Public Sector Personal Data Sharing: Framework and Principles

August 2023

Scottish Government
Riaghaltas na h-Alba

THE UNIVERSITY
of EDINBURGH

Public Sector Personal Data Sharing: Framework and Principles.

December 2021

Overview

In this report, we look in detail at frameworks and practices of providing access to personal data by public sector organisations to private organisations. Currently this practice is extremely rare as it involves considerable legal, moral or ethical risks, including damage to public trust in the public sector. Data protection laws, such as GDPR, serve to balance knowledge exchange and innovation with the protection of personal data, and it is important that one is not abandoned in pursuit of the other.

This report is focused on data held within the Scottish public sector and therefore on the pathways for data sharing that are permissible under GDPR legislation. GDPR requires that 1) consent be obtained from the data subject to allow the processing of their data for new secondary purposes or else 2) the data controller must find another legal basis through which to justify the reuse of the data. In this report, we outline two broad pathways adopted for facilitating personal data sharing by public sector to private sector: first and most common, data sharing agreements, and second, extra legislation surrounding data sharing. We also discuss a third pathway being developed for AI applications.

Across the UK, Europe and other countries around the world, the most common pathway for sharing personal data involves identifying public interest and subsequently drawing up a data sharing agreement. This pathway is currently predominantly used to facilitate sharing personal data held by the public sector to accredited research organisations. For instance, in the UK, the Office for National Statistics Secure Research Services provides access to anonymized unpublished data to accredited researchers, and in Scotland, access to health data is managed through the NHS Scotland Data Safe Haven. At present, there are very few examples of personal data held in the public sector being shared with the private sector, and where this has been done, or is being proposed, the pathway to enable it is the same as that currently used for giving access to researchers (some examples of these public-private agreements are in section 4b).

The second pathway, extra legislation, is often needed to further facilitate or restrict personal data sharing, given the legal requirements outlined in GDPR and the DPA 2018. The Serious Crime Act 2007, for instance, *permits* the disclosure of personal data and sensitive data by public authorities to specified anti-fraud organisations for the prevention of fraud[56]. The Commissioners for Revenue and Customs Act 2005 *limits* data sharing by HMRC to others. Outside of the UK, in Finland the Act on the Secondary Use of Health and Social Data provides a separate legal framework for reusing health and social data, with an amended the pathway for the sharing personal data held by public sector. A separate permit authority will be set up – Findata – that will enable a centralized system for the issuing of the data requests and permits, rather than requiring sharing agreements with each data controller (as is the case in the UK). This Finnish framework is still in the relatively early stages with permits so far only issued for the secondary use of healthcare data.

Finally, the application of AI is creating new demands for larger-scale data sharing, and as such is creating demand for an alternative third pathway to data sharing. Much of the legislation governing the use and application of AI is still in the early stages. Where AI technology has been used for the analysis of big data in healthcare, data has either been accessed through an individual data sharing agreement (as in DeepMind's access to eye scan data), or it is based on obtaining consent from the individual whose data is being accessed. New draft legislation from the EU may provide scope for an alternative model. The proposed AI Act includes details for the development of AI regulatory sandboxes, detailing terms for the re-use of personal data in the sandbox. The AI Act is still under discussion by European member states, however it nonetheless represents potential for a separate pathway for the re-use of personal data in AI applications.

While personal data held by the public sector may have the potential for generating great insights, there remain technical barriers to data sharing, as when data do not harmonise across agencies, and legal barriers, especially where actions are governed by multiple areas of legislation (as is the case with health data). Much of the public around the world remain concerned about the use of their personal data by public authorities and about private sector access to data. Existing pathways for data sharing with researchers – and by implication, other parties – can be improved by creating shared data standards and protocols across agencies, demonstrating public value and involving the public in the designs of infrastructure and data sharing models, marketing the value of data sharing to immediate stakeholders and users, developing a central resource that facilitates data sharing and makes these procedures transparent, and sharing ethical standards and best practices internationally.

Table of Contents

Table of Contents

1. Introduction

Public sector organisations gather, process, and hold large amounts of data that is needed for the operation of public services and government. In this report, we use the term 'data sharing' to describe those times when data is accessed and reused by third parties for new purposes that are outside of the original purpose for the data collection and processing (Thuermer 2019).

Data sharing can take many different forms, depending on the type of data being shared and the parties involved in the sharing. For example, governments and public sector organisations share large amounts of open data with the general public through open data repositories, such as those run by the UK Government[1], the Scottish Government[2] and ONS[3]. It is important to note that data shared in this way does not contain individually identifiable personal details or information that is commercially sensitive (Scottish Government 2015b). Data sharing can also occur between government organisations (sometimes called 'G2G' (Hamza 2011)), from private organisations to government ('B2G' (Richter 2020)), or from public organisations to private organisations (sometimes called 'G2B'). Each of these cases are governed by different codes of practice and laws depending on the type of data that is being shared.

In this report, we look in detail at one type of data sharing: public sector organisations sharing a particular type of data, called personal data, to private organisations. Currently this practice is extremely rare as it involves considerable risks, such as moral or ethical risks, including damage to public trust in the public sector, but also legal risks (Combe 2009), as there are strict laws around the protection of personal data. These laws help safeguard individuals' fundamental human rights, including the right to a private life as described in Article 8 of the Human Rights Act (Equality and Human Rights Commission n.d.).

We present a literature review of three pathways for data sharing around the world and an evaluation of these pathways. The report first defines personal data, then outlines the common legislative frameworks for data sharing that govern the sharing of personal data. Given the restrictions imposed by this legislation, we then outline two broad pathways adopted for facilitating personal data sharing by public sector to private sector: data sharing agreements and extra legislation surrounding data sharing. A third pathway being developed for AI applications is presented in Section 4.

To gain an understanding of how these mechanisms work in practice and to begin to provide some evaluation of these pathways, we have conducted 9 interviews with individuals from the public sector, private sector, and third sector. The findings from these interviews are presented in Section 5. The report ends with a summary of the literature and findings, offering a reflection

[1] https://data.gov.uk/

[2] https://statistics.gov.scot/home

[3] https://www.ons.gov.uk/methodology/geography/geographicalproducts/opengeography

on the next steps and recommendations for future pathways for public sector sharing of personal data.

1. What is 'Personal Data'?

Data that is related to an identified or identifiable individual is called 'personal data'. Our report only looks at the pathways for sharing personal data. Personal data cannot be made freely available in the form of open data unless it is first processed to become fully anonymised, so that individuals are not identifiable. However, fully anonymizing data is difficult as different pieces of data that are each non-identifiable may still be able to be linked together with other datasets to re-identify them (Henriksen-Bulmer & Jeary 2016, Bampoulidis et al 2020).

Personal data can also be 'pseudonymized'. This is where personally identifiable data fields are replaced by unique identifiers. For example, a researcher may need to know which hospital check-in records in a dataset relate to the same individual. Instead of releasing access to individual patient names or NHS/CHI numbers, this information can be replaced by a unique identifier, such as a random number or random sequence of letters and numbers. This process means that fewer personal data is disclosed, and thus reduces the risk of sharing data. However, pseudonymized data is still a form of personal data under GDPR laws (which will be discussed in the next section).

