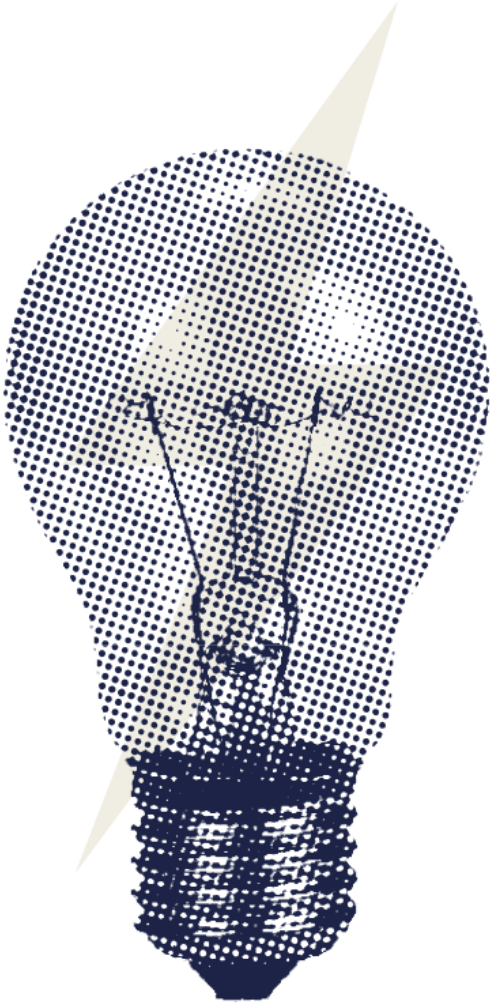
Centre for Data Ethics and Innovation

Unlocking the value of data: Exploring the role of data intermediaries

An exploration of the role intermediaries could play in supporting responsible data sharing



Contents

Executive summary

The government's National Data Strategy sets out to unlock the vast potential of public- and privately-held data in the United Kingdom to drive innovation, boost productivity, create new businesses and jobs, and improve public services. Responsible and efficient data sharing and access are key to realising these goals.

This paper was commissioned by the Department for Digital, Culture, Media, and Sport (DCMS) to support the ambitions set out in the National Data Strategy and subsequent consultation; in particular, the commitment to consider the role of data intermediaries in supporting responsible data sharing, and how the government can intervene to support their adoption.[1] It explores the activity of existing intermediaries across different sectors and considers the role they could play in the future.

Data intermediary is a broad term that covers a range of different activities and governance models for organisations that facilitate greater access to or sharing of data. This paper follows the definition outlined in more detail in Box 1 below, which identifies seven types of data intermediaries, ranging from personal information management systems (PIMS) to industrial data platforms to trusted third parties, among others. They can operate across or within the public, private, and third sectors, encompassing existing institutions in these sectors with new responsibilities or new institutions. In general, data intermediaries enable responsible data access and sharing while managing and protecting individual rights, including preserving privacy. However, data intermediaries may also go beyond legal requirements around data protection and apply additional measures to protect against unethical use of data and ensure it is only used for agreed purposes.

The full value of data could be met if underlying issues inhibiting the sharing of data are addressed. Frontier Economics, in research commissioned by DCMS, identified the following issues: lack of incentives to share data, lack of knowledge around the value of data or the availability of data, commercial, ethical, and reputational risks, legal and reputational risks, costs of data access and sharing, as well as missed opportunities to use data in the public interest. In some cases data intermediaries are already helping to overcome these issues by providing the bespoke governance models and technical infrastructure necessary to facilitate data access and sharing. However, there is scope for intermediaries to play a greater role.

Data intermediaries unlock opportunities to empower individuals and businesses, offering them greater control and choice over who has access to data about them, and the purposes for which it is used. In addition, data intermediaries facilitate data access and sharing for the purposes of analysis. This can include research in the public interest, supporting innovation in commercially-sensitive environments, and enabling the independent audit of data-driven technologies. The COVID-19 pandemic has highlighted the importance of this opportunity, demonstrating how effective use of data can aid decision-making and produce social and economic benefits.

[1] Department for Digital, Culture, Media, and Sport, *National Data Strategy: Policy paper*, 9th May, 2020; Department for Digital, Culture, Media, and Sport, *Government response to the consultation on the National Data Strategy*, May 18, 2021.

There are opportunities to leverage data intermediaries in response to the most pressing social, economic, and environmental challenges today, where greater access to and sharing of data is crucial. This paper discusses three potential opportunities, including: facilitating preventative medicine to improve individuals' wellbeing; enabling better matching of workers with available jobs, particularly as technology transforms the labour market; and enabling the UK to meet its Net Zero targets.

The Centre for Data Ethics and Innovation (CDEI) aims to drive responsible innovation in data-driven technologies. Exploring, developing, and promoting mechanisms for supporting responsible data access and sharing across the economy is a key area of focus for the Centre.

Case studies of data intermediaries and the issues in data access and sharing they address

Table 1 provides an overview of the data intermediary case studies explored in this paper. Each case study has been categorised into a data intermediary type and evaluated against the data sharing issues that they address. Further information can be found by navigating to the relevant Box.

Table 1: Summary table of case studies on data intermediaries included in this paper

Case study	Type of data inter-mediary	Further inform-ation	Lack of incentives to share data	Lack of knowled-ge	Commercial, ethical, and reputational risks	Legal and regulat-ory risks	Costs of data access /sharing
AWS Data Exchange	Data exchange	Box 3		X			X
OpenSAFELY	Data custodian	Box 4			X	X	X
digi.me	PIMS	Box 5	X	X			
SOLID	PIMS	Box 6	X	X			
Open Banking Implemen-tation Entity (OBIE)	Trusted third party	Box 7	X		X		X
Pensions Dashboard	Data custodian	Box 8	X		X	X	
Genomics England	Data custodian	Box 9	X		X	X	
Advanced Product Concept Analysis Environment (APROCONE)	Industrial data platform	Box 10	X	X	X		
MK Data Hub	Industrial data platform	Box 11	X	X	X		X
National Institute of Standards and Technology (NIST)	Data custodian	Box 12			X	X	

Hypothetical pen portrait: Data donation for healthcare research

"Christopher wants to support research into mental health conditions, as they have affected his family in the past. He knows that there are a variety of inherited and environmental factors that may be associated with mental health conditions and believes mental health researchers could benefit if he donates data about himself. However, he is concerned about his privacy, given the sensitive nature of the data.

Health Hub is a non-profit organisation established to mediate the sharing of information between individuals wanting to donate their data, and accredited research organisations that would most benefit from access to it. Rather than ask Christopher to manually provide his data, Health Hub maintains a list of trusted data providers - including banks, GPs, and local authorities - from which it collects relevant information about Christopher on his behalf. Additionally, Christopher is able to easily share relevant data from his smart devices and apps, including information from his smart watch and sleep tracking app. Health Hub then provides accredited organisations secure access to this data for research and charitable purposes.

Christopher trusts that Health Hub will steward his data effectively because it has been certified as a "data donation service provider" by an independent accreditor. This ensures Health Hub is compliant with data protection and cybersecurity standards, and has been able to provide sufficient evidence that the data sharing it facilitates will be for societal benefit. In addition, Christopher must consent to Health Hub collecting data on his behalf and approve who has access to his data. Through this intermediary, researchers are able to make breakthroughs in identifying and treating mental health conditions."

Introduction

The government's National Data Strategy identifies five potential opportunities for data to transform the UK including: boosting productivity and trade, supporting new businesses and jobs, increasing the speed, efficiency, and scope of scientific research, driving better delivery of policy and public services, and creating a fairer society for all.[2] However, the vast potential of public and privately-held data in the UK has yet to be realised.

This paper was commissioned by DCMS to support the ambitions set out in the National Data Strategy and subsequent consultation, in particular, the commitment to consider the role of data intermediaries in supporting responsible data sharing, and how the government can intervene to support their adoption.[3] It considers the different opportunities that data intermediaries, in their various forms, may unlock by enabling data sharing while protecting individual rights and interests. Some of these opportunities reflect areas in which data intermediaries currently operate and could be built upon, whereas others are speculative examples of where intermediaries could prove transformative in the future.

This paper is organised into five sections. Section 1 establishes the definition and types of data intermediaries. Section 2 explores the issues in data access and sharing, and considers how data intermediaries can overcome these issues. Section 3 outlines how data intermediaries could empower individuals and businesses with more control and choice. Section 4 outlines how data intermediaries could enable analysis through data access and sharing, which includes opportunities to conduct analysis in the public interest, in commercially-sensitive environments, and to audit data and data-driven technologies. Section 5 outlines future data sharing scenarios that could be facilitated by data intermediaries. Through case studies and fictional pen portraits, the paper outlines how data intermediaries can help to overcome some of the most significant issues in data sharing. It should be noted that the approaches adopted by these intermediaries are context-specific. What works in one context may not work in another. These case studies and pen portraits are illustrative and are not intended to endorse or criticise the activities undertaken by these intermediaries.

The Centre for Data Ethics and Innovation operates to support responsible data-driven innovation. This paper is the CDEI's first output on the topic of data intermediaries. But facilitating responsible data sharing across the economy, including piloting new forms of data stewardship and governance, is one of three themes that will guide the CDEI's work over the next year, alongside public sector innovation and AI assurance.[4]

---

[2] Department for Digital, Culture, Media, and Sport, *National Data Strategy: Policy paper,* 9th May, 2020; Centre for Data Ethics and Innovation, *Addressing trust in public sector data use*, July 2020.
[3] Department for Digital, Culture, Media, and Sport, *National Data Strategy: Policy paper,* 9th May, 2020; Department for Digital, Culture, Media, and Sport, *Government response to the consultation on the National Data Strategy*, May 18, 2021.
[4] Centre for Data Ethics and Innovation, *Centre for Data Ethics and Innovation: Two year review*, July 2021; Centre for Data Ethics and Innovation, *The need for effective AI assurance*, 15 April 2021.

Section 1. What are data intermediaries?

