds for public benefit.

This emphasis on use maintains the focus on making data as accessible as possible, as articulated in the pre-existing policy. However, it also partners accessibility more explicitly with other considerations, such as enhanced awareness-raising and training (as detailed further below).

Additionally, this vision should also be informed by, and contribute to, maintaining public and research community trust in the ESRC and its activities, including upholding academic freedom.

2 Enhance data deposit

We recommend that the ESRC maintains the requirement that ESRC-funded grant holders deposit data by default, where possible. Specifically:

- While the ESRC should acknowledge privacy, confidentiality and licensing issues which may challenge data deposit, funded researchers should factor such issues into their planning and ensure as much data can be shared as possible.

- To help mainstream and normalise data citation, the ESRC should issue guidance on how to document data and mandate the usage of persistent identifiers. We recommend that such identifiers are used every time reference is made to the datasets.

- Where researchers must make available data related to publications prior to the end of a funded grant, i.e., prior to the customary data deposit stage, and the data cannot be deposited and made available on a timely basis with UKDS, we recommend that the data are made available via the researchers' own institutional repositories, in preference to third party platforms like Mendeley and Zenodo. In this way, institutional access to and maintenance of the data will be better enabled.

3 Better acknowledge tiers of sensitivity in data

We recommend that the ESRC in an updated policy better acknowledges, highlights, and facilitates understanding of the 'tiers of sensitivity' and 'access spectrum' across social science research data.

While the existing policy makes some reference to differences in data access (i.e., open, safeguarded and controlled), greater attention to these categories would assist in moving beyond an open vs closed dichotomy perceived by several of our respondents, and rather it would contribute to operationalising more effectively the vision of data being 'as open as possible, as secure as necessary'.

Towards operationalising this vision, and to ensure consistency in its operationalisation, the ESRC should also work with its investments and with researchers on situations where a more restricted, less than fully open, approach to data access may be warranted.

4 Include rights and responsibilities of data users

We recommend that the ESRC adds a new and separate section to the research data policy, addressing rights and responsibilities of data users, which are not separately addressed in the current version of the policy.

This new policy section should include, for example, data users making appropriate efforts to include citations to dataset uses, and to notify the dataset creators or licensers.

5 Clarify rights of human participants

We also recommend that in the research data policy, the ESRC more clearly acknowledges the rights of human participants in research data funded by the ESRC.

In addition, the ESRC should consider providing updated guidance on how such rights may best be given effect, in line with the **ESRC Framework for Research Ethics**.

6 Include demographic and diversity characteristics in funded data

We recommend that, where possible, demographic and diversity (covering equality, diversity and inclusion – EDI) characteristics are included in datasets of human subjects created with ESRC funding.

We recognise that including such characteristics while retaining conformity with the legal requirement of data minimisation is an issue which would also need consideration. However, incorporating them into data where they are not already included could help reveal biases and inform future (appropriate and sensitive) onwards use of the data.

A summary of demographic and diversity characteristics could also be provided in the accompanying metadata record.

4.2 Policy Implementation and Promotion

7 Enhance Data Management Plans (DMPs)

We recommend that DMPs should be retained as part of the process for applying for and holding ESRC research grant funding, as a key foundation for data deposit, access, sharing and re-use.

We also advocate scoping a re-working of DMPs to enhance their effectiveness, and to link better with the requirements associated with the new UKRI Funding Service.

Given the number of detailed DMP-related recommendations, they are set out in a separate section, 4.5, below.

8 Build awareness of the policy

Since we find that knowledge of the current policy tends to be somewhat lacking, we recommend that ESRC takes measures to better profile it, both among its key constituents and the wider world.

Increasing awareness of the policy could add to public confidence and trust in research data sharing and re-use, as well as increasing awareness of research data issues among researchers. The ESRC could partner in awareness-raising efforts with academic societies and stakeholders.

Furthermore, we recommend that the ESRC ensures that all requirements and best practice are met regarding the accessibility of the policy. In addition to EDI considerations, this includes enhancing the policy's user experience and interface.

9 Promote public benefit

We recommend that the ESRC addresses the following issues in order to carry forward successfully the visions of research data funded for the public benefit and building public trust outlined above:

- The ESRC should consult with the public and stakeholders on what the public benefit of research is, how it may be implemented in research funding decision-making, and how risks are managed.

Such consultations may be informed by cognate efforts, for example ADR's Strategy and report on the public good in the context of data (ADR UK, 2022), and the 'panels' created in some circumstances to assist public benefit decision-making processes (see, e.g., National Data Guardian, 2022; Tetley-Brown et al., 2024).

- It should work with TREs and other relevant stakeholders (such as public authorities providing administrative data) to enhance public benefit criteria for accessing and exporting data respectively within and out of controlled settings (see, e.g., Taylor et al., 2021).