Personal data also includes 'special category data', which cover information relating to a person's race or ethnic origin, sexual orientation, political opinions, religious beliefs, trade union membership, health data (including genetic and biometric data). A full list is given in Table 1. These types of data have been defined under UK GDPR to be particularly sensitive and so require extra protection.

Terms surrounding data and data sharing can sometimes be used in different ways, for example anonymized and pseudonymized can sometimes be used interchangeably. To help clarify how terms are used in this report, Table 1, provides a summary of some of the key terms used throughout this report.

Table 1: Key data terms used in this report

Data Type	Description
Personal data	Data that relates to an identified or identifiable individual. This can be directly identifiable information, or information about individuals that can be indirectly identified through combining it with other information (UK Information Commissioner's Office n.d.b). Understood as such, this includes pseudonymized data.
Special category data	This is a special category of personal data which is defined under UK GDPR as: • "personal data revealing racial or ethnic origin; • personal data revealing political opinions; • personal data revealing religious or philosophical beliefs; • personal data revealing trade union membership; • genetic data; • biometric data (where used for identification purposes); • data concerning health; • data concerning a person's sex life; and • data concerning a person's sexual orientation." (UK Information Commissioner's Office n.d.c).
Data controllers	Data controllers have control over the data purposes and decisions over the processing of personal data (UK Information Commissioner's Office n.d.d).
Data processors	Data processors "act on behalf of, and only on the instruction of, the relevant controller" (Ibid.)
Anonymized data	Data that has had any identifiable information removed from the data. Anonymized data cannot be linked back to an individual, because of this fully anonymized data is hard to achieve.
Pseudonymized data	This is personal data that has undergone further processing to remove direct identifying individual information and replace these with artificial identifiers.

2. Legislation Framing Data Sharing

GDPR and International Frameworks

In the UK, personal data is protected by the UK General Data Protection Regulation (GDPR) (UK Information Commissioner's Office 2018), the Data Protection Act (DPA) 2018 (UK Public General Acts 2018) and, where data is shared with or concerned with individuals in Europe, the EU GDPR (European Parliament 2016). UK GDPR law is based upon the EU GDPR law, which is one of the strictest privacy and security laws in the world. These GDPR laws set out seven key principles for the processing of personal data:

- Lawfulness, fairness and transparency – data must be processed lawfully, fairly and transparently to the data subject
- Purpose limitation – data must be collected for specified, explicit and legitimate purposes
- Data minimisation – only the minimum data necessary for the specified purpose should be collected or processed
- Accuracy – data must be kept accurate and up to date
- Storage limitation – data must be kept for only as long as necessary for the specified purpose
- Integrity and confidentiality – data processing must be done in a way that ensure appropriate security, integrity and confidentiality
- Accountability – the data controller is responsible for demonstrating compliance with these principles

Full details of these principles can be found in Article 5(1) and Article 5(2) of the UK GDPR and Article 5.1-2 of the EU GDPR law. One of the most important features of GDPR is that it is tied to the location of the citizens or residents whose data is being processed, and not to the location of where that data is itself stored. This means that GDPR applies even if the data is being processed or collected outside of the EU or UK, e.g. if a company outside of Europe processes data of EU or UK citizens or residents GDPR law still applies. The effect of this has been that many countries outside of Europe have also updated, or are in the process of updating, their own data protection laws to mirror that of the GDPR framework. For example, see legislation and work in Japan (Kumazawa 2019, Nishimura 2021, Japan Personal Information Protection Commission 2020), Singapore (Singapore Government 2012, Singapore Personal Data Protection Commission 2019, Singapore Competition and Consumer Commission 2019), Australia (Adams & Allen 2014, Australian Government, n.d.).

In the USA, at the time of writing, they are still in the process of updating laws which previously applied to specific sectors to be cross-sectoral as is seen in GDPR (U.S. Federal Data Strategy 2020). Federal legislation in USA is also supplemented by state level legislation in some cases (e.g. California) (Tierney 2019). The U.S. Federal Data Strategy will be similar to the framework adopted in Canada, where there is both federal and state level guidance for data protection.

As many of these other national data protection legislations align with current GDPR legislation and as this report is focused on data held within the Scottish public sector, the discussion below focuses on the pathways for data sharing that are permissible under GDPR legislation.

GDPR and Personal Data Sharing

Given the widespread impact of GDPR laws, we examine in detail the current laws before then going on to outline how data sharing currently operates under these laws. The first two principles of GDPR (Lawfulness, fairness and transparency; and Purpose limitation) regulate the sharing of personal data across the UK and Europe.

Firstly, data must be processed lawfully, as covered in the first principle. Article 6 of the UK GDPR sets out the six lawful bases, at least one of which must apply when processing personal data:

- Consent is given by the individual for the data to be processed for the specified purpose
- The processing is necessary for a contract the data controller has (or is about to have) with the individual whose data is involved
- The processing is needed because of a legal obligation
- The processing is needed to protect someone’s life
- The processing is necessary to perform a public task, be that a task that is in the public interest or one that is part of official functions that are clearly based in law
- The processing is needed for the legitimate interests of the data controller or a third party. Note: This lawful base cannot apply to public authorities’ data processing.

The second principle outlined in GDPR is that of purpose limitation. In full, this principle states that personal data will be:

> collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall not be considered to be incompatible with the initial purposes.[4]

Taken together these principles considerably limit the scope for different pathways of data sharing. In practice, they mean that one of the following must be done: 1) consent must be obtained from the data subject to allow the processing of their data for this new secondary purpose or 2) the data controller must find another legal basis through which to justify the reuse of the data. In the following sections, we describe pathways for data sharing where the consent of the data subject is not possible. In section five, we provide a brief reflection on options that are used where consent is possible.

[4] https://gdpr-info.eu/art-5-gdpr/

When the data being processed is special category data, there are further stipulations set out in the UK GDPR and UK Data Protection Act 2018. First, special category data can only be processed if it meets one of the ten conditions outlined in Article 9 of the UK GDPR, and additional conditions set out in Part 1 and 2 of the DPA 2018. Broadly, these conditions cover data processing for: cases where explicit consent is given; where the data has been made public by the data subject; reasons of substantial public interest; and a variety of sector specific reasons, such as health and social care or research and statistics that have a legal basis. Where reasons of public interest are given, the DPA 2018 outlines 23 public interest conditions, the most appropriate of which should be selected. These conditions are outlined in Box 1

Box 1: Substantial public interest conditions outlined in DPA 2018

The 23 public interest conditions are:

- Statutory and government purposes
- Administration of justice and parliamentary purposes
- Equality of opportunity or treatment
- Racial and ethnic diversity at senior levels
- Preventing or detecting unlawful acts
- Protecting the public
- Regulatory requirements
- Journalism, academia, art and literature
- Preventing fraud
- Suspicion of terrorist financing or money laundering
- Support for individuals with a particular disability or medical condition
- Counselling
- Safeguarding of children and individuals at risk
- Safeguarding of economic well-being of certain individuals
- Insurance
- Occupational pensions
- Political parties
- Elected representatives responding to requests
- Disclosure to elected representatives
- Informing elected representatives about prisoners
- Publication of legal judgments
- Anti-doping in sport
- Standards of behaviour in sport

As well as being able to meet one of the various conditions outlined in the legislation, GDPR also requires deciding if an ‘appropriate policy document’ is needed (UK Information Commissioner's Office, n.d.c), which in some cases will need to outline why it is not possible to get individuals‘ consent for the data processing. In addition, across both special category data and personal data processing, a Data Protection Impact Assessment (DPIA) is usually required as the processing of these data types may be of high risk to the individual. A DPIA outlines the purpose and scope of the data processing, identifies risks to the individual and measures what will be taken to mitigate those risks (UK Information Commissioner's Office, n.d.e) and should be made available to the public.