Data intermediary is a broad term that covers a range of different activities and governance models for organisations that facilitate greater access to or sharing of data. Box 1 below outlines the definition adopted in this paper, which draws on definitions by the Open Data Institute (ODI), the Organisation for Economic Cooperation and Development (OECD), among others.[5]

Box 1: What are data intermediaries?

Each time data is shared, accessed, used or protected, a number of stewardship activities would typically take place at the intersection of the data sharing and access journeys. They can include, for example, finding data that is fit-for-purpose, managing transfers and usage rights, and ensuring that the right protections are in place.

Data intermediaries - operating in the public, private or third sectors - could help absorb some of the costs and risks that would be normally associated with performing data processing activities in-house. There is already a vibrant ecosystem of innovative data intermediaries, which act between those sharing and accessing data. Many of these organisations are creating novel, technology-enabled solutions to allow safe and frictionless data sharing.

Intermediaries can provide technical infrastructure and expertise to support interoperability between datasets, or act as a mediator negotiating sharing arrangements between parties looking to share, access, or pool data. They can also provide rights-preserving services - for example, by acting as a data custodian allowing remote analysis through privacy-enhancing technologies, or providing independent analytical services in a siloed environment.

Data intermediaries could assume the roles and obligations of a data controller and/or processor,

Data intermediaries can facilitate data sharing for both personal and non-personal data for commercial and not-for-profit purposes. There is a tendency to focus on personal data when discussing data sharing solutions, yet non-personal data can also be immensely valuable and brings additional challenges, including commercial sensitivities that need to be addressed.[6] Data intermediaries can operate across or within the public, private, and third sectors. They can also be established by organisations operating in the public or private sectors, or operate as new institutions.

[5] Fionntán O'Donnell and Rebecca Ghani, "Matchmakers of the data world: How 'data intermediaries' can bring decision makers and data together to help combat Covid-19", *Open Data Institute*, October 13, 2020; Organisation for Economic Cooperation and Development (OECD), *Enhancing access to and sharing of data: Reconciling risks and benefits for data re-use across societies*, November, 2019.

[6] The Open Data Institute, *The data spectrum*, 2021.

Types of data intermediaries

Data intermediaries cover a range of different types of institutions that address different issues and can open up new opportunities. Table 2 provides the definitions of different data intermediary types, which are a subset of intermediary types identified by the ODI and the OECD. [7]

Table 2: Types of data intermediaries

Type	Definition
Data trusts	Provide fiduciary data stewardship on behalf of data subjects.[8]
Data exchanges	Operate as online data platforms where datasets can be advertised and accessed - commercially or on a not-for-profit basis.
Personal information management systems	Seek to give data subjects more control over their personal data.
Industrial data platforms	Provide shared infrastructure to facilitate secure data sharing and analysis between companies.
Data custodians	Enable privacy-protecting analysis or attribute checks of confidential data, for example, via the application of Privacy-Enhancing Technologies (PETs).
Data cooperatives	Enable shared data spaces controlled by data subjects.
Trusted third parties	Provide assurance to those looking to access confidential datasets that the data is fit-for-purpose (e.g. in terms of quality or ethical standards).

The exhibits below showcase some of the different data flows and architectures supported by these different types of data intermediary.

[7] The Open Data Institute, *Mapping the wide world of data access*, 2021; Organisation for Economic Cooperation and Development (OECD), *Enhancing access to and sharing of data: Reconciling risks and benefits for data re-use across societies*, November, 2019

[8] See Ada Lovelace Institute and the UK AI Council, *Exploring legal mechanisms for data stewardship*, March 2021.

Exhibit 1: Personal Information Management System (PIMS)

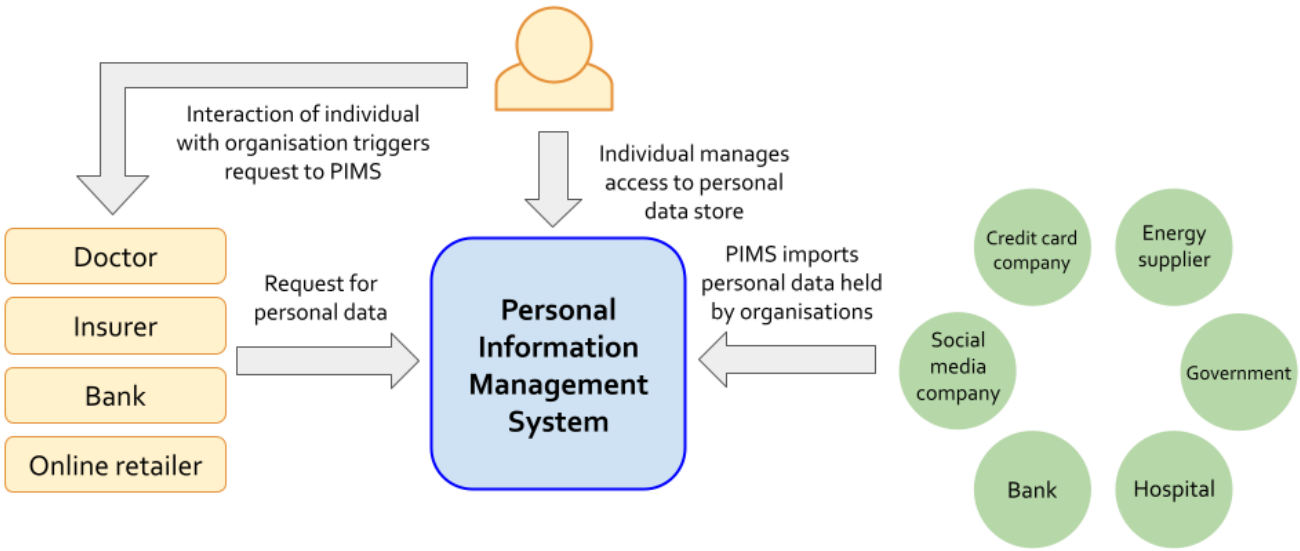


Exhibit 1 shows a personal information management system (PIMS), where an individual imports personal data from providers such as social media companies, banks, hospitals, and the government, among others. The individual is able to manage who has access to their personal data store, granting or revoking access to organisations such as GPs, banks, and online retailers, among others. The case studies on digi.me and SOLID in Boxes 5 and 6 below provide real-world examples of this type of data intermediary.

Exhibit 2: Data custodian

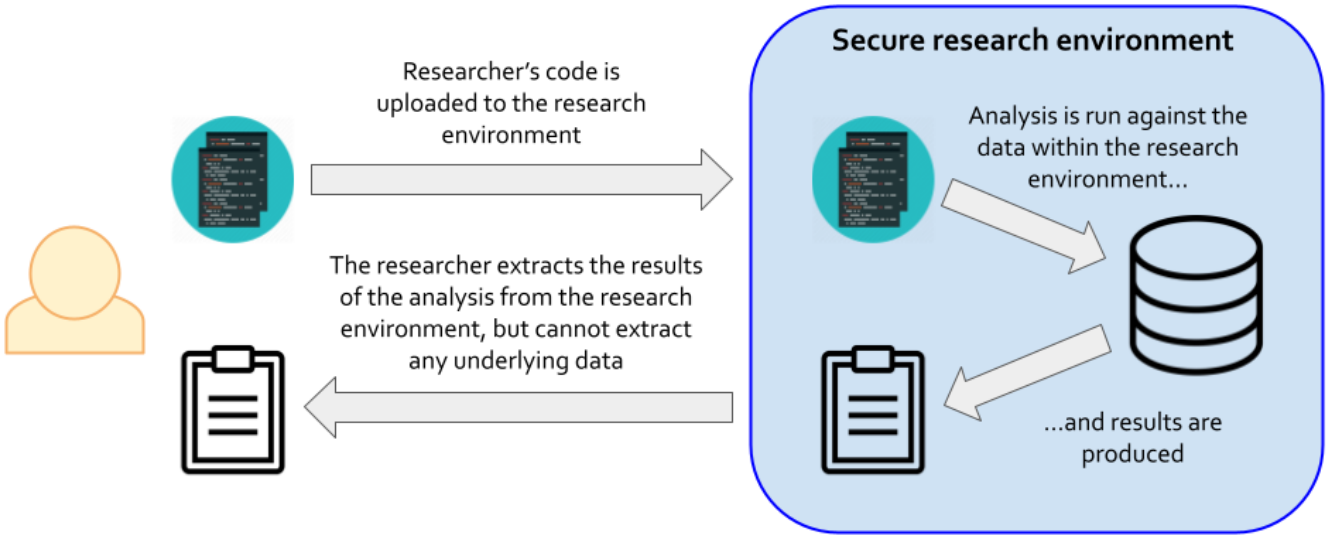


Exhibit 2 shows a high-level overview of a data custodian that manages a secure research environment holding highly sensitive data, such as healthcare data. In this example, a researcher uploads their analytic code to the environment where it is executed against the data. The researcher is unable to copy or extract any of the raw data from the environment, only the results of the analysis.