- The ESRC should also explore with the relevant stakeholders the scope for augmenting use and awareness of different types of safe setting now available (i.e., virtual Safe Havens versus full Safe Haven settings) for different types and combinations of data.

10 Resource support

The ESRC should support the updated policy with enhanced provision of training and guidance available for key policy constituents: researchers (data creators and data users), research organisations, and research support staff. More specifically, we recommend that:

- The ESRC take steps to improve awareness of existing guidance that it, UKDS and UKRI and its constituent bodies provide (e.g., the Good research resource hub, UKDS Learning Hub and the MRC Regulatory Support Centre).

- New guidance should also be devised where needed, to help translate the policy into practical contexts, ideally including real examples and case studies as well as reflecting different disciplinary norms with respect to working with data.

- The ESRC should make more funding available to cover the costs of data management involved in a research grant – for example, for documenting data and preparing data for deposit at the end of the grant. This funding should not come from, e.g., research organisation overheads but instead there should be a specific research grant budget line funded by ESRC for such work and acknowledged by peer reviewers as an admissible cost.

11 Support data re-use

To maximise the returns on investments in making data available for re-use, we recommend that the ESRC continues to enhance its incentivisation, through funding opportunities for data re-use research projects, such as via its secondary data analysis initiative. In our research, some participants were unaware there was such funding despite its existence for several years.

The ESRC also needs to ensure sufficient time for applicants to consider pre-existing data sources and incentivise their re-use for all its funding opportunities.

12 Increase researcher recognition

We recommend that the ESRC takes further steps to instil a culture of research data sharing and re-use for the public good by better incentivising and recognising researchers sharing data. The ESRC should work on this with UKRI and the other constituent UKRI research councils, the UK Government, devolved governments, Research England, the Scottish Funding Council, the Higher Education Funding Council for Wales and the Northern Ireland Executive, higher education institutions and other relevant stakeholders. Better researcher incentivisation and recognition could be pursued both through clearer guidance in policy frameworks such as the REF and in career trajectories such as institutional promotion processes increasing the weight associated with appropriate data sharing and publication practices.

13 Improve policy monitoring and evaluation

Monitoring and evaluating relevance and effectiveness of the policy will be important. On this point, we specifically recommend to the ESRC the following measures:

- The ESRC should review its research data policy on a regular basis, to ensure that it remains fit for purpose. It should also monitor support (guidance, training, etc.) related to the policy.

- The ESRC should also devise a clear basis for assessing and monitoring researchers' compliance with the policy, from beginning to end of funded research. Here we endorse a Software Sustainability Institute (SSI) report recommendation to focus on "how good [i.e., policy-compliant] practice can be increased and incentivised" (Aragon et al., 2023, p. 6). The ESRC should also consider sanctions for non-compliance with the policy.

4.3 Scope for Alignment

14 General

We recommend that the ESRC consider alignment across the following topics:

- **Data concepts and terminology** – towards enhancing levels of awareness and knowledge within a diverse and fast-changing data landscape, we recommend that ESRC and its investments lead on producing a common set of data-related terminology. This terminology should also consider that adopted by other UKRI research councils (e.g., MRC) and other key data-producing and data-holding bodies (e.g., Office for National Statistics and government departments).

- **Up-to-date laws:** we recommend that the ESRC aligns the policy and other relevant documents with current applicable laws, where this has not already been done. For instance, the ESRC should update the **ESRC Data requirements** to remove references to repealed sections of legislation, e.g., the Data Protection Act 1998 sections 7 and 33.

- **Metadata** – we recommend that the ESRC and its investments explore options for increased harmonisation and standardisation on metadata, such as through a federated discovery approach, also informed by and aligned to leading principles, standards and best practice. We include with this consideration of metadata for software used to generate and/or analyse datasets, as per the SSI's recommendation (Aragon et al., 2023).

- **User-driven** – we recommend that the ESRC pursues better alignment with users' needs in its investment in new and emerging forms of data and in supporting data access, through more user engagement.

- **Equality, Diversity and Inclusion (EDI), sustainability and net zero** – we recommend that development of the updated ESRC research data policy should be consistent with relevant UKRI and ESRC policies in the aforementioned areas. The ESRC may need to engage in further stakeholder discussion so that the updated policy upholds EDI, sustainability and net zero considerations among the research data community.

15 Role of UK Data Service (UKDS)

We recommend that UKRI and ESRC promote UKDS as the default repository of social science research data in the UK, including data funded by ESRC or other funders – recognising that research funded by non ESRC funding bodies may have a social science element, especially research funded by UKRI research councils such as AHRC, MRC or NERC.

Social science research data should be offered for deposit in UKDS, which will actively refer researchers to alternative ESRC data infrastructures or other repositories when such alternatives better suit the data's specific needs as per the **UKDS Collections Development Policy**.