Finally, when data being processed is personal data, ICO's data sharing code of practice based on GDPR specifies that a data sharing agreement must be drawn up that outlines: the parties involved, the purpose of the data sharing (the aim, why the data being shared is key to achieving those aims, the benefits to the data subjects or society), details of the data that will be shared, and justification for the lawful basis of sharing (UK Information Commissioner's Office, n.d.f). The agreement can also include details of liability, any limitations to data use, and details of the duration of the agreement. In the UK, these agreements are kept in publicly searchable lists maintained by each of the data controllers; for example, see lists maintained by the Department for Education (UK Department of Education 2022) or the NHS (NHS Digital 2020). Data sharing agreements must be made with each of the data controllers that are responsible for the data seeking to be shared. As data in Scotland and the UK are not usually stored within one centralized database, this can mean multiple data sharing agreements will be needed with each of the relevant public bodies.

3. Data Sharing Pathways

Pathway 1: Data Sharing Agreements around the World

This framework for data sharing, where public interest/public benefit is identified and a data sharing agreement is drawn up, is common practice in UK, Europe, Australia, Canada, USA and elsewhere; it comprises what we call the first pathway for data sharing and the most common. However, currently, this pathway is predominantly used for sharing public sector personal data with trusted research centres for academic analysis. For example, Statistics Denmark provides access to anonymized personal data to researchers accessing the data through Danish research environments[5]. Where this data is to be linked to other datasets not held by Statistics Denmark, approval must be sought from The Danish Data Protection Agency (Danish Data Protection Agency, n.d.), the national independent supervisory authority in Denmark who operate a similar role to the Information Commissioner‘s Office (ICO) in the UK.

In USA, a federal statistics research data centre is used to facilitate the sharing of restricted or non-public data that is governed under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) (U.S. Office of the Chief Technology Officer 2018). Individuals must travel to the centres to access the data securely, with any outputs from analysis being reviewed

[5] https://www.dst.dk/en/TilSalg/Forskningsservice

before release to protect against misinterpretation or inappropriate disclosures. As in the UK, records are kept of data access and data disclosures. There are currently 31 Federal Statistical Research Data Centre's located across the USA, operating in partnership with more than 50 research organisations, such as universities, government agencies and not for profit research institutions[6]. In Australia, another example of these research centres is the Secured Unified Research Environment (SURE) (SaxInstitute, n.d.). SURE is a national online workspace used by over 500 researchers and 25 government and health data custodians that facilitates the analysis and sharing of health and human service data.

In the UK, an equivalent environment is operated by the Office for National Statistics Secure Research Services (SRS), which provides access to anonymized unpublished data to accredited researchers. SRS operates in accordance with the five safes framework of safe people, safe projects, safe settings, safe outputs, safe data (UK Office for National Statistics, n.d.a). While many datasets are available via remote access on the SRS, some data may only be accessible in an approved safe setting. While the data provided to researchers is anonymized, accredited processors may carry out data linkage before data is then de-identified and shared (UK Office for National Statistics, n.d.b).

A similar platform operating in Wales is the Secure Anonymised Information Linkage (SAIL) Databank. SAIL enables the robust storage of anonymized, person-based records for the improvement of health and wellbeing and associated services (Sail Data Bank, n.d.). It is accessible only for the purposes of research, providing a secure environment for the analysis of population data.

In Scotland, access to health data is managed through the NHS Scotland Data Safe Haven (NHS Scotland, n.d.). Safe Havens are secure environments for the use of electronic NHS data for research. The data is stored in a de-identified form which can be linked to other datasets, including non-health data, by trained staff. The data remains under the control of the NHS, ensuring compliance with NHS policy and legislation. Safe Havens are required to work to a set of seven principles that govern their operation across Scotland (Scottish Government 2015a). Importantly, one of these principles stipulates that personal data cannot be sold by the Safe Haven nor transferred to a commercial organisation, reflecting concerns over public trust.

At a regional level, DataLoch is a data service that allows academic researchers and health service managers to access health and social care data from the South-East of Scotland (DataLoch, n.d.). Potential users of the data must complete an application process along with relevant accredited training (as used in the Safe Havens). Each application is assessed by NHS employees to ensure that only the minimum amount of data needed to answer the specified research question is requested and that requests are in the public interest. This assessment stage is vital as the data stored within the DataLoch is unconsented, patient level data.

[6] https://www.census.gov/about/adrm/fsrdc/locations.html

In each of the above examples, personal data is first processed to be in an anonymized form that is then moved to the safe research environment where the researcher can access the data for their analysis. An alternative to this is seen in the OpenSAFELY platform. OpenSAFELY is an open-source platform facilitating the analysis of electronic health records in England (OpenSAFELY, n.d.). Currently, the platform is only used to support research that is related to the COVID-19 pandemic, with all activity on the platform publicly logged, including analysis code. Proposals to use the data are reviewed by both NHS England and the OpenSAFELY group to assess the public benefit of the proposals and ensure that the proposals are from accredited analysts. Through collaborations with the organisations that supply and maintain electronic health record systems in England (OpenSAFELY-TPP and OpenSAFELY-EMIS), data can be analysed 'in situ'. This means that researchers first develop their analysis on randomly generated data, which shares the same characteristics as the raw patient data. This first step serves to reduce unnecessary disclosure of individual data. Once the analysis code has been developed and tested, it is then sent to the 'live' data environment, with researchers only seeing the tables and graphs that are the results of their analysis and not the original patient records. This means that patient data remains in the original administrative database where it was first gathered, instead of requiring additional processing to create an anonymized or pseudonymized dataset that is then moved to a secure data store (as is needed for other safe environments).

In sum, across the UK, Europe and other countries around the world, there is a common pathway for sharing personal data which involves the identification of public interest/public benefit and the subsequent drawing up of a data sharing agreement. This pathway is currently predominantly used to facilitate the sharing of personal data from public sector to accredited research organisations (mostly hosted in collaboration with universities). While the overall process is broadly the same, the technical approaches used to manage the data sharing process vary. In the first case, as seen in the DataLoch, and Safe Havens, the personal data is further processed and then stored in a separate safe environment in which researchers can run their analysis. In the second, as seen in the OpenSAFELY platform, data remains in the original database held by the public organisation, with researchers only seeing the results of their analysis returned.