Exhibit 3: Data collaborative

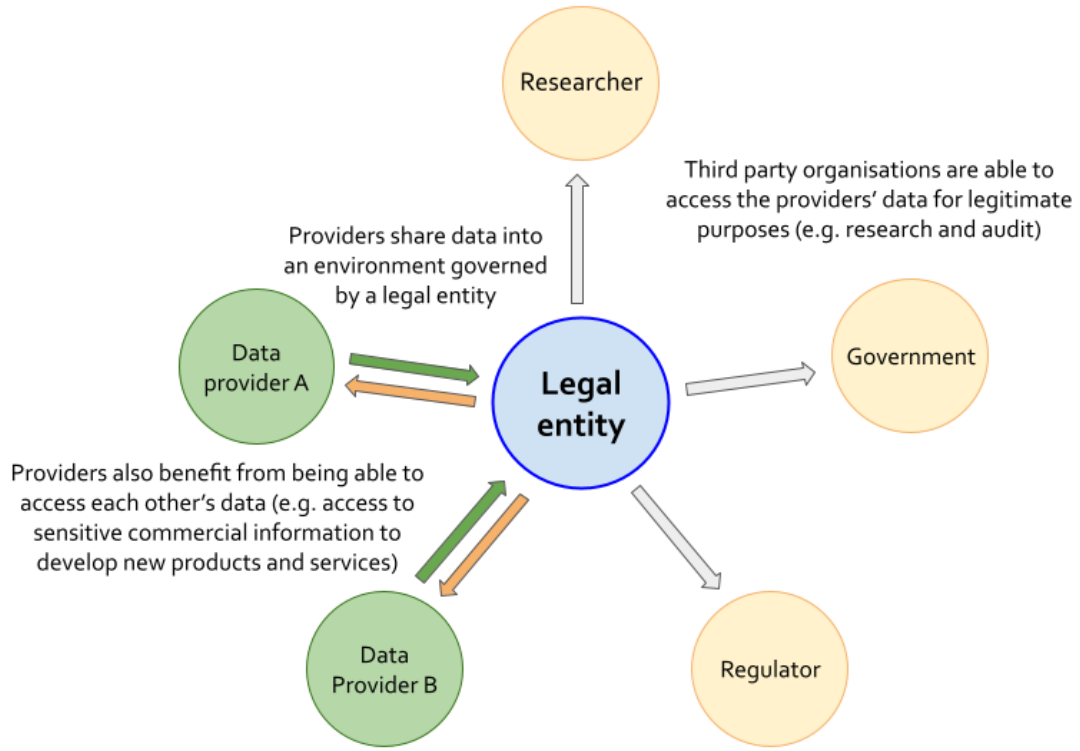


Exhibit 3 shows how a data intermediary (the legal entity) could act as a data collaborative enabling data providers (e.g. A and B) to share their data with other parties (e.g. researchers, the government, or regulators) that have a legitimate interest in accessing their data. The data intermediary manages access rights to the data it holds and ensures compliance with relevant data protection regulations.

Section 2. Exploring issues in data access and sharing

According to Frontier Economics, in research commissioned by the Open Data Institute (ODI), increased public and private sector data sharing could generate social and economic benefits worth between 1 and 2.5 percent of GDP.[9] However, these benefits have not yet been realised due to underlying issues that inhibit the sharing of data. In some areas, data intermediaries are already being used as a mechanism to facilitate data access and sharing by providing governance frameworks and technical infrastructure that tackle these issues.

Understanding the issues holding back data sharing

Frontier Economics, in separate research commissioned by DCMS, identified several issues holding back data sharing that are outlined in Table 3.[10] There are opportunities for data intermediaries to overcome these issues.

Table 3: Issues that may prevent optimal data sharing

Issue	Description	Potential role of data intermediary
Lack of incentives to share data	Data providers may not be sufficiently incentivised to share or provide access to their data (e.g. because sharing requires them to incur costs or effort that they are not able to recoup from those that benefit).	Operate a "pay to play" model where individual organisations can only access the larger datasets if they too have contributed data (see Online Safety Data Initiative (OSDI)).
Lack of knowledge	Data providers may lack sufficient knowledge of the potential uses of their data, while data users lack sufficient knowledge of what data could be made available, and how.	Enable easier discovery of particular datasets, enabling data providers to generate value from their data and enabling data users to identify potential applications (see AWS Data Exchange in Box 3).
Commercial, ethical, and reputational risks	Perceived or actual risk of losing competitive advantage, suffering reputational damage from data uses that breach others' trust, or enabling ethically questionable uses of data may deter data access and sharing.	Provide assurance to data subjects and holders that data will only be accessed for defined purposes by approved individuals/organisations (see APROCONE in Box 10).

[9] Frontier Economics, *Economic impact of trust in data ecosystems: Report prepared for the ODI*, February 2021.
[10] Frontier Economics, *Increasing access to data across the economy: Report prepared for the Department for Digital, Culture, Media, and Sport*, March 2021.

Legal and regulatory risks	Perceived or actual risks of breaching data protection, intellectual property rights, or regulatory requirements may also provide a deterrent to sharing.	Provide legal expertise and common data standards to facilitate legally-compliant data sharing (see Pensions Dashboard in Box 8).
Costs of data access/sharing	Costs may be prohibitive because of a lack of common foundations, infrastructure, and technologies that are needed for data sharing to be cost effective.	Provide data sharing environments that make effective use of technology including applying privacy-enhancing technologies (see OpenSAFELY in Box 4).
Missed opportunities to use data in the public interest	Cases where data sharing may be particularly likely to lead to economic and social benefits.	Provide secure access to sensitive data in a trusted research environment for the purpose of research in the public interest (see Genomics England in Box 9).

Establishing trustworthy data sharing

The Centre for Data Ethics and Innovation has previously highlighted that many of the challenges associated with data access and sharing are linked to a lack of trust between parties.[11]

Concerns around who has access to data, how it will be used, and potential asymmetries in the benefits accrued by parties all affect trust. In addition, uses of data that may be legal but could be considered irresponsible or unethical undermine public trust more widely. Such tensions come to a head when data sharing and requests for access involve managing competing interests, such as balancing an individual's desire for privacy against analysing data to understand the potential harm an intervention could have on particular groups. Such a challenge can be viewed as a collective action problem whereby individual data providers (e.g. private companies, public sector bodies, etc.) lack incentives, or face misaligned incentives, to cooperate with others to share data, even though cooperation may bring benefits for all participants. Data intermediaries are able to manage such competing interests and establish trustworthy, and often collaborative, data sharing arrangements.

Limiting the purposes for which data may be shared is crucial to securing the trust of organisations or individuals whose data is being accessed. Intermediaries can establish clear purpose limitations on data use, and ensure no onward sharing. Furthermore, there are cases where complex considerations need to be made before awarding access to data. Data intermediaries are able to establish the mechanisms for making such decisions and ensure data is used responsibly. For example, Genomics England can allow data to be used for research purposes (the purpose limitation) but has a relatively broad remit to determine which research proposals are legitimate (see Box 9 for further information).

[11] Centre for Data Ethics and Innovation, *Addressing trust in public sector data use*, 20 July, 2020.

Finally, lack of quality data may inhibit data sharing by undermining data users' confidence that the available data will be fit for purpose and enable them to make informed decisions using the data.[12] From a developer perspective, poor quality data may require substantial investment in cleaning and formatting before it can be used. In addition, data providers may be unwilling to share data if it is of poor quality as it poses a reputational risk for their organisation. Data intermediaries are able to provide assurances around the quality of data and impose particular standards to make working with it easier and less costly.

Establishing the technical infrastructure for data access and sharing

Data intermediaries can provide the technical infrastructure necessary to facilitate data access and sharing. They can help organisations to lower the costs of and knowledge required to share data, particularly if building this infrastructure is prohibitively costly for data providers or data users.

By enhancing interoperability and data portability, data intermediaries can build bridges between different data providers. Interoperability can be encouraged by creating common data standards between data providers, enabling the efficient aggregation of datasets by allowing developers to implement a single, consistent code library across multiple partners, or creating Application Programming Interfaces (APIs) that facilitate more efficient and automated data sharing. This could include facilitating data transfers across borders. Data portability allows customers to easily transfer their data from one provider to another. The combination of enhanced interoperability and portability has helped to facilitate more competition in sectors such as banking.

Data intermediaries can also provide the technical infrastructure that enables secure sharing of data, particularly personal data. This could include embedding effective cybersecurity measures, applying privacy-enhancing technologies (PETs) (explored in Box 2 below), or creating trusted research environments (TREs), among others. Data intermediaries such as OpenSAFELY (outlined in Box 4), the ONS Secure Research Service, and the Pensions Dashboard (outlined in Box 8) have adopted different approaches to sharing data securely, which protect data subjects yet also enable data to be used in new and innovative ways. There may be opportunities for data intermediaries to take more action in this space going forward.

Box 2: What are privacy-enhancing technologies (PETs)?

A key challenge of providing access to sensitive data is balancing privacy and transparency. In order to make use of data it needs to be accessible to the data user (transparency), which necessarily compromises data subjects' privacy to some degree. A set of emerging privacy-enhancing technologies (PETs) are beginning to shift this tradeoff so that high levels of both privacy and transparency can be achieved. These technologies can enable an individual or

[12] Government Data Quality Hub, *What is data quality?*, May 6, 2021.

organisation to answer questions using data they cannot see.[13] The CDEI's PETs Adoption Guide explores how PETs can be applied in practice.[14]

Data intermediaries could leverage PETs to enable sensitive data to be more widely utilised, or to enable data to be accessed in a more privacy-focused way. For example, intermediaries could use PETs to enable sensitive transactional data from financial institutions to be collectively analysed in order to identify fraud coordinated across those institutions.[15] A real-world example of an intermediary using PETs is OpenSAFELY (discussed in Box 4), which implements a federated analytics system to enable research to be conducted on large volumes of healthcare data without researchers ever being able to directly observe the data.

Box 3: AWS Data Exchange

Data sharing issues addressed by the data intermediary (descriptions outlined in Table 3)

Lack of incentives to share data		Legal and regulatory risks	
Lack of knowledge	X	Costs of data sharing and access	X
Commercial, ethical, and reputational risks		Missed opportunities to use data in the public interest	X

AWS Data Exchange is a commercial data exchange that aggregates and stores non-personal data from public and private organisations such as Reuters, the news agency, Change Healthcare, a company that processes healthcare transactions, and Dun & Bradstreet, who maintain a database of global business accounts, among others.

AWS Data Exchange simplifies finding, subscribing to, and using data from third-party providers through a web portal. It enables data users to explore what data is available for their specific needs. It also provides the infrastructure to share and aggregate datasets easily. This includes creating a common API that allows users to integrate data from multiple sources as well as providing cloud computing services that enable users to store, maintain, and run analyses on their datasets.

During the COVID-19 pandemic, AWS Data Exchange offered the COVID-19 Data Exchange to facilitate research on the pandemic. Data from the exchange has been used to better predict COVID epidemiology by the Chan Zuckerberg Biohub.[16]

[13] OpenMined, 2021
[14] Centre for Data Ethics and Innovation, *PETs Adoption Guide*, 2021
[15] The FFIS Privacy-Enhancing Technology (PET) project, 2021
[16] AWS Data Exchange, 2021.