The ESRC and UKRI should ensure adequate resourcing is provided to UKDS so that it remains capable of providing sufficient input and expertise on both conventional and newer and emerging types of data and related deposit situations – the latter ranging potentially widely, e.g., ethical archiving of data co-produced with vulnerable groups, to handling synthetic data and ML models.

16 Coordinate with recent data policy developments within UKRI

We recommend that UKRI and its constituent research councils including the ESRC, and other relevant bodies work towards achieving the goal of maximum coordination, alignment and cross-referencing with each other's policies on research data.

Within the wider UKRI landscape, there are particular opportunities for aligning the ESRC research data policy with the MRC's refreshed data sharing policy, published in 2023. As such, we recommend that the ESRC follow up on opportunities to learn from the MRC on the processes involved in the production, implementation and monitoring of research data policy changes.

4.4 Further Work

17 Legal and ethical aspects of combining datasets

We recommend that the ESRC and UKRI commission more cross-disciplinary research into legal and ethical issues associated with combining different datasets involving human participants, including synthetic data derived from such data, and whether and how confidentiality and data protection can best be preserved for such participants, in ways that facilitate high quality and appropriate social science research. This research should also include a consideration of the compatibility of collecting EDI data with the data protection requirement of data minimisation to facilitate the implementation of Recommendation #6.

18 Guidance on specific issues

We recommend that the ESRC stands ready to initiate research into specific issues, new circumstances or ongoing developments which potentially or actually have implications for the research data policy, and to use this research as a basis for offering guidance to researchers and other stakeholders.

Such circumstances include developments in AI research in the social sciences, and compliance with changing data protection law (see, e.g., Boyd et al., 2018). Other circumstances include synthetic data, on which the ESRC could draw on **ADR UK's interim position statement on synthetic data** plus the results of **research it has funded on synthetic data**.

19 Data definition and software code

We recommend that the ESRC launches consultations with stakeholders on:

- a clear definition of 'data' for the purposes of the ESRC's data deposit funding requirement; and
- how and where best to facilitate the appropriate deposit of software code associated with data generated by ESRC-funded research, echoing the recommendation to this effect in the recent SSI report (Aragon et al., 2023).
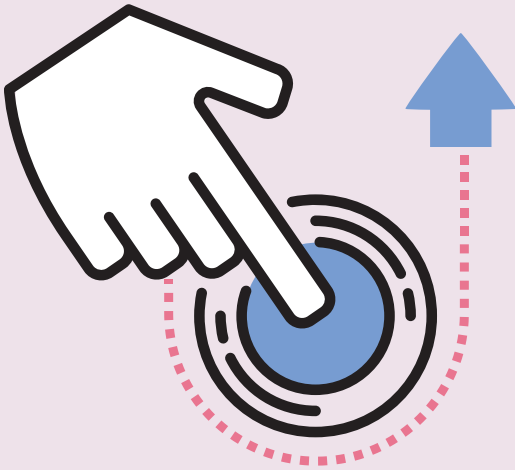
4.5 Detailing Data Management Plan (DMP) enhancement

Recommendation #7 concerns DMPs. Here we elaborate on that recommendation with some specific detail on how DMPs may be enhanced:

i. We recommend that the ESRC implements a living DMP process – of which the DMP at application stage is just the first step – which would cover the entire research data lifecycle and help ensure the DMP remains up to date at all times. In this living DMP, trigger points for reviewing and updating a DMP might include the following situations:

- when research/research data collection deviates from the original plan (e.g., when online interviews planned in one platform like Microsoft Teams are or must instead be conducted on a different platform);

- in the case of a long-term 5-year project, for example, when an opportunity to publish early arises, and underlying data for the paper must be made available for peer review and open access purposes; and/or

- when there is a new or additional purpose for research/research data collection.

The ESRC should promote the living DMP process to researchers, research data management staff and research organisations as a helpful aid for planning, including for resource allocation and preservation of data.

ii. The living DMP process should be facilitated by easily accessible (online) software tools for writing and editing DMPs, with the current standard in this regard being the **Digital Curation Centre's DMPonline tool**. We recommend that the ESRC, UKRI, UKDS and DMPonline all work together to ensure there are appropriate and sustainable tool(s) – whether an adapted DMPonline tool and/or new developments – supporting documenting of the living DMP process. In addition, such tools could provide a basis to enhance the extent to which DMPs are 'machine actionable', in accordance with the **Research Data Alliance (RDA) DMP Common Standard (RDCS)** (see Philipson et al., 2023). DMP documentation supported by appropriately designed software tools can also lead to better reproducibility, replicability and analysis, as recognised in the **Goldacre Report** (Goldacre & Morley, 2022).