Examples of data sharing agreements with private companies

While at present many of the data centres only allow accredited academic researchers to access and use the data, there is scope for the same pathways to be used to facilitate private sector access to personal data held within the private sector. For example, DataLoch are currently in the process of developing governance to allow researchers from third and private sector organisations to access data extracts from August 2022 (DataLoch, n.d.). This would follow a similar process to that currently adopted for academic researchers with an assessment being made over the legitimacy and benefits to patients or that it is in the public interest.

Another example is the early access release of education data to accredited education analytics suppliers by the Department for Education. Here data from the National Pupil Database is shared with six different accredited suppliers, which includes charitable companies and private companies, to allow them to provide data services for schools and local authorities (UK Department of Education 2022).

At a larger scale, in 2016 DeepMind Health formed a partnership with Moorfields Eye Hospital to access a data set of one million eye scans and related health data, including clinical diagnosis, treatment, model of the eye scanning machine, and patient age (Moorfields Eye Hospital, n.d.). The project aimed to investigate how machine learning could be used to analyze patient eye scans to help improve early detection of eye disease (DeepMind 2016). Early results from the study were published in 2018 (Fauw 2018), demonstrating that the AI technology could match the accuracy of the clinical ophthalmologist after being trained on 14,884 eye scans, and that it could be applied to images from a range of eye scanners. It important to note that the project involved the use of de-personalised data and so consent from individuals was not necessary.

Outside of health data, the National Data Analytics Solution (NDAS) project provides an example of crime and justice data sharing (The Alan Turing Institute 2017). This project is led by the West Midlands Police on behalf of the Home Office to develop a new scalable data analytics capability that would be owned by the UK law enforcement agencies. The project aims to apply the resulting analytics capability to explore issues of modern slavery, including prevention, detection, prosecution, and the safeguarding of victims. In addition, data analytics developed may be applied to other crime and policing issues such as serious violence, organized crime, firearms, domestic abuse and demand and resourcing. To complete this, West Midlands Police are working with Accenture who are the data processor for the project. Personal data is shared with the NDAS in accordance with the Law Enforcement purpose of the Data Protection Act. However, while private sector are involved in the data processing, it is unclear the extent to which data is shared with the company Accenture.

At present, there are very few examples of personal data held in the public sector being shared with the private sector. Where this has been done, or is being proposed, the pathway to enable it is the same as that currently used for giving access to researchers. That is to say, a public interest condition is first identified and then a data sharing agreement is drawn up between the parties involved. As with researcher access, often these proposals will be approved by a nominated panel who assess the application prior to data being shared. This helps to ensure that the five safes as described by ONS (safe people, safe projects, safe settings, safe outputs, safe data are met (UK Office for National Statistics, n.d.a).

Pathway 2: Extra Legislation Surrounding Data Sharing

Given the legal requirements outlined in GDPR and the DPA 2018, extra legislation is often needed to further facilitate or restrict personal data sharing. For instance, the Serious Crime Act 2007 permits the disclosure of personal data and sensitive data by public authorities to specified anti-fraud organisations for the prevention of fraud (UK Office for National Statistics

2015). Nonetheless, the data sharing must still be in line with DPA, meaning that considerations must be made over the fairness, transparency, accuracy and security of the data sharing. Furthermore, many of the existing data shares that operate under this legislation are governed by data sharing agreements, thus the underlying framework for the data sharing remains broadly similar that described in the previous section.

In contrast, the Commissioners for Revenue and Customs Act 2005 limits data sharing by HMRC to others (UK Public General Acts 2005). Under this act, data can only be shared outside of HMRC for a restricted set of reasons, such as the fulfilment of HMRC's functions, legal compliance, public interest or where the individual has given their consent (UK HM Revenue & Customs, n.d.). This sharing is further amended by the Digital Economy Act 2017, which permits under certain conditions, the sharing of non-identifying information (UK Department for Digital, Culture, Media & Sport 2016). Along with facilitating the sharing of limited HMRC data, the Digital Economy Act also sets out conditions for sharing public data for research purposes (UK Public General Acts 2017). However, this still maintains the requirement for compliance with the DPA.

Outside of the UK, in Finland the Act on the Secondary Use of Health and Social Data (Finland Government 2019) provides a separate legal framework for the reuse of health and social data. This act allows the reuse of data for:

- scientific research
- statistics
- development and innovation activities
- steering and supervision of authorities
- planning and reporting duties of authorities
- teaching
- knowledge management

The act has also amended the pathway for the sharing of personal data that is held by public sector (Finland Ministry of Social Affairs and Health, n.d.). As described in the previous section, in the current UK framework, data sharing agreements must be made with each data controller. This creates an administrative burden across multiple organisations. In the new Finnish framework, a separate permit authority will be set up – Findata – that will enable a centralized system for the issuing of the data requests and permits. This will allow those who wish to use data from several different bodies or those who want to use data from Finish health and social care records to make one application. For those seeking to access datasets, a separate data utilisation plan will also be required. Findata will also provide and manage a secure environment through which the data can be accessed (Deloitte 2020). Where the data requested is individual level data, the data will be anonymized or pseudonymized. According to our interview with Findata, external organisations apply to Findata, who would contact the relevant data controllers for the datasets; the data controllers securely send the filtered data to the Findata, which does the linking and combining of multiple datasets, then provides a secure link to the applicant. The value of the service is rooted in the curating, linking and combining of the data, not the raw data itself, as the data does not leave Findata.

This Finnish framework is still in the relatively early stages with permits so far only issued for the secondary use of healthcare data (Zajc 2021). However, it demonstrates an alternative second pathway for the sharing of personal data whereby extra legislation is adopted to enable the sharing of personal data held in the public sector. It is worth noting that in Finland, according to our interview with Findata, citizens must opt-out of their data being used rather than opt-in, so that by default citizen data can be used in this manner.

Pathway 3: Artificial Intelligence (AI) and Data Sharing

The application of AI is creating new demands for larger-scale data sharing, and as such is creating demand for an alternative third pathway to data sharing. As stated in Norway's AI strategy document 'access to high-quality datasets is essential for exploiting the potential of AI' (Norwegian Ministry of Local Government and Modernisation 2020, p.6). However, currently there are very few mechanisms to allow that, given the constraints outlined in earlier sections. The UK AI strategy summarises this challenge:

> "Some of the most valuable data – in terms of its potential for enabling innovation, improving services of realising public sector savings – cannot be made open because it contains nationally critical, personal or commercially sensitive information. This includes data which could be used to identify individuals. Organisations looking to access or share data can often face a range of barriers, from trust and cultural concerns to practical and legal obstacles. To address these issues, we are working with industry to pioneer mechanisms for data sharing such as Data Trusts." (UK Department for Business, Energy & Industrial Strategy 2019, n.p.).

Despite the interest in this area much of the legislation governing the use and application of AI is still in the early stages (European Commission 2022). Where AI technology has been used for the analysis of big data in healthcare, data has either been accessed through an individual data sharing agreement (as in DeepMind's access to eye scan data), or it is based on obtaining consent from the individual whose data is being accessed. For example, the UK BioBank[7] is a large-scale database containing genetic and health information from 500,000 consenting UK participants. Data is anonymized enabling researchers from around the world to access the information.