Box 4: OpenSAFELY

Data sharing issues addressed by the data intermediary (descriptions outlined in Table 3)

Lack of incentives to share data		Legal and regulatory risks	X
Lack of knowledge		Costs of data sharing and access	
Commercial, ethical, and reputational risks	X	Missed opportunities to use data in the public interest	X

In 2020, OpenSAFELY was developed by the DataLab at the University of Oxford and a number of electronic health record providers (who already manage NHS patients' records), working on behalf of NHS England and NHSX.

OpenSAFELY is a secure analytics platform for electronic health records that allows independent researchers to run analyses without ever being able to see the data directly, and without transferring it out of the secure data centre in which it resides. OpenSAFELY enables research in the public interest using sensitive data, while maintaining the privacy and confidentiality of that data. Effective response to the COVID-19 pandemic has relied on large-scale, real-time analysis of sensitive datasets, such as patient healthcare records.

One of OpenSAFELY's first projects identified patients most at risk of dying from COVID-19 with significantly higher accuracy than other approaches, enabling relevant organisations to focus their resources on the most vulnerable patients.[17]

[17] OpenSAFELY, 2021.

Section 3. Empowering individuals and businesses in data access and sharing

Existing data intermediaries already empower individuals and businesses to have control and choice over who has access to data about them and how it is used. However, this is a relatively nascent area and there may be opportunities for further developments.

More control could ensure individuals are better placed to exercise their data rights and give meaningful consent. In addition, individuals and businesses could be empowered to use data about themselves to make more informed decisions, benefit from new and/or improved products and services, or support innovation in new data-driven markets.[18]

Enabling individuals to control how data about them is used and for what purposes

Personal information tends to be fragmented and dispersed across many locations, making it difficult for individuals to have a full overview of and control over the data they share. This is reflected in the CDEI's research highlighting that people believe they have limited control over their online experiences. Only 36 percent of respondents in a survey commissioned by the CDEI believed that they have meaningful control over online targeting systems, while only 33 percent believe that companies will do what users request through their settings and preferences.[19] Similarly, the Open Data Institute highlights that there is a greater demand for mechanisms to "rebalance the current power imbalance between individuals and corporations".[20]

Personal information management systems (PIMS) - also known as personal data stores or data lockers - have emerged in response to individuals' real or perceived loss of control over how data about them is stored and used. These data intermediaries aim to return control to users by enabling all information related to them to be stored in a single location and managed through a single interface. Individuals are given fine-grained control over who has access to data about them, and are empowered to withdraw that access at any time. Although some PIMS, such as Nesta Decode, are non-profit enterprises, others such as digi.me, HATDeX, and SOLID are commercial enterprises, suggesting that a competitive marketplace may emerge in response to individuals' preferences to have more control over data about themselves. Boxes 5 and 6 below provide an overview of how digi.me and SOLID work in principle. Other examples of PIMS, such as Mine, facilitate an individual's right to erasure under existing data protection legislation by automating the process of requesting data deletion from an organisation.[21] Looking beyond current uses, PIMS could enable

[18] Centre for Data Ethics and Innovation, *Addressing trust in public sector data use*, July 2020.
[19] Centre for Data Ethics and Innovation, *Online targeting: Final report and recommendations*, February 4, 2020.
[20] Jack Hardinges and Dr. Jared Robert Keller, "What are 'bottom-up' data institutions and how do they empower people?", *Open Data Institute*, June 25, 2021.
[21] Mine, 2021

automatic opt-outs on data sharing for uses that may be considered undesirable, such as gambling or political advertising.

In the future, data intermediaries could provide the secure infrastructure that enables individuals to verify their identity without having to disclose personally identifying information (known as attributes) to other individuals or organisations. For example, data intermediaries could support digital identity services by enabling real-time attribute verification by authoritative bodies, while also ensuring individuals have control over who has access to their data. If an individual needs to confirm their age for age-restricted goods and services, a third party could scan a QR code on their mobile phone that generates a unique link to a subset of a record on a government database, such as the individual's age on their driver's license record. All of the personal data remains in the government database and only the necessary information has to be verified for any situation. In this scenario, the results could confirm that the individual is over the age requirement without disclosing their exact age.[22]

Box 5: digi.me

Data sharing issues addressed by the data intermediary (descriptions outlined in Table 3)

Lack of incentives to share data	X	Legal and regulatory risks	
Lack of knowledge	X	Costs of data sharing and access	
Commercial, ethical, and reputational risks		Missed opportunities to use data in the public interest	X

digi.me is a personal information management system (PIMS) that provides users with a personal data store in which to collate their own data and share with apps and companies that are integrated with the digi.me platform.

Through digi.me, users can import data about themselves from online providers such as social media platforms. The data is encrypted within the digi.me Private Sharing app, and stored in a personal cloud of choice (e.g. Dropbox, Google Drive, etc.). Individuals are then able to control what data about them is shared with private companies. digi.me provides an overview of individual companies' purpose limitations on personal data when they request access and identifies whether a copy of their data will be stored. Individuals can also choose to revoke apps and companies' access to their personal data store.[23]

digi.me has facilitated data sharing and access in the healthcare, financial services, and social media sectors. For example, digi.me has enabled Dutch citizens to share data about themselves securely with the government, while HealthyMe allows Icelandic citizens to download their

[22] Department for Digital, Culture, Media and Sport, *The UK digital identity and attributes trust framework*, February 2021
[23] digi.me, 2021.

electronic health records and, Consentry transfers COVID-19 test results and vaccine certificates back to individuals using digi.me's technology.[24]

Box 6: SOLID

Data sharing issues addressed by the data intermediary (descriptions outlined in Table 3)

Lack of incentives to share data	X	Legal and regulatory risks	
Lack of knowledge	X	Costs of data sharing and access	
Commercial, ethical, and reputational risks		Missed opportunities to use data in the public interest	X

SOLID is a personal information management system where individuals are the primary stewards of their own data - they decide what data is stored, where it is stored, and who has access to it. Individuals can also revoke access to their data.[25] SOLID is currently being piloted in Greater Manchester to provide individuals control over their health and social care records, collating them in a personal data store and granting authorised medical professionals access to these records. SOLID can enable individuals to build a holistic representation of their healthcare, by collating disparate healthcare data that currently exists in silos across the NHS, as well as potentially uploading information from fitness or nutrition apps.

The Flanders Government is piloting SOLID for government services. It has provided citizens with a personal digest of government services that can be accessed via a front-end application, My Citizen Profile. My Citizen Profile collates data from different areas of government (e.g. departments, agencies, and localities, etc.) into a single accessible application and can be used to guide citizens to the services they need, from wherever they enter the network. Using My Citizen Profile, a citizen can also share personal information with government departments while their data remains in the personal data store.[26] NatWest Bank and the BBC are also piloting this technology, suggesting opportunities to expand this approach to other sectors in future.[27]

[24] Ibid.
[25] SOLID, 2021.
[26] Ruben Verborgh and Katrien Mostaert, *Streamlining governmental processes by putting citizens in control of their own data - Solid*, 2019
[27] Digi.Me, *What is digi.me?*, 2021; Greater Manchester Digital Platform, *New SOLID technology aims to support dementia patients*, November, 2020; BBC News, *NHS data: Can web creator Sir Tim Berners-Lee fix it?*, November 2020; Financial Times, *World wide web founder scales up efforts to reshape internet*, February 2020.

Providing individuals and businesses with more choice over data-driven products and services

Providing individuals and businesses with more choice over goods and services drives competition between providers, while also stimulating innovation in new goods and services. Data intermediaries already facilitate more choice for individuals and businesses, but there are opportunities to do more.

The financial services sector is leading in this regard with Open Banking (outlined in Box 7 below). Opening up access to banking data has helped foster a new ecosystem of fintech startups providing a range of customer and business-facing services that are regulated by the Financial Conduct Authority (FCA). It has enabled services for personalised price comparison, for finding competitive exchange rates when making payments abroad, and for seamlessly making charitable "micro-donations" by rounding up purchases.[28] By the end of 2020, there were over 300 licensed fintechs and service providers using Open Banking, and over 2.5 million consumers and SMEs using Open Banking-enabled services to manage their finances.[29]

The government established the Smart Data Working Group to support the development and delivery of smart data infrastructure and standards for the benefit of consumers, particularly vulnerable consumers, in energy, communications, and pensions (explored in Box 8 on the Pensions Dashboard Programme). This would extend the successes of Open Banking to these sectors. The Smart Data Working Group aims to develop systems and standards - including digital identification - that enable smart data innovations, thereby encouraging the development of new, innovative services for consumers and SMEs.[30] If this approach is extended to other sectors, it has been estimated that consumers could save four billion pounds a year by shifting to the best available offer.[31] At an aggregate level, greater personal data mobility could increase UK GDP by an estimated £27.8 billion.[32] Current efforts to establish a new pro-competition regime in digital markets may provide further opportunities for intermediaries to support data sharing. This could include driving interoperability and better enabling consumers to control and share their data.[33]

[28] Open banking, 2021
[29] Open banking, 13 January 2021
[30] Department for Business, Energy, & Industrial Strategy, *Smart Data Working Group*, 2021.
[31] Department for Business, Energy, & Industrial Strategy, *Next steps for Smart Data: Putting consumers and SMEs in control of their data and enabling innovation*, September 2020.
[32] CtrlShift, *Data mobility: The personal data portability growth opportunity for the UK economy*, 2018.
[33] In April 2021 a Digital Markets Unit was established in the CMA to operationalise the future pro-competition regime for digital markets.