iii. The DMP section in the UKRI Funding Service should be extended beyond a 500-word limit to ensure the DMP is sufficient, enabling all relevant issues for effective data management to be captured, yet without becoming overly burdensome. Based on our study, we recommend allowing up to 1500 words.

iv. There should be enhanced interoperability between the Funding Service and DMPonline or subsequent DMP tool to help populate the DMP section in the Funding Service automatically through applicants being prompted by questions (reflecting the most pertinent and useful aspects of the ESRC research data policy), guiding them to add the issues and content relevant to their data management planning. Some institutions' modified versions of DMPonline already do this, but not all.

v. The ESRC should consider the introduction of another principle in the research data policy and pertinent to DMPs, on risk of harm, which should be considered in a broad sense, including risk and harm to research participants, collaborators, and society at large. Riskier and more harmful data may, broadly speaking, be subject to extra restrictions on access and (re)use. These concepts may aid researchers in writing their DMPs and planning and managing their research data throughout its lifecycle.

vi. The ESRC, along with other relevant stakeholders such as UKDS, should enhance its guidance to researchers and support staff on the rationale and process of producing a DMP, to address deficits of knowledge and resources, including but not limited to topics of consent, lawful basis, anonymisation, IP and international collaboration during data deposit and use. The ESRC and relevant stakeholders should also work together to align the structure and content of DMPs templates, including embedded guidance.
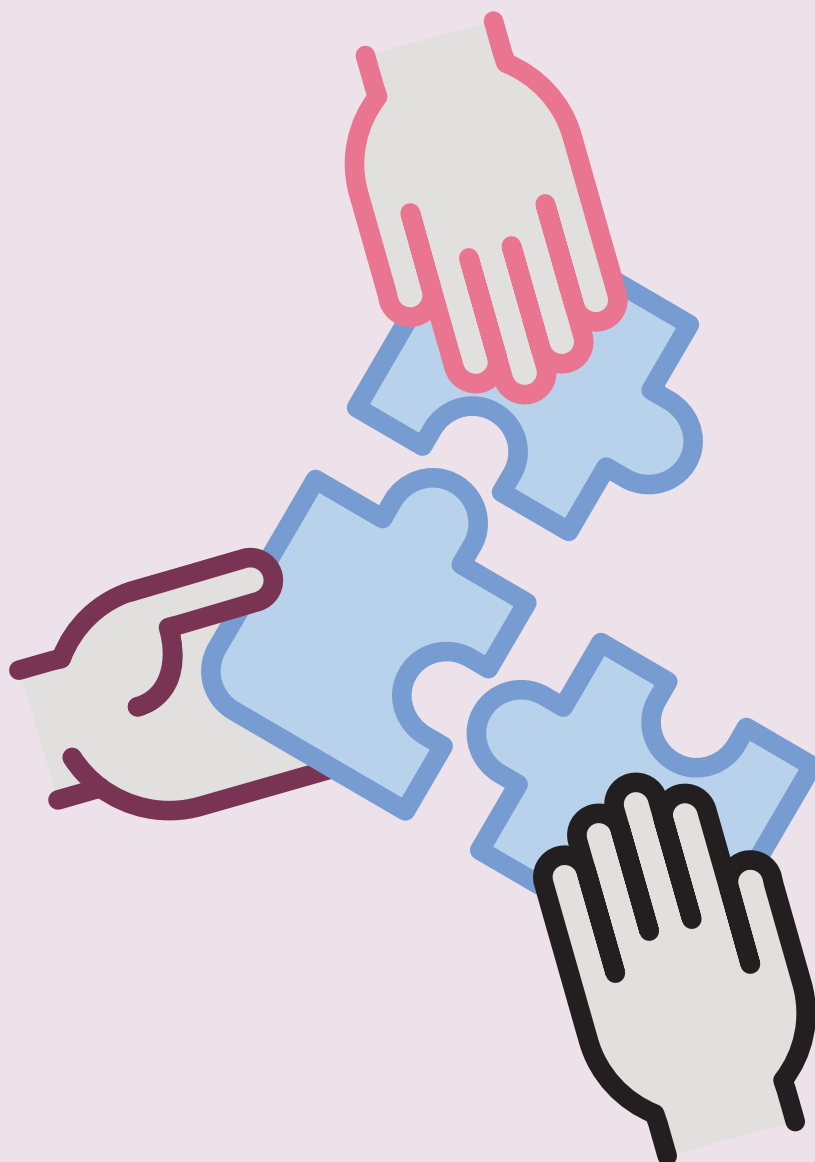
Conclusion

This report has considered the ESRC research data policy within its broader context, recent developments and through the generation of new empirical research on the policy and social science research data in the UK. The latter was undertaken through DMP analysis, a survey and focus groups, alongside engagement with ESRC stakeholders. We have listed a series of recommendations, mainly for the ESRC, to strengthen its policy and practices, ensuring the social science research data that it funds is appropriately accessible as an open public good, and for the purposes of reproducibility and replicability.