Another example of big data access can be found in the Sentinel system, which is led by the US Food and Drug Administration (FDA) (Sentinel, n.d.). Like the OpenSAFELY system in the UK, the Sentinel system enables the analysis of data that remains in situ with queries being

[7] https://www.ukbiobank.ac.uk/

sent to organisations which can opt in to return query results. Thus data partners keep control over their data. As in BioBank, any directly identifiable patient data is not shared.

Finally, the Harmony Alliance operate a public-private partnership for the analysis of data for blood cancer research (Harmony Alliance, n.d.). This data is gathered from pharmaceuticals, biobanks, hospitals, interventional and non-interventional trails. From interviews with Harmony alliance, they described their approach as a 'de-facto anonymisation process' where any data provider who submits to their data lake must exclude all personal data identifiers (names, addresses, IDs, etc). The data is passed to a third party for processing and standardising, then is harmonised and introduced onto the platform. Both public and private organisations submit data to the data sharing platform.

While many current AI projects that use personal data do so using consent-based frameworks, new draft legislation from the EU may provide scope for an alternative model. The proposed AI Act includes details for the development of AI regulatory sandboxes, which would provide:

> a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan. This shall take place under the direct supervision and guidance by the competent authorities with a view to ensuring compliance with the requirements of this Regulation and, where relevant, other Union and Member States legislation supervised within the sandbox (European Commission 2021, p. 69, Article 53: 1).

The proposed act provides an article explicitly detailing terms for the re-use of personal data in the sandbox. Article 54 states that:

> "the innovative AI systems shall be developed for safeguarding substantial public interest in one or more of the following areas:
>
> (i) The prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security, under the control and responsibility of the competent authorities. The processing shall be based on Member State or Union law;
> (ii) Public safety and public health, including disease prevention, control and treatment;
> (iii) A high level of protection and improvement of the quality of the environment" (Ibid. pg 70, Article 54: 1)

The article continues by specifying terms for use, including processes for storage, processing logs, data retention and erasure. It is emphasized that the re-use of personal data should only be used where other data types (anonymized, synthetic, or other non-personal data) would not be suitable. At the time of writing, the AI Act is still under discussion by European member states, however it nonetheless represents potential for a separate pathway for the re-use of personal data in AI applications.

4. Issues and Barriers to Personal Data Sharing

As the last two of these pathways are yet to be implemented at scale, there is limited public evaluation of these processes. Given that the proposed pathways for public to private sector sharing would broadly follow the model currently used for the re-use of data by academic research, a short summary of the issues identified from this sector are provided below.

Technical, Legal and Economic Barriers

While personal data held by the public sector may have the potential for generating great insights, there remain technical barriers to data sharing. For example, data held by different public bodies is often not harmonized and scope for analysis can be limited by the quality of the underlying data, such as when they lack metadata and standards (van Panhuis 2014), while different public bodies may have different security requirements (UK Centre for Data Ethics and Innovation 2020).

Alongside this, there is often considerable confusion and uncertainty over the legal frameworks, both from the general public and from those involved in the negotiation of data sharing agreements, especially where actions are governed by multiple areas of legislation (Big Data Value Association 2019, NHS Digital 2006, Savirimuthu 2021).

In Finland, the government has addressed some of these technical and legal challenges through Findata, which manages the technical needs and data context. Findata told us, for instance, that they worked with data controllers to identify local and national data standards and standards across agencies, which they then worked to harmonise.

A final concern are economic barriers, should data sharing require payment to the data controller or other agencies. Our interviewee with Findata expressed concern that costs set by perceived value and demand could limit access to certain stakeholders, such as the public sector, third sector and universities.

Organisational Transparency and Cultural barriers

The re-use of personal data will usually require approval from an ethics council, public benefit panel or similar internal assessment structure. In some cases, these assessment structures may not be required to provide feedback on the application, and the reasoning behind a decision will not be made available to those requesting the data, as described by Adams & Allen (2014) regarding requests made in Australia:

> Government data custodians are ultimately responsible for the confidentiality and security of the health information datasets they hold and for making the final decision on whether or

> not to release such information. Like Human Research Ethics Committee's (HRECs), government data custodians consider the risks and benefits associated with disclosing the information but, unlike HRECs, they are not required to provide reasons for rejecting a proposal. These decisions go largely unscrutinised by any independent authority and unchallenged by researchers. (p. 958)

This situation may be further complicated by different bodies' requiring different approval processes (Cavallaro 2020) or having varying willingness to share data, even when it is legal to do so. This hesitancy can be due to organisational constraints, such as the cost or resources needed to enable sharing, or wider cultural factors, such as concerns over the impact on reputation and potential damage to the public authority (Laurie & Stevens 2016) and different attitudes to risk held by public and private sectors (Mikhaylov et al 2018). The Population Research Network (Australia), for instance, cited the issue of conservative data custodians and controllers who, if in doubt, say "no" to releasing data. PRN said it is usual for custodians and controllers to feel very unsure whether they should or could release data, and that there are widespread assumptions that private sector organisations are not allowed to access public data at all.

In the case of Findata, they created a live website to address such issues of transparency that shows all their applicants, what the application was for, what the decision was (whether approved or rejected), any costs to access, what the time limit of the approved permits was, as well as outputs and outcomes of any project. Our Findata interviewee also recommended an early strategic communication strategy with stakeholders who would use the service, to communicate its purpose and value to research and innovation up front. Our interviewee with Population Research Network (Australia) also recommends creating concrete policies and acts within government that give political accountability to the transparency and availability of data for the public interest.

Public Trust

Factors for organisational reluctance to share data can also be linked to wider issues over public trust. Evidence from surveys continues to suggest that much of the public around the world remain concerned about the use of their personal data by public authorities (The Paypers 2020), and about private sector access to data (Scottish Government 2013, Ipsos MORI Social Research Institute 2016, Street 2021, Biddle 2018). In the UK, the DeepMind and Royal Free London NHS Foundation Trust controversy[8] and the failed care.data project[9], serve as

[8] The Royal Free NHS Foundation Trust were found to have failed to comply with data protection laws by the ICO when they shared around 1.6 million patient details with Google DeepMind as part of a project to develop a diagnosis and detection system for acute kidney injury. The investigation found that patients had not been adequately informed that their data would be used in such a way. For more information on the inquiry and the resulting actions (Hern 2017)

[9] NHS England's care.data programme aimed to create a national database of patient's medical records in England. This would have enabled doctors to access patient details from wherever they were being

reminders of both the risks of data sharing but also the difficulties of successfully communicating and maintaining public trust in sharing personal data (Godlee 2016).

A notable exception to this can be seen in results from Finland, where public trust in others and government is high. In 2019, data from the OECD found that 64% of the Finnish population reported trusting the government, noticeably higher than the OECD average of 45% (OECD 2021). While levels of trust were seen to vary by institution, trust in the civil service and national government was high. Data from other surveys has also previously shown there is a high trust in public social and health services (Sitra 2016) and that a higher proportion would be willing to share information about their health if it were to be used for scientific research than in other European countries (Vanska & Halenius 2019). While assessing public trust in data sharing continues to be hard to measure, these statistics do give some insight into why Finland has been able to implement the extra legislation highlighted in Section 3.