Box 7: Open Banking Implementation Entity (OBIE)

Data sharing issues addressed by the data intermediary (descriptions outlined in Table 3)

Lack of incentives to share data	X	Legal and regulatory risks	
Lack of knowledge		Costs of data sharing and access	X
Commercial, ethical, and reputational risks	X	Missed opportunities to use data in the public interest	X

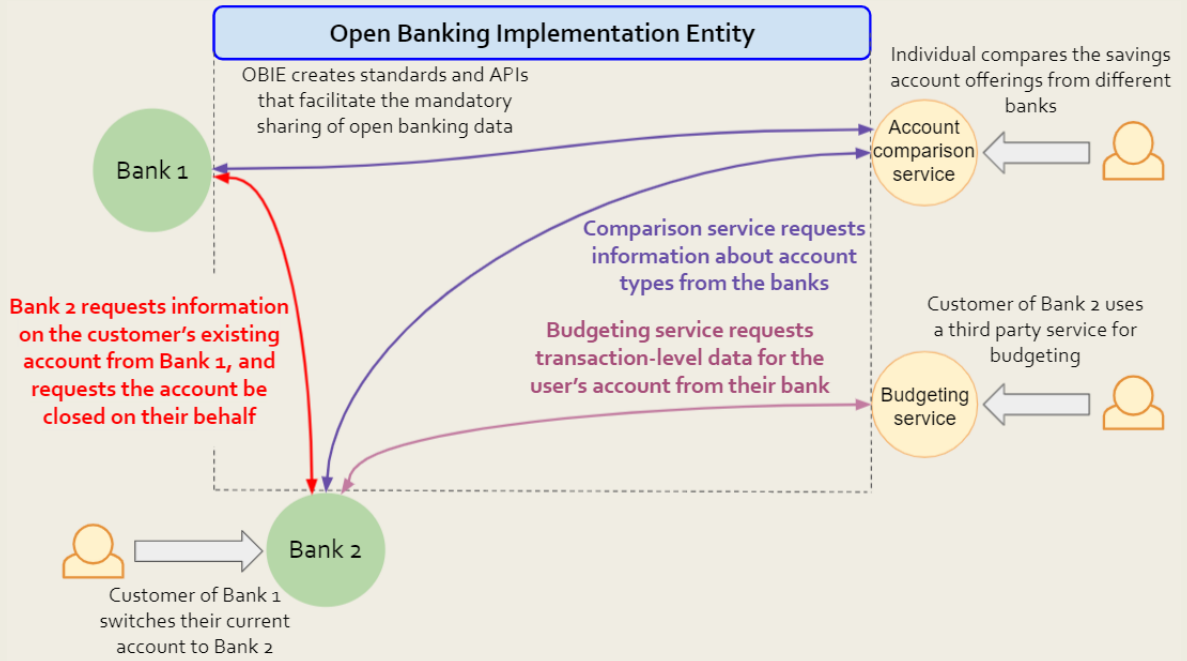
The Open Banking Implementation Entity (OBIE) acts as a trusted third party, creating the technical standards and industry guidelines that support Open Banking. In this ecosystem, the Financial Conduct Authority (FCA) manages the licensing of third party providers (TPPs) that are authorised to access consumer data.[34] Consumers can determine who has access to their data and can revoke access at any point.[35]

OBIE ensures that all banks and building societies comply with standardised formats for APIs, security profiles, customer experience, and operational guidelines in order to facilitate trustworthy and secure data sharing, while also supporting these businesses in complying with existing regulations.[36] Standardising the underlying requirements improves efficiency and levels the playing field for all providers, ensuring that small fintechs - as well as large banks - can access the data they need to develop new products and services and compete in the market.

Open Banking - facilitated by OBIE - offers individuals and SMEs more choice by enabling them to share data about themselves to identify the most appropriate financial product or service for their individual needs. These include easily switching bank accounts to new providers, using budgeting apps that provide personalised recommendations, and using price comparison websites to receive tailored quotes for financial products.[37] **Exhibit 4** provides an overview of how OBIE works in practice. Research by Experian suggests that consumers are taking advantage of greater choice in financial markets: requests to share data through Open Banking tripled during the pandemic.[38]

[34] Open Banking, 2021
[35] Open Data Institute and Fingleton, *Open Banking, preparing for lift off: Purpose, progress, and potential*, 2019.
[36] Open Banking, 2021
[37] Open Banking, *What is Open Banking?*, 2021.
[38] Open Banking, *May 2021: Monthly Highlights*, 2021.

Exhibit 4: Open Banking Implementation Entity (OBIE)



Box 8: Pensions Dashboard

Data sharing issues addressed by the data intermediary (descriptions outlined in Table 3)

Lack of incentives to share data	X	Legal and regulatory risks	X
Lack of knowledge		Costs of data sharing and access	
Commercial, ethical, and reputational risks	X	Missed opportunities to use data in the public interest	X

The Pensions Dashboard Programme (PDP), established by the government, will act as a data custodian, providing secure access to pensions information in a single dashboard and linking data from private pension schemes, government pension schemes, and personal identifiers in a single ecosystem.[39] The PDP is creating the underlying digital architecture and governance framework to enable the sharing of pensions data. This will include creating common data standards, establishing an identity verification service, and facilitating consent and authorisation to share an individuals' data.

The PDP will empower individuals to make more informed decisions about their financial security in retirement through combining individuals' pensions in a single dashboard, projecting an individuals' retirement income based on their pensions, and providing impartial advice on

[39] Pension Dashboard Programme, 2021.

retirement planning. According to research by Aegon, 73 percent of individuals surveyed have multiple pension pots, while 17 percent have misplaced one or multiple pension pots.[40] Over 25 percent of retired individuals aged 55 and over do not know the size of their pension savings.[41] Consolidating an individual's pension information in one place will enable savers to make informed choices that help them better prepare for retirement.

[40] Amy Austin, "Fewer savers losing track of pension pots", *FT Adviser*, May 4, 2011.
[41] Financial Conduct Authority, *Data Bulletin*, March 2018.

Section 4. Enabling analysis through data access and sharing

The response to COVID-19 has exemplified the importance of data sharing for the purpose of analysis. The pandemic created a pressing need to access and share data between the public and the private sector, and necessitated new forms of collaboration. As part of the response, new data intermediaries, such as the NHS COVID-19 contact tracing app and the COVID-19 Data Store, were developed to facilitate rapid evidence-based decision-making to tackle the pandemic.[42]

The COVID-19 response also exposed challenges in relation to data access and sharing, particularly around ensuring data quality, addressing methodological differences, and a lack of infrastructure, particularly around the need for new data sharing agreements.[43] For example, as part of the effort to identify vulnerable people during the pandemic, the 33 London boroughs would have had to establish potentially hundreds of data sharing agreements with each other to share sensitive data on individuals eligible for free school meals.[44] Other issues included inconsistencies in the availability and quality of statistics on COVID-related deaths across hospital trusts and nations within the UK.[45] Data intermediaries could help to address such challenges, bringing wider benefits such as data-informed decision-making, and supporting research and innovation.

Enabling data access and sharing in the public interest

Surveys by the CDEI highlight that individuals tend to be more comfortable sharing their data for purposes in the national interest.[46] Data intermediaries can overcome issues in data access and sharing where there is a clear public benefit in doing so. The healthcare sector is leading in this regard. Data intermediaries such as the UK Biobank and Genomics England have invested in collecting and storing data on the human genome at scale.

Genomics England operates as a trusted research environment (TRE) that grants accredited researchers access to its data for approved projects (outlined in Box 9 below). It has contributed to the growth of a thriving genomic industry in the United Kingdom, driving innovations in personalised medicine and new treatments for major diseases. Other TREs operating in the healthcare space include the Scotland Data Safe Haven programme, the UK Secure eResearch Platform in Wales, and the UK Data Service Secure Lab.[47]

[42] Ada Lovelace Institute and the Royal Society, *Learning data lessons: data access and sharing during COVID-19*, November 2020.
[43] Ibid.
[44] Ibid.
[45] Ibid.
[46] Centre for Data Ethics and Innovation, *CDEI publishes polling data on data sharing*, May 2021; Centre for Data Ethics and Innovation, *Online targeting: Final report and recommendations*, February 2020.
[47] UK Health Data Research Alliance, *Trusted Research Environments: A strategy to build public trust and meet changing health data science needs*, 2020.

The Office for National Statistics' Secure Research Service (ONS SRS) is another example of a data intermediary that facilitates research and development (R&D) in the public interest. Similar to Genomics England, it accredits both researchers and research projects, providing access to anonymised data for specific purposes based on criteria outlined in the Research Code of Practice and Accreditation Criteria.[48] Through this platform, researchers have assessed the efficacy of prostate cancer screening, saving the NHS up to £1 billion a year in screening and follow-up treatments, assessed burglary security effectiveness, and analysed the prevalence of COVID-19 infection within and across households, among other projects. The Integrated Data Programme (IDP), proposed by the ONS in the National Data Strategy, builds on the ONS SRS to facilitate the discovery of public sector-based Open Data. The IDP aims to bring together data on a single platform and in a standard format that will save time, effort, and skills on the part of data users.[49]

Data intermediaries can enhance capacity for undertaking data intensive research, particularly by organisations that may not have the financial capacity to invest in the data infrastructure. For example, Stanford University in the United States has proposed establishing a national research cloud for academic and non-profit researchers, which would democratise access to the cutting-edge computing power, data, expertise, and capital necessary to train machine learning models and conduct valuable research.[50]

[48] UK Statistics Authority, *Research code of practice and accreditation criteria*, February 2020; Office for National Statistics, 2021.
[49] Office for National Statistics, *Publishing with the Integrated Data Programme*, March 11, 2021.
[50] Stanford University, *National Research Cloud*, 2021; John Thornhill, "A public research cloud would stimulate innovation", *Financial Times*, October 19, 2020.

Box 9: Genomics England

Data sharing issues addressed by the data intermediary (descriptions outlined in Table 3)

Lack of incentives to share data	X	Legal and regulatory risks	X
Lack of knowledge		Costs of data sharing and access	
Commercial, ethical, and reputational risks	X	Missed opportunities to use data in the public interest	X

Genomics England is a data custodian that collects sensitive data on the human genome for research purposes as part of the 100,000 Genomes Project. It sequences genomic data from individuals with rare diseases or cancer in order to facilitate breakthroughs in medical treatment, particularly personalised medicine.[51]

Genomics England works with the NHS to identify potential participants (e.g. those with rare diseases or cancer). Participants must consent to having their DNA sequenced and can opt out from participating in the project at any time. If a participant consents, the DNA sample collected at their hospital or GP is sent for genome sequencing to extract their genomic data. Genomics England analyses a participant's health data alongside their genomic data to create a report that is then shared with the participant's medical professionals to better understand their condition and enable more targeted treatment.