We advocate the maximum coherence among different UKRI funding councils as regards research data, and consider that a move towards policies for different types of research data or contexts is considered by UKRI, rather than funder-specific research data policy documents – where they exist – as is currently the case. Social science research in interdisciplinary projects may be funded by the EPSRC and the MRC, with their own, separate policies, or the AHRC which does not have a bespoke policy. Maximum coherence should also be sought with other UK funding bodies as regards social science research data management, planning and sharing.

Further work is required on the topics mentioned in Section 4.4 above, alongside additional topics beyond the scope of our inquiry, notably international and cross-border data issues and the monetisation of research data. This work should be led by UKRI, and its findings and recommendations shared widely with the research community in the UK.

References

ADR UK. (2022). *A UK-wide public dialogue exploring what the public perceive as 'public good' use of data for research and statistics.*
https://www.adruk.org/news-publications/publications-reports/public-perceptions-of-public-good-a-dialogue/

ADR UK. (2023). *The legal framework for using administrative data for research purposes.*
https://www.adruk.org/fileadmin/uploads/adruk/Documents/The_legal_framework_for_accessing_data_April_2023.pdf

Albert, J. (2022). *A guide to the EU's new rules for researcher access to platform data.* Algorithm Watch.
https://algorithmwatch.org/en/dsa-data-access-explained/

Aragon, S., Antonioletti, M., Walker, J., & Chue Hong, N. (2023). *Understanding the software and data used in the social sciences: A study for the Economics and Social Research Council.* **https://doi.org/10.5281/zenodo.7785707**

Aufderheide, P., Butler, B., & Anastácio, K. (2024) The Chilling Effects of Obstacles to Accessing, Using, and Sharing In-copyright Data for Quantitative Research. *Information & Culture*, forthcoming. **https://dx.doi.org/10.2139/ssrn.4662397**

Boyd, A. (2017). *Annex H: Data ownership in the UK.* Longitudinal Studies Strategic Review: Annexes.
https://www.ukri.org/wp-content/uploads/2022/03/ESRC-020222-Longitudinal-Studies-Strategic-Review-Annexes.pdf

Boyd, A., Woollard, M., Macleod, J., & Park, A. (2018). The destruction of the 'Windrush' disembarkation cards: a lost opportunity and the (re)emergence of Data Protection regulation as a threat to longitudinal research. *Wellcome Open Research*, 3, 112, 1-14. **https://doi.org/10.12688/wellcomeopenres.14796.1**

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
https://doi.org/10.1191/1478088706qp063oa

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, 12(3), 297-298.
https://doi.org/10.1080/17439760.2016.1262613

Dankar, F.K., & Ibrahim M. (2021). Fake It Till You Make It: Guidelines for Effective Synthetic Data Generation. *Applied Sciences*, 11(5), 2158, 1-18. **https://doi.org/10.3390/app11052158**

Desai, T., Ritchie, F., & Welpton, R. (2016). *Five Safes: designing data access for research.* University of the West of England Economic Working Paper Series 1601. **https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf**

Digital Science, Hahnel, M., Smith, G., Schoenenberger, H., Scaplehorn, N., & Day, L. (2023). *The State of Open Data* 2023. Digital Science Report. **https://doi.org/10.6084/m9.figshare.24428194.v1**

ESRC. (2022). *Data Infrastructure Strategy 2022–2027.* **https://www.ukri.org/publications/esrc-data-infrastructure-strategy/**

ESRC. (2023). *Future data services.*
https://www.ukri.org/what-we-do/browse-our-areas-of-investment-and-support/future-data-services/

Eve, M.P. (2024). Digital Scholarly Journals Are Poorly Preserved: A Study of 7 Million Articles. *Journal of Librarianship and Scholarly Communication*, 12(1), 1-23. **https://doi.org/10.31274/jlsc.16288**

Fiil-Flynn, S., Butler, B., Carroll, M., Cohen-Sasson, O., Craig, C., Guibault, L., Jaszi, P., Jütte, B.J., Katz, A., Quintais, J.P., Margoni, T., Rocha de Souza, A., Sag, M., Samberg, R., Schirru, L., Senftleben, M., Tur-Sinai, O., & Contreras J.L. (2022). Legal reform to enhance global text and data mining research. *Science*, 378, 951-953. **https://doi.org/10.1126/science.add6124**

Gal, M., & Lynskey, O. (2023). Synthetic Data: Legal Implications of the Data-Generation Revolution. *Iowa Law Review* 109, forthcoming. **http://dx.doi.org/10.2139/ssrn.4414385**