High levels of public trust on their own should not be used as justification for data sharing though, as highlighted by comments made in the ethics advisory report for the National Data Analytics Solution, which stated:

> "people's 'comfort' with something should not be considered to be weighty evidence as far as the ethics of data analytics is concerned or, indeed, in the light of a recognition that the legal rights of individuals (most of whom are likely to be innocent) are involved." (The Alan Turing Institute 2017, p. 15).

At a technical level, confidence and privacy can be maintained through appropriate privacy enhancing technologies, such as federated learning, which lets an authority design machine learning models without needing to store individual's data in a central location. According to the UK government's Centre for Data Ethics and Innovation, the lack of public trust could also be overcome in part by giving citizens more agency to control the data held about them and making clear when it is appropriate and in the public interest to share data, especially when it is shared without consent. A CDEI report on 'Addressing Trust in Public Sector Data Use' calls for 'clear principles for determining what constitutes public interest in the sharing and use of data' and clear rules on protecting privacy even when sharing is deemed in the public interest' (UK Centre for Data Ethics and Innovation 2020, n.p.).

One non-standard approach that can address public trust are consent-based models of data-sharing (Shahaab & Khan 2020). However, there was no evidence to suggest that this is used to facilitate the sharing of data to private organisations. Another example of consent-based sharing is using personal data stores and other personal data service companies. These companies enable individuals to manage their own personal data in secure ways.[10] At present

treated and scope for new research that used anonymized datasets of records from across primary and secondary care settings. However, the project failed to win support from the public or doctors and was ultimately scrapped.

[10] One example of these intermediaries is MyDex (https://mydex.org).

these intermediaries offer specialized services and are not a standard approach for the sharing of personal data that is held by public sector.

From the literature, we find that existing pathways for data sharing with researchers – and by implication, other parties – can be improved by creating more awareness among agencies about the different legislative requirements placed on their data, streamlining and increasing public transparency around the decisions behind accepting or rejecting data sharing agreements, and exploring consent-based models of data sharing that address issues of public trust.

A large part of the interviews also discussed the need for legitimacy and trust across data sharing processes. This ranges from the trust with government itself, to trust in the technical process and also trust that outcomes will be useful, valuable, safe and worth the risks. Our interviews showed that public trust is a matter of national context. In Finland, Findata came about in large part because of the government's long history of using data for the public good and being able to clearly define the value (our interviewee cited that this track record goes back to the cancer registry in 1954, which had clear public value). Nearly all interviewees (GovTech Polska, Mydex, FinTech Scotland and Findata) said data sharing processes and data sharing models should be legitimised by the public and stakeholders through a consultation, at a minimum, or through more advanced methods of civic engagement, such as co-design. Both GovTech Polska and Findata, for instance, carried out public consultations on data sharing infrastructures and strategies; our Findata interviewee favoured meaningful dialogue and conversations with the public, over more performative or transactional consultations.

Many of our interviewees (Population Research Network - Australia), GovTech Polksa and Fintech Scotland) also all expressed a desire to share best ethical practices across national borders, pointing out that this would help create broader harmony of standards and an effective community of practice.

5. Conclusion and Recommendations

This report has outlined three different pathways for the re-use of personal data held by the public sector. Firstly, in accordance with GDPR much of the current sharing of personal data is undertaken through the establishment of data sharing agreements, which must be set up between each of the data controllers. This pathway is predominantly used by academic researchers for the completion of research that is in the public interest. Given the wide-reaching scope of GDPR and the DPA 2018, a second pathway has emerged through the drafting of extra legislation to facilitate the re-use of data in certain circumstances. The example of Findata represents the most innovative approach taken to this. Under Finland's Act on the Secondary Use of Health and Social Data, personal data can be reused for the purposes of development and innovation, widening the scope for the private and third sector involvement. Finally, early draft legislation from the EU Commission suggests there may be future scope for the reuse of personal data through AI sandboxes, which would be applied to crime, public security, public health and safety and environmental issues.

The report has also identified the ongoing issues with current practices of public sector data sharing for purposes of research, including technical barriers related to data quality and harmonisation of datasets across agencies, along with legal barriers when actions are governed by multiple areas of legislation, as is the case with health data. Culturally organisations will often not make the reasoning behind their decision to reject a data sharing request transparent, and they may be hesitant to share data even when legally allowed to. Public trust continues to be a major barrier; to address this issue, we find innovation in consent-based models, such as personal data stores.

We want to conclude by arguing for the importance of exploring pathways that remain in harmony with GDPR. Such a route avoid proposals by a UK government consultation that ran in autumn 2021 (UK Department for Digital, Culture, Media & Sport 2021). This consultation included a wide range of proposed reforms to UK data protection laws and the scope and roll of the Information Commissioner's Officer (ICO). The authors of this report wish to state clearly that while in this report we do acknowledge the difficulties of sharing data under current data protection laws, these difficulties do not warrant the widespread amendment or dismantling of these laws as proposed in the UK Government consultation. As summarized in an excellent chapter by Savirimuthu (2021) that examines AI and data protection laws in healthcare:

> "GDPR helps empower relevant actors in the healthcare environment [… by]: (i) framing the subjects and objects of regulation, (ii) providing mechanisms for determining transparency, accountability and legitimacy (iii) and enhancing the ideals of responsiveness through processes for managing risks from the use of digital health solutions." (pg 8).

Although Savirimuthu is considering data protection laws a healthcare context, we believe these three points of empowerment apply to any context where these laws apply. Data protection laws serve to balance knowledge exchange and innovation with the protection of personal data, and it is important that one is not abandoned in pursuit of the other.

Ultimately, based on this literature review and our interviews, we have pulled out five recommendations for models and approaches for public sector data sharing of private data:

1. Focus on creating shared data standards and protocols across agencies and local and national contexts - a public agency could be dedicated to this role. These data standards should create confidence in the quality of the data as well as the consistency of the data sets.

2. Develop a clear way to define and demonstrate public interest and public value. This includes involving the public meaningfully from early stages in the designs of the data sharing infrastructures and models to generate public licence.
3. From earliest phases, develop ways to market the value and utility of the data sharing infrastructure to immediate stakeholders and users (i.e. researchers, private sector innovators) and be transparent about the risk and opportunities. Ways of doing this include involving stakeholders in the designs of the infrastructure or creating a typology of data and datasets that may be of value.
4. Develop a central resource or agency, such as a data permit authority, that helps aggregate, combine and link data and has the autonomy to decide which permissions to grant, as well as the resources needed to provide quality data. Make the process of this as transparent as possible.
5. Share ethical standards and best practices internationally. Develop and maintain an international community of practice that explores the "Futures of GDPR". Many countries and regions developing data-sharing approaches will have similar needs for support and learnings, which they could do from each other.

List of Organisations Interviewed:

- Population Research Network (Australia)
- M&C SAATCHI (Global)
- Scottish Centre for Administrative Data Research (Scotland)
- GovTech Polska (Poland)
- FinTech Scotland (Scotland)
- MyDex (Scotland)
- Harmony Alliance (France and Spain)
- Findata (Finland)

References

Adams, C., & Allen, J. (2014). Government databases and public health research: Facilitating access in the public interest. Journal of Law and Medicine, 21(4), 957–972.