However, some medical conditions are not fully understood and require further research. Genomics England operates as a trusted research environment (TRE), granting public and private researchers access to its anonymised data for specific research projects. Approval for these projects goes through multiple stages to protect data subjects and ensure that the project conforms to the purpose limitations and governance model adopted by Genomics England. In addition, Genomics England never allows the data to be removed from its secure data centre, monitors all activity undertaken by researchers, and establishes strict legal repercussions if researchers re-identify individuals from the database.

Genomics England is contributing to the growth of a thriving genomics industry in the United Kingdom, which Deloitte estimated contributed 10 percent to the total market for genomics globally in 2015.[52] Startups and smaller companies are playing an important role in driving this industry, and formed part of a growing £1.9 billion economic contribution in 2018/19.[53]

[51] Genomics England, 2021.
[52] Deloitte, *Genomics in the UK: An industry study for the Office of Life Sciences*, 2015.
[53] Department for Business, Energy, and Industrial Strategy and the Department of Health and Social Care, *Genome UK: The future of healthcare*, September 26, 2020.

Facilitating data sharing in commercially-sensitive environments to drive innovation

Data intermediaries can enable the brokering of data sharing solutions in commercial environments where lack of trust between competing parties is particularly acute. In these environments, data sharing brings benefits to all parties involved - such as improving supply chain management, enabling collaborative product development and design, or tackling industry-wide issues - yet there may be competing interests such as a desire to protect sensitive intellectual property or market position that inhibit data sharing. One example of this is Advanced Product Concept Analysis Environment (APROCONE), which is an industrial data platform that facilitates collaborative product design in the airline industry. Box 10 provides more detail. Another example of a data intermediary operating in this space is HiLo Maritime Risk Management (HiLo), a not-for-profit initiative set up by major players in the maritime industry to benchmark the health and safety performance of shipping companies against their competitors. HiLo acts as a trusted third party which aggregates data from these companies to improve the health and safety of the industry as a whole. In return for providing data, individual companies receive specialised insights to improve their own safety practices. HiLo has reduced lifeboat accidents by 72 percent, engine room fires by 65 percent, and bunker spills by 25 percent.[54]

The approach adopted by APROCONE and HiLo could be extended to other industries, such as transport, construction, or logistics. These are industries where there are multiple parties producing individual components as part of a wider supply chain, yet data sharing may be inhibited by a need to protect sensitive data such as intellectual property. Other examples include facilitating secure data sharing between financial institutions to detect financial crimes such as fraud or money laundering, where coordination across multiple firms that typically compete for business is required.[55] Further opportunities for data intermediaries to facilitate data sharing may exist across a supply chain, particularly where there may be a push from consumers to ensure they are consuming ethical goods and services.

Data intermediaries can build bridges between data providers in commercially sensitive environments. MK Data Hub, outlined in Box 11, is one such example operating in Milton Keynes as part of a Smart City initiative that brings together public- and private-sector organisations using new technologies to reduce congestion and resource use in the city.[56] Future opportunities in this space could include tackling a lack of data sharing on electric vehicle charging points, which are currently siloed in private company databases. The Office for Zero Emission Vehicles recently consulted on improving the consumer experience at public electric vehicle (EV) chargepoints. This included opening chargepoint data to enable consumers to easily locate a chargepoint that suits their needs.[57] A data intermediary could act as a trusted third party between the companies

[54] James Maddison and Josh D'Addario, "Case study: The value of sharing data for benchmarking and insights", *The Open Data Institute (ODI)*, March 3, 2020.
[55] Future of Financial Intelligence Sharing, *Innovation and discussion paper: Case studies of the use of privacy preserving analysis to tackle financial crime*, January 8, 2021.
[56] Royal Academy of Engineering, *Towards trusted data sharing: Guidance and case studies*, 2018.
[57] Department for Transport and the Office for Electric Vehicles, *The consumer experience at public electric vehicle chargepoints*, February 13, 2021.

involved in order to aggregate the public EV chargepoint data, enabling consumers to understand where charging points are located, their reliability, the connector type, and the charging speed. This may facilitate greater electric vehicle adoption by allowing consumers to more easily integrate their vehicles into their life.

Data intermediaries could also support the growth of new industries, as in the case of the government's Online Safety Data Initiative (OSDI). The OSDI aims to increase access to sensitive data held by different, often competing, companies working to build new tools designed to address harmful content online. The OSDI will test methodologies for providing safe access to high quality data in order to support the development of technologies that identify and remove harmful and illegal content from the internet.[58] Providing safety tech companies and researchers access to such data could foster increased innovation in the sector and more effective tools for combating online harms.[59] Such a model could be replicated in other industries. The OSDI will also create more valuable datasets by increasing the volume of data through aggregating individual organisations' data in a single dataset. There are many cases outside of online safety where the value of larger aggregated datasets that an intermediary could host and manage far exceeds the value that can be extracted from data siloed in each company.

Box 10: Advanced Product Concept Analysis Environment (APROCONE)

Data sharing issues addressed by the data intermediary (descriptions outlined in Table 3

Lack of incentives to share data	X	Legal and regulatory risks	
Lack of knowledge	X	Costs of data sharing and access	
Commercial, ethical, and reputational risks	X	Missed opportunities to use data in the public interest	

Advanced Product Concept Analysis Environment (APROCONE) is a collaboration between private and public organisations including Airbus, Rolls-Royce, academic partners, and supply chain companies in the aircraft industry. It seeks to improve collaborative product design for aircrafts, while protecting participants' intellectual property through a digital platform that allows the secure exchange and sharing of product data.

The industrial data platform operates between consortium partners who are able to control their intellectual property, allowing other parties access to minimally required information to support their own designs. They can choose to add or remove partners, thereby protecting their sensitive commercial information. In addition, consortium partners are able to use their existing analysis tools, with the platform performing the required actions that ensure interoperability between partners, overcoming barriers to efficient and cost effective data sharing.[60] This platform

[58] PUBLIC, *"Online Safety Data Initiative" launches to transform data access for online harms*, February 4, 2021
[59] Centre for Data Ethics and Innovation, *Review into bias in algorithmic decision-making*, November 2020.
[60] Royal Academy of Engineering, *Towards trusted data sharing: Guidance and case studies*, 2018.

improves the efficiency of the product design process by removing the need for regular contract-based negotiations on handing over data, enabling different aspects of the product design process to happen simultaneously.

The data sharing facilitated by APROCONE has enabled an innovative approach to initial aircraft/engine sizing that is at least ten times faster and could deliver significant fuel burn savings. It has led to manufacturing cost savings, and has enhanced design processes by making valuable data available earlier in the design lifecycle.[61]

Box 11: MK Data Hub

Data sharing issues addressed by the data intermediary (descriptions outlined in Table 3)

Lack of incentives to share data	X	Legal and regulatory risks	
Lack of knowledge	X	Costs of data sharing and access	X
Commercial, ethical, and reputational risks	X	Missed opportunities to use data in the public interest	X

MK:Smart Hub is a Smart City initiative that aims to use data generated by the Internet of Things (IoT) devices alongside many other data sources to create new city services and businesses, particularly in the energy, water, and transportation sectors.[62]

The MK Data Hub is an industrial data platform that forms the backbone of MK:Smart Hub. It provides developers access to data from a range of different sources including: local and national open data, data streams from infrastructure networks, data from sensor networks such as weather and pollution data, as well as data crowdsourced from social media and mobile applications. The Data Hub also ensures that data is easily discoverable for its developers, helping to build bridges between different data ecosystems to improve city services and drive innovation. It provides a trustworthy infrastructure offering clear routes to value creation by granting data providers control over who has access to their data and the conditions of using their data.

VivaMK, a smart parking scheme using intelligent camera technology allowing drivers to book a parking space via an app, is the first large-scale commercial development emerging from MK:Smart City.[63] Vivacity Labs has also been used to monitor social distancing during the COVID-19 crisis using its AI sensor network.[64]

[61] Aerospace Technology Institute, *Annual Review*, 2019
[62] Royal Academy of Engineering, *Towards trusted data sharing: Guidance and case studies*, 2018.
[63] Tracsis, *AI investment to end MK traffic jams*, May 17, 2017
[64] Ada Lovelace Institute and the Royal Society, *Learning data lessons: data access and sharing during COVID-19*, November 2020.

Facilitating data sharing to enable independent auditing of data-driven technologies

There is growing recognition and concern about the potential for algorithmic bias to replicate or amplify human biases and particularly affect vulnerable groups.[65] In recent research, both the Brookings Institute and the CDEI highlight that an important source of bias in algorithmic decision-making is the data upon which the model relies.[66] One example of a data intermediary operating in this space is the National Institute of Standards and Technology (NIST) in the United States, outlined in Box 12, which provides sensitive personal information to test the performance of facial recognition algorithms across demographic groups.[67]

As the volume and variety of data used in algorithmic decision-making grows, there is an increasing need to quality-assure the data used. This is an area in which data intermediaries could play an important role by resolving the tension between the need for sensitive personal data to test models for bias and the risk of infringing individuals' privacy and undermining trust.[68] OpenSAFELY is one approach that could possibly be more widely adopted to address this tension. In addition, data intermediaries could potentially license or accredit datasets, assuring data users of the quality of the data.[69]

The CDEI's review into bias in algorithmic decision-making highlighted the need to perform rigorous testing of new technologies to ensure that platforms do not unintentionally discriminate against groups of people. To do this, organisations need access to demographic data, including protected characteristics, to monitor how models perform.[70] However, there are a number of obstacles to organisations collecting this data, including concerns about the likely reaction of the public, and about how to manage the data in a manner that protects individuals' privacy. There are opportunities for data intermediaries to support this space by collecting and managing demographic data on behalf of organisations, reducing the burden on them and giving individuals' confidence that their data will only be used for responsible and ethical monitoring of algorithmic decision-making. Given the sensitivities in this area, the governance structure and technical infrastructure would need to be carefully designed to foster public trust. In this way, data intermediaries can manage the risks that come with technological innovation to protect data providers and maintain their trust.