Goldacre, B & Morley, J. (2022). *Better, Broader, Safer: Using health data for research and analysis: A review commissioned by the Secretary of State for Health and Social Care*. Department of Health and Social Care. **https://assets.publishing.service.gov.uk/media/624ea0ade90e072a014d508a/goldacre-review-using-health-data-for-research-and-analysis.pdf**

Goldstein, H. (2017). Editorial: The Digital Economy Act and Statistical Research. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 180(4), 945–946. **https://doi.org/10.1111/rssa.12317**

Heyburn, R., Bond, R.R., Black, M., Mulvenna, M., Wallace, J., Rankin, D., & Cleland, B. (2018). Machine learning using synthetic and real data: Similarity of evaluation metrics for different healthcare datasets and for different algorithms. In J. Liu, J. Lu, Y. Xu, L. Martinez & E. Kerre (Eds.), *Data Science and Knowledge Engineering for Sensing Decision Support: Proceedings of the 13th International FLINS Conference* (pp. 1281-1291). World Scientific Proceedings Series on Computer Engineering and Information Science Data Science and Knowledge Engineering for Sensing Decision Support **https://www.worldscientific.com/doi/abs/10.1142/9789813273238_0160**

Institute for Information Law (2023). *Information Law and the Digital Transformation of the University. Part II. Access to Data for Research*. University of Amsterdam. **https://www.ivir.nl/part-ii-access-to-data-for-research/**

Iramina, A., Perel, M., & Elkin-Koren, N. (2023). *Paving the Way for the Right to Research Platform Data*. **http://dx.doi.org/10.2139/ssrn.4484052**

Jefferson, E., Liley, J., Malone, M., Reel, S., Crespi-Boixader, A., Kerasidou, X., Tava, F., McCarthy, A., Preen, R., Blanco-Justicia, A., Mansouri-Benssassi, E., Domingo-Ferrer, J., Beggs, J., Chuter, A., Cole, C., Ritchie, F., Daly, A., Rogers, S., & Smith, J. (2022). *GRAIMATTER Green Paper: Recommendations for disclosure control of trained Machine Learning (ML) models from Trusted Research Environments (TREs)*. **https://doi.org/10.5281/zenodo.7089491**

Jennings, L., Anderson, T., Martinez, A., Sterling, R., Chavez, D.D., Garba, I., Hudson, M., Garrison, N.A., & Russo Carroll, S. (2023). Applying the 'CARE Principles for Indigenous Data Governance' to ecology and biodiversity research. *Nature Ecology and Evolution*, 7, 1547–1551. **https://doi.org/10.1038/s41559-023-02161-2**

Kerasidou, C., Malone, M., Daly, A., & Tava, F. (2023). Machine learning models, trusted research environments and UK health data: ensuring a safe and beneficial future for AI development in healthcare. *Journal of Medical Ethics*, 49, 838-843. **https://jme.bmj.com/content/49/12/838**

Lievesley, D. (2024). *Independent Review of the UK Statistics Authority*. **https://assets.publishing.service.gov.uk/media/65ef2f733649a23451ed63a9/UKSA_Review.pdf**

Lin, D., Crabtree, J., Dillo, I. Downs, R.R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M.E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D.V., Stockhause, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data*, 7, 144, 1-5. **https://doi.org/10.1038/s41597-020-0486-7**

Morten, C., Nicholas, G., & Viljoen, S. (2024). Researcher Access to Social Media Data: Lessons from Clinical Trial Data Sharing. *Berkeley Technology Law Journal*, 39(1), 109-204. **https://btlj.org/wp-content/uploads/2024/04/0003_39-1_Morten.pdf**

National Data Guardian. (2022). *What do we mean by public benefit? Evaluating public benefit when health and adult social care data is used for purposes beyond individual care.* https://www.gov.uk/government/publications/what-do-we-mean-by-public-benefit-evaluating-public-benefit-when-health-and-adult-social-care-data-is-used-for-purposes-beyond-individual-care

OECD. (2006, updated 2021). *Council Recommendation on Access to Research Data from Public Funding.* OECD/LEGAL/0347. https://www.oecd.org/sti/recommendation-access-to-research-data-from-public-funding.htm

OECD. (2022). *Declaration on Access to Research Data from Public Funding.* OECD/LEGAL/0321. https://legalinstruments.oecd.org/public/doc/157/157.en.pdf

Philipson, J., Hasan, A., & Moa, H. (2023). Making Data Management Plans Machine Actionable: Templates and Tools. *Data Science Journal*, 22(29) 1-13. https://doi.org/10.5334/dsj-2023-029

Pownall, M., Talbot, C.V., Henschel, A., Lautarescu, A., Lloyd, K.E., Hartmann, H., Darda, K.M., Tang, K.T.Y., Carmichael-Murphy, P., & Siegel, J. A. (2021). Navigating Open Science as Early Career Feminist Researchers. *Psychology of Women Quarterly*, 45(4), 526-539. https://doi.org/10.1177/03616843211029255

Royal Society. (2012). *Science as an open enterprise.* The Royal Society Science Policy Centre report 02/12. https://royalsociety.org/-/media/policy/projects/sape/2012-06-20-saoe.pdf

StataCorp. (2023). *Stata Statistical Software: Release 18.* StataCorp LLC.