The Alan Turing Institute (2017). Independent Digital Ethics Panel for Police. Ethics Advisory Report for West Midlands Police. July. https://www.turing.ac.uk/sites/default/files/2018-11/turing_idepp_ethics_advisory_report_to_wmp.pdf

Australian Government - Department of the Prime Minister and Cabinet. (2018) New Australian Government

Australian Government (n.d.). Australian Privacy Principles guidelines. https://www.oaic.gov.au/privacy/australian-privacy-principles-guidelines

Bampoulidis, A., Bruni, A., Markopoulos, I., & Lupu, M. (2020). Practice and Challenges of (De-Anonymisation for Data Sharing. In F. Dalpiaz, J. Zdravkovic, & P. Loucopoulos (Eds.), Research Challenges in Information Science (pp. 515–521). Springer International Publishing. https://doi.org/10.1007/978-3-030-50316-1_32

Biddle, N, Edwards, B, Gray, M, McEachern, S. Public attitudes towards data governance in Australia (2018). Centre for Social Research and Methods, Australian National University. 12. https://csrm.cass.anu.edu.au/research/publications/public-attitudes-towards-data-governance-australia-0

Big Data Value Association (2019). Towards a European Data Sharing Space. April. https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpace%20PositionPaper_April2019_V1.pdf

Cavallaro, F., Lugg-Widger, F., Cannings-John, R., & Harron, K. (2020). Reducing barriers to data access for research in the public interest—Lessons from covid-19. The BMJ. July 6. https://blogs.bmj.com/bmj/2020/07/06/reducing-barriers-to-data-access-for-research-in-the-public-interest-lessons-from-covid-19/

Combe, C. (2009). Observations on the UK transformational government strategy relative to citizen data sharing and privacy. Transforming Government: People, Process and Policy, 3(4), 394–405. https://doi.org/10.1108/17506160910997892

Danish Data Protection Agency (n.d.). What We Do. https://www.datatilsynet.dk/english/about-us/what-we-do

DataLoch (n.d.). FAQs. https://dataloch.org/about-us/faqs

De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., … Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Medicine, 24(9), 1342–1350. https://doi.org/10.1038/s41591-018-0107-6

DeepMind (2016). Announcing DeepMind Health research partnership with Moorfields Eye Hospital. 5 July. https://deepmind.com/blog/announcements/announcing-deepmind-health-research-partnership-moorfields-eye-hospital

Deloitte (2020). D01. Study on public sector data strategies, policies and governance. Data analytics for Member States and Citizens. 15 May. https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/2020-06/DIGIT%20-%20D01%20-%20Study%20on%20public%20sector%20data%20strategies%2C%20policies%20and%20governance%20v3annexes.pdf

Equality and Human Rights Commission (n.d.). The Human Rights Act 1998. https://www.equalityhumanrights.com/en/human-rights/human-rights-act.

European Commission (2022). AI Watch - National strategies on Artificial Intelligence. https://publications.jrc.ec.europa.eu/repository/handle/JRC129123.

European Commission (2021). The AI Act. https://artificialintelligenceact.eu/the-act/

European Parliament (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679

U.S. Federal Data Strategy (2020). 2020 Action Plan. https://strategy.data.gov/action-plan/

Finland Government (2019). Act on the Secondary Use of Health and Social Data (552/2019). https://www.finlex.fi/fi/laki/alkup/2019/20190552

Finland Ministry of Social Affairs and Health (n.d.). Secondary use of health and social data. https://stm.fi/en/secondary-use-of-health-and-social-data

Godlee F. (2016). What can we salvage from care.data? BMJ. 354:i3907 doi:10.1136/bmj.i3907

Hamza, H., Sehl, M., Egide, K., & Diane, P. (2011). A Conceptual Model for G2G Relationships. In M. Janssen, H. J. Scholl, M. A. Wimmer, & Y. Tan (Eds.), Electronic Government (pp. 285–295). Springer. https://doi.org/10.1007/978-3-642-22878-0_24

Harmony Alliance (n.d.). FAQ. https://www.harmony-alliance.eu/bigdata-platform/faq

Henriksen-Bulmer, J., & Jeary, S. (2016). Re-identification attacks—A systematic literature review. International Journal of Information Management, 36(6, Part B), 1184–1192. https://doi.org/10.1016/j.ijinfomgt.2016.08.002

Hern, A. (2017). Royal Free breached UK data law in 1.6m patient deal with Google's DeepMind. *The Guardian*. 3 July. https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act

Ipsos MORI Social Research Institute (2016). The One-Way Mirror: Public attitudes to commercial access to health data. March. https://wellcome.org/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

Japan Personal Information Protection Commission (2020). Act on the Protection of Personal Information. June. https://www.ppc.go.jp/en/legal/

Kumazawa, H. (2019). Joint Statement by Haruhi Kumazawa, Commissioner of the Personal Information Protection Commission of Japan and Věra Jourová, Commissioner for Justice, Consumers and Gender Equality of the European Commission. https://www.ppc.go.jp/files/pdf/310123_pressstatement_en.pdf

Laurie, G. & Stevens, L. (2016). Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom. September. Journal of Law and Society, Vol. 43, Issue 3, pp. 360-392, 2016, Available at SSRN: https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1467-6478.2016.00759.x

Mikhaylov, S. J., Esteve, M., & Campion, A. (2018). Artificial intelligence for the public sector: Opportunities and challenges of cross-sector collaboration. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2128), 20170357. https://doi.org/10.1098/rsta.2017.0357

Moorfields Eye Hospital (n.d.). DeepMind Health Q&A. https://www.moorfields.nhs.uk/faq/deepmind-health-qa

NHS Digital (2020). NHS Digital Data Uses Register. https://digital.nhs.uk/services/data-access-request-service-dars/data-uses-register

NHS Digital (2006). Protecting patient data. https://digital.nhs.uk/services/national-data-opt-out/understanding-the-national-data-opt-out/protecting-patient-data#section-251-nhs-act-2006-approval

NHS Scotland (n.d.). Data Safe Haven. https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens

Nishimura, K. (2021). Diet enacts data bills despite concerns raised over privacy. The Asahi Shimbun: Breaking News, Japan News and Analysis. 12 May. https://www.asahi.com/ajw/articles/14346985

Norwegian Ministry of Local Government and Modernisation (2020). National Strategy for Artificial Intelligence. https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf

OECD (2021). Drivers of Trust in Public Institutions in Finland, OECD Publishing, Paris, https://doi.org/10.1787/52600c9e-en.