[65] Nicol Turner Lee, Paul Resnik, and Genie Barton, "Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms", *The Brookings Institute*, May 22, 2019.
[66] Centre for Data Ethics and Innovation, *Interim report: Review into bias in algorithmic decision-making*, July 25, 2019; Nicol Turner Lee, Paul Resnik, and Genie Barton, "Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms", *The Brookings Institute*, May 22, 2019.
[67] Charles Boutin, "NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software", *National Institute of Standards and Technology*, December 19, 2019.
[68] Centre for Data Ethics and Innovation, *Review into bias in algorithmic decision-making*, November 2020.
[69] Ben Snaith, Deborah Yates, Ed Evans, "Assurance, trust, confidence - what does it all mean for data?",*Open Data Institute*, June 2021. (ODI refer to licensing datasets for assurance, while accreditation points to a future opportunity)
[70] Centre for Data Ethics and Innovation, *Review into bias in algorithmic decision-making*, November 2020.

Box 12: NIST Facial Recognition Vendor Testing

Data sharing issues addressed by the data intermediary (descriptions outlined in Table 3)

Lack of incentives to share data		Legal and regulatory risks	X
Lack of knowledge		Costs of data sharing and access	
Commercial, ethical, and reputational risks	X	Missed opportunities to use data in the public interest	X

The National Institute of Standards and Technology (NIST) is a data custodian that manages four collections of databases provided by the State Department, Department of Homeland Security and the FBI.[71] NIST facilitates data sharing by providing a trustworthy infrastructure for sharing sensitive, personally identifying information - photographs as well as personally identifying metadata from tags, indicating subjects' age, sex, race, and country of birth - for the purposes of testing the performance of facial recognition algorithms. It provides researchers with access to sensitive personal and demographic data that enables them to quality-assure algorithms for fairness. By facilitating the trustworthy management of photographs from databases, researchers using NIST can avoid scraping test data from online sources, which has undermined public trust and damaged the reputations of the companies involved.[72]

Nearly 200 facial recognition algorithms have quantified demographic differences in their results using NIST.[73] These studies have measured each algorithm's error rate for false positives and negatives as well as the comparative difference in errors for photographs with different demographic tags. This has enabled researchers to significantly expand the evidence base on algorithmic bias and helped developers improve the performance of their algorithms.

[71] Charles Boutin, "*NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*", *National Institute of Standards and Technology*, December 19, 2019.
[72] Olivia Solon, Facial recognition's "dirty little secret": Millions of online photos scraped without consent, *NBC*, March 2019
[73] NIST, *Ongoing FVRT Activities*, April 2021

Section 5. Looking to the future for data intermediaries

Data intermediaries already facilitate increased access to and sharing of data across different sectors. In addition to the use cases explored above, there is considerable scope to apply data intermediaries to drive responsible innovation so that it brings benefits to people and supports economic growth. Three potential use-cases for data sharing are explored in more detail below. They are not intended to be exhaustive, but rather identify the breadth and depth of intermediary activity.

Driving innovation in preventative medicine

Data intermediaries could unlock data access and sharing for preventative medicine by leveraging the vast quantities of data produced by individuals when they interact with networked digital technologies such as smartphones, apps, social media, wearables and loyalty cards, etc.[74] Currently, data that could be used for preventative medicine tends to be siloed across healthcare, financial services, retail, wearable technologies, and social media platforms.

Sharing this data could bring benefits to individuals. It could enable them to better manage their health risks and lower the risk of chronic disease (e.g. diabetes, obesity, etc.). It could also facilitate improvements in healthcare by enabling more rapid and targeted treatments and disease management. In turn, this could extend people's quality of life and could lead to financial savings by avoiding costly clinical care.[75]

However, making sensitive data about people's health available, and providing the ability to link and analyse datasets about individual's health brings privacy and security concerns. Much of this data is not collected in a healthcare context and is not subject to all of the same strict processing requirements as health data collected in medical trials or clinical contexts. Such data must be stewarded in a safe and responsible manner in order to maintain public trust while also enabling innovation. A data intermediary could provide a secure environment within which to aggregate and analyse this data, while also providing effective governance mechanisms.

Preparing workers for the Future of Work

New developments in robotics and machine learning could transform the labour market by automating the tasks and jobs typically performed by humans. Research by the OECD estimates that 12 percent of jobs are currently at risk of automation in the United Kingdom.[76] The McKinsey Global Institute highlights that the automation of tasks within jobs will necessitate transforming the

[74] Ada Lovelace Institute, *The data will see you now*, October 2020
[75] World Health Organisation, *The case for investing in public health: The economic case for prevention*, 2014
[76] Organisation for Economic Cooperation and Development, *What happened to jobs at high risk of automation?*, January 2021.

workforce's skills, moving away from basic cognitive skills (e.g. basic data inputting and processing) and physical and manual skills (e.g. general equipment management) towards advanced technological skills (e.g. programming), social and emotional skills (e.g. leadership and managing others), and higher cognitive skills (e.g. creativity).[77]

Research from the OECD and the International Labour Organisation (ILO) highlights that better matching of employees' skills with available opportunities could deliver improved wellbeing for workers through higher job satisfaction and pay, while employers benefit from increased employee retention and higher productivity.[78] In the Future of Work, these benefits could be enhanced by reducing unemployment among vulnerable workers and transforming the labour market to better meet the demand for specific skills.

Data intermediaries could play an important role in improving matching between employment needs, skills availability, and education data to better prepare workers for the Future of Work. They could act as a trustworthy third party aggregating data from potential employers on job vacancies, data from government departments on educational outcomes and unemployment, and from training providers on skills availability. They could then match potential employees with available opportunities or further training needs. These datasets are currently siloed across these organisations and contain sensitive personal data that needs to be managed with sufficient care.[79] A data intermediary could ensure the data is managed responsibly by adopting an approach similar to OpenSAFELY, Genomics England, or ONS SRS which would maintain the privacy of individuals, while also delivering benefits for society.

Enabling the UK economy to meet its Net Zero targets

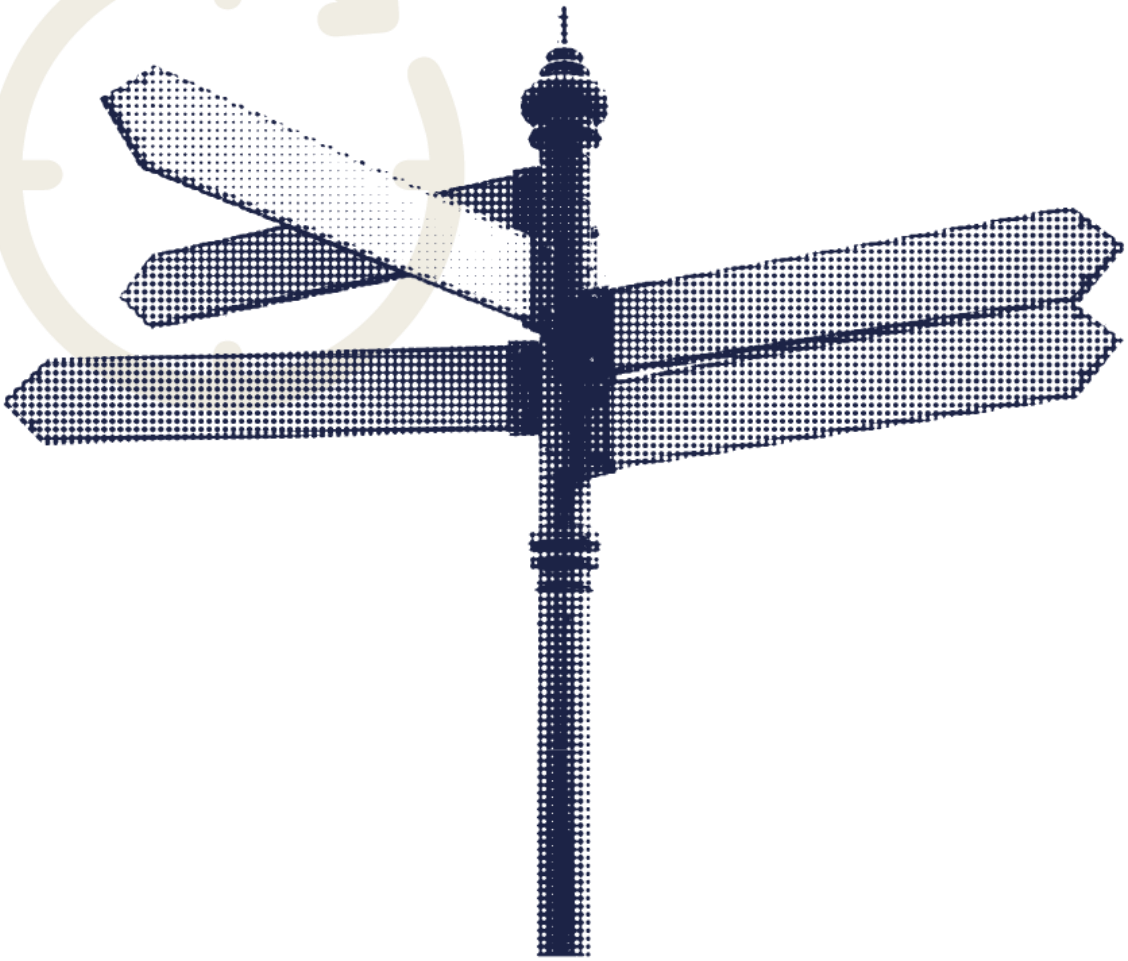
According to research by the Royal Society, using digital technologies could play a major role in driving the carbon emissions reduction the United Kingdom needs to make by 2030.[80] Further research by the Ellen MacArthur Foundation argues that adopting the principles of the Circular Economy - "designing out waste and pollution, keeping products and materials in use, and regenerating natural systems" - could further reduce the UK's carbon emissions, particularly by focusing on urban centres.[81]

Improved data access and sharing is critical to enabling the United Kingdom to meet its Net Zero targets by combining increased knowledge over carbon emissions, greater innovation, and the opportunity to develop the concept of the circular economy. Data intermediaries could play an important role in facilitating these objectives. For example, a data intermediary that adopts a similar approach to that pioneered by Smart Data could enable individuals and businesses to better

[77] McKinsey Global Institute, *Skill shift: Automation and the future of the workforce*, May 2018.
[78] Organisation for Economic Cooperation and Development and the International Labor Organisation, *Better use of skills in the workplace: Why it matters for productivity and local jobs*, 2017.
[79] The exception to these silos is the Higher Education Statistics Authority (HESA), which collects data on employment outcomes for university graduates and acts as a data intermediary between these organisations.
[80] The Royal Society, *Digital technology and the planet: Harnessing computing to achieve net zero*, 2020.
[81] Ellen MacArthur Foundation, *What is the circular economy?*, 2021.

understand their carbon footprint, switch to goods and services that reduce that carbon footprint, and drive innovation in new, more environmentally-friendly goods and services.