Taylor, J.A., Crowe, S., Pujol, F.E., Franklin, R.C., Feltbower, R.G., Norman, L.J., Doidge, J., Gould, D.W., & Pagel, C. (2021). The road to hell is paved with good intentions: the experience of applying for national data for linkage and suggestions for improvement. *BMJ Open*, 11(8), E047575, 1-10. https://bmjopen.bmj.com/content/11/8/e047575

Tetley-Brown, L., Sorbie, A., Miyake, E., & Daly, A. (2024). Private sector use of public sector personal data? An insider perspective on Scotland's approach. In J. Hunsinger (Ed.), *Handbook of Public Policy and the Internet* (forthcoming). Edward Elgar.

Thomson, R., Lacey, A. Nasrawy, M., Boddy, J., Morrice, L., & Brannen, J. (2024). *Scoping Longitudinal Qualitative Studies with Seldom-Heard Families: Literature Review and Academic Expert Consultation.* University of Sussex. Forthcoming.

Tickell, A. (2022). *Independent Review of Research Bureaucracy – Final Report.* https://assets.publishing.service.gov.uk/media/62e234da8fa8f5033275fc32/independent-review-research-bureaucracy-final-report.pdf

UK Data Service. (n.d.). *Digital Economy Act.* https://ukdataservice.ac.uk/learning-hub/research-data-management/data-protection/data-protection-legislation/digital-economy-act/