In addition, a data intermediary could play an important role in enhancing access to essential climate data, and support innovation focused on tackling climate change. This could include using digital twins of the physical environment to improve resource use and efficiency or facilitating the creation of Circular Cities.[82] They could provide the technical infrastructure that builds bridges between private providers (e.g. energy companies, construction companies) and public providers (e.g. public transport, waste management companies, recycling services, supermarkets and restaurants) to enable city resources to continue to be reused rather than lost. A data intermediary could also facilitate the sharing of sensitive data using privacy-enhancing technologies, while also enabling data providers to have more control over who has access to their data and the potential uses.



[82] The Royal Society, *Digital technology and the planet: Harnessing computing to achieve net zero*, 2020.

Data intermediaries have been proposed as one solution to improving data access and sharing across the UK economy by facilitating trustworthy, safe, and efficient sharing of data, while also protecting individual rights, including preserving privacy. They can unlock opportunities for data sharing that empower individuals and businesses, offering them more control and choice over who has access to data about them, and the purposes for which it is used. In addition, data intermediaries facilitate data sharing for the purposes of analysis. This can include research in the public interest, supporting innovation in commercially-sensitive environments, and enabling the independent auditing of data-driven technologies. This paper explores a range of case studies to highlight intermediary activity that is currently happening, while also examining potential opportunities. Data intermediaries could play an important role in facilitating responsible data-driven innovation that improves lives through greater competition and economic growth. This could include going beyond legal requirements around data protection and applying additional measures to protect against unethical use of data and ensure it is only used for agreed purposes.

Hypothetical pen portrait: Collaborative research environment for the transport sector

"A range of public, private and third sector organisations in the transport sector have a common interest in developing data-driven technologies to improve how they deliver products and services, and to support their operations. Greater access to high-quality transport and mobility data can enable these organisations to train and develop new data-driven technologies, or to enhance existing systems. Furthermore, greater access to live transport data could help inform real-time decision-making.

For example, central government and local authorities could use such data to develop traffic management processes; vehicle manufacturers could develop innovations that improve vehicle safety and performance; environmental organisations could gain a better understanding of the climate impact of different vehicles and traffic management approaches; insurers could develop fairer and more efficient methods for processing claims. Such developments could provide significant benefits for society. However, it may be challenging to share this data due to privacy and competition concerns. Sharing mobility data or data from connected vehicles could violate individuals' privacy, and vehicle manufacturers may be unwilling to share information for fear of disclosing valuable performance data with their competitors.

The organisations come together to create a secure research environment in which they pool their pseudonymised datasets, and are then able to develop systems using this shared data. Each organisation utilises the data through the privacy-preserving technique of federated analytics, meaning they can run analysis against the data without ever being able to see it. The environment is operated by an accredited intermediary organisation that operates as an independent entity, and is funded by the partner organisations."

Bibliography

Ahmed Abdulla, Ewa Janisewska-Kiewra, and Jannik Podlesny, "Data ecosystems made simple", *McKinsey & Company*, March 8, 2021.

Annabelle Gawer, "Bridging differing perspectives on technological platforms: Toward an integrative framework", *Research Policy*, Vol. 43 (7), pp. 1239-1249.

Ada Lovelace Institute, *The data will see you now*, October 2020

Ada Lovelace Institute and the AI Council, *Exploring legal mechanisms for data stewardship*, March 2021.

Ada Lovelace Institute and the Royal Society, *Learning data lessons: data access and sharing during COVID-19*, November 2020.

Behavioural Insights Team, Doteveryone, and Centre for Data Ethics and Innovation, *Active online choices: Designing to empower users*, November 2020.

Bertin Martens, Alexandre de Steel, Inge Graef, Thomas Tombal, and Néstor Duch-Brown, *JRC Digital Economy Working Paper 2020-05: Business-to-business data sharing: An economic and legal analysis*, European Commission, 2020.

Ben Snaith, Deborah Yates, Ed Evans, "Assurance, trust, confidence - what does it all mean for data?",*Open Data Institute*, June 2021

Bob Bailey, "Guest post: How "data intermediaries" can build trust in alternative data", *Open Data Institute*, January 19, 2021.

Centre for Data Ethics and Innovation, *Addressing trust in public sector data use*, July 2020.

Centre for Data Ethics and Innovation, *CDEI publishes polling data on data sharing*, May 2021.

Centre for Data Ethics and Innovation, *COVID-19 repository and public attitudes: 2020 in review*, March 2021

Centre for Data Ethics and Innovation, *Online targeting: Final report and recommendations*, February 2020.

Centre for Data Ethics and Innovation, *Review into bias in algorithmic decision-making*, November 2020.

Competition and Markets Authority, *Online platforms and digital advertising market study final report*, 2020.

CtrlShift, *Data mobility: The personal data portability growth opportunity for the UK economy*, 2018.

Deloitte, *Assessing the value of TfL's open data and digital partnerships*, 2017.

Deloitte, *Genomics in the UK: An industry study for the Office of Life Sciences*, 2015.

Department for Business, Energy, and Industrial Strategy, *Next Steps for Smart Data: Putting consumers and SMEs in control of their data and enabling innovation*, September 2020.

Department for Business, Energy, and Industrial Strategy, *Smart data working group*, 2021.

Department for Business, Energy, and Industrial Strategy and the Department of Health and Social Care, *Genome UK: The future of healthcare*, September 26, 2020.

Department for Digital, Culture, Media, and Sport, *National Data Strategy: Policy paper*, May 9, 2020.

Department for Transport and the Office for Electric Vehicles, *The consumer experience at public electric vehicle chargepoints*, February 13, 2021.

Diane Coyle, Stephanie Diepeveen, Julia Wdowin, Lawrence Kay, Jeni Tennison, *The value of data: Policy implications*, February 2020.

Digital Competition Expert Panel, *Unlocking digital competition: Report of the Digital Competition Expert Panel*, March 2019.

Ellen MacArthur Foundation, *What is the circular economy?*, 2021.

European Commission, *Analytical Report no.15: High-value datasets: Understanding the perspective of data providers,* 15 July, 2020.

European Commission, *Impact assessment report and support study accompanying the proposal for a regulation on data governance*, November, 2020.

Fionntán O'Donnell and Rebecca Ghani, "Matchmakers of the data world: How 'data intermediaries' can bring decision makers and data together to help combat Covid-19", *Open Data Institute*, October 13, 2020.

Frontier Economics, *Economic impact of trust in data ecosystems: Report prepared for the ODI*, February 2021.

Frontier Economics, *Increasing access to data across the economy*, March 2021.

Future of Financial Intelligence Sharing, *Innovation and discussion paper: Case studies of the use of privacy preserving analysis to tackle financial crime*, January 8, 2021.

Government Data Quality Hub, *What is data quality?*, May 6, 2021.

Ipsos MORI, Centre for Data Ethics and Innovation, and Sciencewise, *Public attitudes towards Online Targeting: A report by Ipsos MORI for the Centre for Data Ethics and Innovation and Sciencewise*, February 2020.

Jack Hardinges, "What is a data trust?", *Open Data Institute,* July 2018.

Jack Hardinges and Dr. Jared Robert Keller, "What are 'bottom-up' data institutions and how do they empower people?", *Open Data Institute*, June 25, 2021.

James Maddison and Josh D'Addario, "Case study: The value of sharing data for benchmarking and insights", Open Data Institute (ODI), March 3, 2020.

Matthias Kässer, Andreas Tschiesner, and Thibaut Müller, "Competing in a world of digital ecosystems", *McKinsey Quarterly*, February 1, 2018.

McKinsey Global Institute, *Skill shift: Automation and the future of the workforce*, May 2018.

Michael Tisne, "Collective data rights can stop big tech from obliterating privacy", *MIT Technology Review*, May 25, 2021.

Mohammad Aaser, Kumar Kanagasabai, Marcus Roth, and Asin Tavakoli, "Four ways to accelerate the creation of data ecosystems", *McKinsey & Company*, November 23, 2020.

Nathan Bookbinder-Ryan, "Empowering the user to make active online choices", *Centre for Data Ethics and Innovation Blog*, November 25, 2020.

Nicol Turner Lee, Paul Resnik, and Genie Barton, "Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms", *The Brookings Institute*, 22 May, 2019.

Open Data Institute, *Mapping the wide world of data access*, 2021.

Open Data Institute and Fingleton, *Open Banking, preparing for lift off: Purpose, progress, and potential*, 2019.

Organisation for Economic Cooperation and Development and the International Labor Organisation, *Better use of skills in the workplace: Why it matters for productivity and local jobs*, 2017.

Organisation for Economic Cooperation and Development (OECD), *Enhancing access to and sharing of data: Reconciling risks and benefits for data re-use across societies*, November, 2019.

Organisation for Economic Cooperation and Development, *Measuring the economic value of data and cross-border data flows: A business perspective*, August 2020.

Organisation for Economic Cooperation and Development, *What happened to jobs at high risk of automation?*, January 2021.

Professor Dame Wendy Hall and Jérôme Pesenti, *Growing the Artificial Intelligence industry in the UK*, 2017.

Royal Academy of Engineering, *Towards trusted data sharing: Guidelines and case studies*, 2018.

Shota Ichihashi, "Competing Data Intermediaries", *Rand Journal of Economics* (Forthcoming).

The Royal Society, *Digital technology and the planet: Harnessing computing to achieve net zero*, December 2020.

UK Health Data Research Alliance, *Trusted Research Environments: A strategy to build public trust and meet changing health data science needs*, 2020.

World Health Organisation, *The case for investing in public health: The economic case for prevention*, 2014.