Van den Eynden, V., & Corti, L. (2017). Advancing research data publishing practices for the social sciences: from archive activity to empowering researchers. *International Journal on Digital Libraries*, 18, 113-121. https://doi.org/10.1007/s00799-016-0177-3
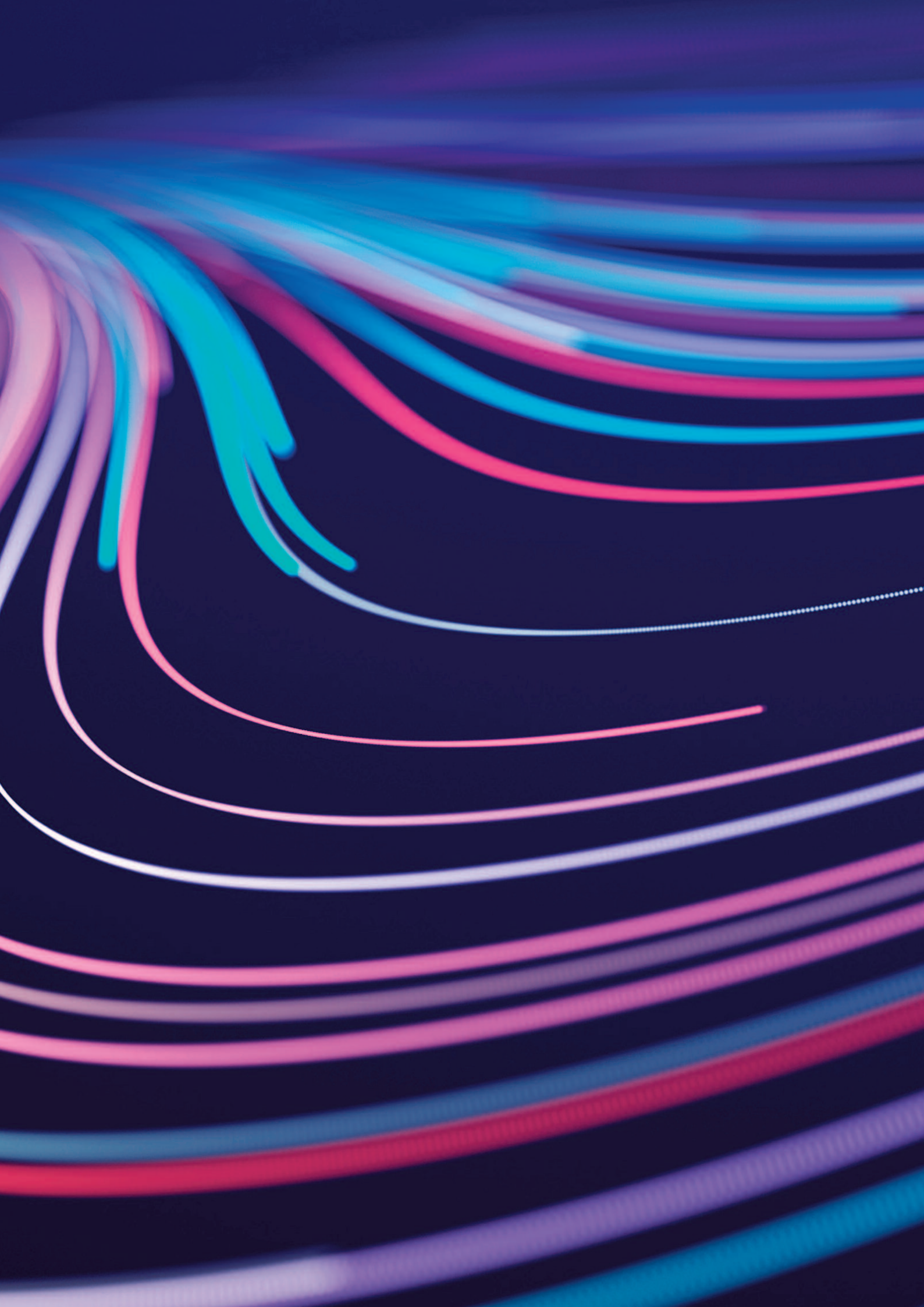
Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., Bonino da Silva Santos, L., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018, 1-9. https://doi.org/10.1038/sdata.2016.18
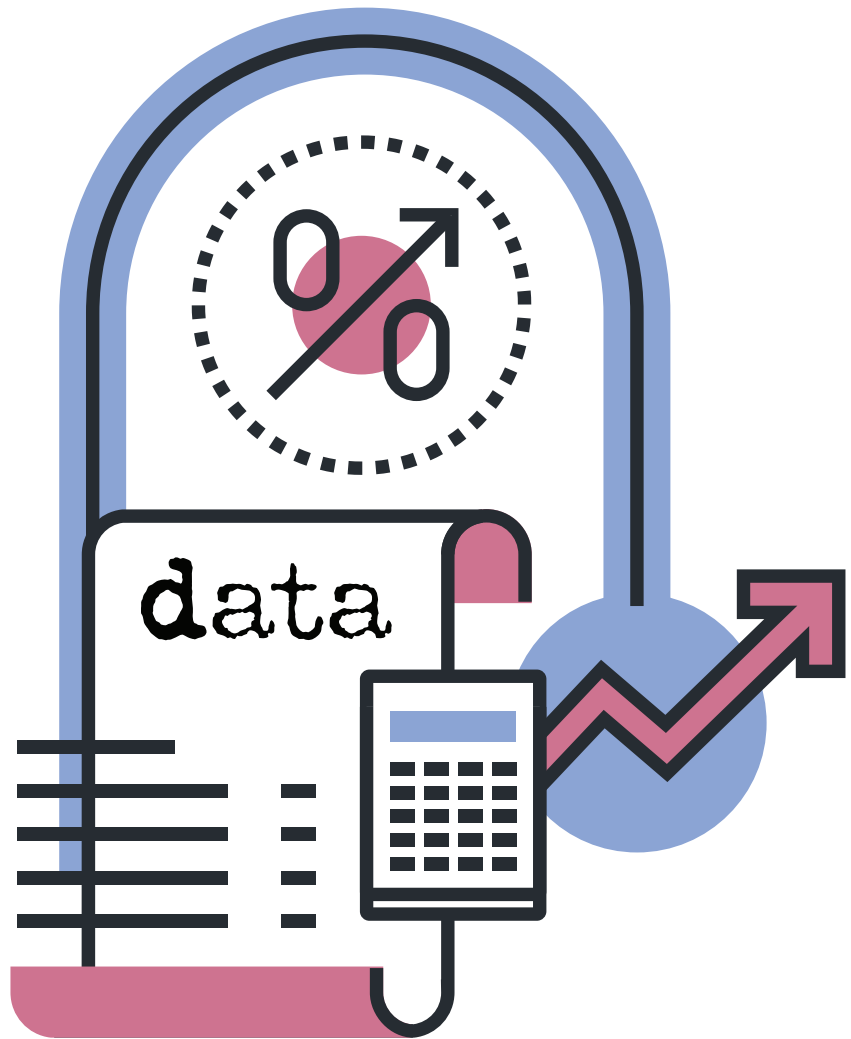
Zeitlyn, D. (2021). For Augustinian Archival Openness and Laggardly Sharing: Trustworthy Archiving and Sharing of Social Science Data From Identifiable Human Subjects. *Frontiers in Research Metrics and Analysis*, 6, 736568, 1-5. https://doi.org/10.3389/frma.2021.736568

https://www.dundee.ac.uk/leverhulme