U.S. Office of the Chief Technology Officer (2018). The State of Data Sharing at the U.S. Department of Health and Human Services. U.S. Department of Health and Human Services, September. https://www.hhs.gov/sites/default/files/HHS_StateofDataSharing_0915.pdf

OpenSAFELY (n.d.). About OpenSAFELY. https://www.opensafely.org/about/

Parliament of Australia (2020). Data Availability and Transparency Bill 2020. https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r6649

The Paypers (2020). More than half of Europeans don’t trust public authorities with their data. December 21. https://thepaypers.com/payments-general/more-than-half-of-europeans-dont-trust-public-authorities-with-their-data--1246377

Richter, H. (2020). The Law and Policy of Government Access to Private Sector Data (‘B2G Data Sharing’) (SSRN Scholarly Paper ID 3594109). Social Science Research Network. https://doi.org/10.2139/ssrn.3594109

Sail Data Bank (n.d.). Overview. https://saildatabank.com/about-us/

Savirimuthu, J. (2021). The GDPR, AI and the NHS Code of Conduct for data driven health and care technology. In: Personal Data Protection and Legal Developments in the European Union. https://livrepository.liverpool.ac.uk/3098248/

SaxInstitute (n.d.). SURE. https://www.saxinstitute.org.au/our-work/sure/

Scottish Government (2015a). Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics. 16 November.

Scottish Government (2015b). Open Data Stategy. 25 Feb. https://www.gov.scot/publications/open-data-strategy/

Scottish Government (2013). Public Acceptability of Data Sharing Between the Public, Private and Third Sectors for Research Purposes. 4 October. https://www.gov.scot/publications/public-acceptability-data-sharing-between-public-private-third-sectors-research-purposes/pages/1/

Sentinel (n.d.). About the Food and Drug Administration (FDA) Sentinel Initiative. https://www.sentinelinitiative.org/about_us

Shahaab, A., & Khan, I. (2020). Estonia is a ‘digital republic’ – what that means and why it may be everyone’s future. The Conversation. 7 October. http://theconversation.com/estonia-is-a-digital-republic-what-that-means-and-why-it-may-be-everyones-future-145485

Singapore Competition and Consumer Commission (2019). Data: Engine for Growth – Implications for Competition Law, Personal Data Protection, and Intellectual Property Rights. https://www.cccs.gov.sg/resources/publications/occasional-research-papers/data-engine-for-growth

Singapore Government (2012). Personal Data Protection Act 2012. https://sso.agc.gov.sg/Act/PDPA2012#pr12-

Singapore Personal Data Protection Commission (2019). Trusted Data Sharing Framework. https://www.imda.gov.sg/-/media/Imda/Files/Infocomm-Media-Landscape/SG-Digital/Tech-Pillars/Artificial-Intelligence/Trusted-Data-Sharing-Framework.pdf

Sitra (2016). Survey of attitudes to welfare data in Finland. https://www.sitra.fi/artikkelit/survey-attitudes-welfare-data-finland/

Statistics Denmark (2020). Data for research. Addressing trust in public sector data. 20 July. https://www.dst.dk/en/TilSalg/Forskningsservice

Street, J., Fabrianesi, B., Adams, C., Flack, F., Smith, M., Carter, S. M., Lybrand, S., Brown, A., Joyner, S., Mullan, J., Lago, L., Carolan, L., Irvine, K., Wales, C., & Braunack-Mayer, A. J. (2021). Sharing administrative health data with private industry: A report on two citizens’ juries. Health Expectations, 24(4), 1337–1348. https://doi.org/10.1111/hex.13268

Tierney, M. (2019). Data Privacy Laws by State: The U.S. Approach to Privacy Protection. August 27. Https://Blog.Netwrix.Com/. https://blog.netwrix.com/2019/08/27/data-privacy-laws-by-state-the-u-s-approach-to-privacy-protection/

Thuermer, G., Walker, J., Simperl, E. (2019). Data Sharing Toolkit. Data Pitch. https://datapitch.eu/wp-content/uploads/2019/10/7770-Final-Data-Sharing-Toolkit-Web.pdf

UK Centre for Data Ethics and Innovation (2020). Addressing Trust in Public Sector Data Use. https://www.gov.uk/government/publications/cdei-publishes-its-first-report-on-public-sector-data-sharing/addressing-trust-in-public-sector-data-use

UK Department for Digital, Culture, Media & Sport (2021). Data: a new direction. 10 September. https://www.gov.uk/government/consultations/data-a-new-direction

UK Department for Digital, Culture, Media & Sport (2016). Digital Economy Bill Part 5: Digital Government. 5 July. https://www.gov.uk/government/publications/digital-economy-bill-part-5-digital-government

UK Department for Business, Energy & Industrial Strategy (2019). AI Sector Deal. 21 May. https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal

UK Department of Education (2022). External data shares. https://www.gov.uk/government/publications/dfe-external-data-shares

UK HM Revenue & Customs (n.d.). internal manual. Information Disclosure Guide. https://www.gov.uk/hmrc-internal-manuals/information-disclosure-guide/idg40120

UK Information Commissioner's Office (2018). Guide to the UK General Data Protection Regulation (UK GDPR). https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/

UK Information Commissioner's Office (n.d.a). What we do. https://ico.org.uk/about-the-ico/what-we-do/

UK Information Commissioner's Office (n.d.b). What is personal data? https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/

UK Information Commissioner's Office (n.d.c). Special category data. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/

UK Information Commissioner's Office (n.d.d). Controllers and processors. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/controllers-and-processors/

UK Information Commissioner’s Office (n.d.e). Data protection impact assessments. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/

UK Information Commissioner’s Office (n.d.f). Data sharing agreements. https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/data-sharing-a-code-of-practice/data-sharing-agreements/

UK Information Commissioner's Office (2017). Royal Free - Google DeepMind trial failed to comply with data protection law. 3 July. ico-response-national-data-strategy-consultation.pdf

UK Office for National Statistics (n.d.a). Accessing secure research data as an accredited researcher. https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme#the-five-safes

UK Office for National Statistics (n.d.b). ONS research and data access policy. https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/onsresearchanddataaccesspolicy

UK Office for National Statistics (2015). ICO review: Data sharing between the public and private sector to prevent fraud. 16 April. https://ico.org.uk/media/action-weve-taken/audits-and-advisory-visits/1043719/ico-review-data-sharing-to-prevent-fraud.pdf

UK Public General Acts (2018). Data Protection Act 2018. https://www.legislation.gov.uk/ukpga/2018/12/schedule/1/enacted

UK Public General Acts (2017). Digital Economy Act 2017. https://www.legislation.gov.uk/ukpga/2017/30/part/5/chapter/5/enacted

UK Public General Acts (2005). Commissioners for Revenue and Customs Act 2005. https://www.legislation.gov.uk/ukpga/2005/11/contents

United States Census Bureau. Research Data Centres. https://www.census.gov/about/adrm/fsrdc/locations.html

van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., Heymann, D., & Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. BMC Public Health, 14(1), 1144. https://doi.org/10.1186/1471-2458-14-1144

Vanska, R., & Halenius, L. (2019). People value having the power to make decisions about the use of their data. Sitra. January 17 https://www.sitra.fi/en/articles/people-value-power-make-decisions-use-data/

Zajc, T. (2021). From eHealth week: Open data models – enablers for easier secondary use of data. 14 September. https://blog.better.care/from-ehealth-week-open-data-models-enablers-for-easier-secondary-use-of-data

Scottish Government
Riaghaltas na h-Alba

This publication is available at **www.gov.scot**

Any enquiries regarding this publication should be sent to us at

The Scottish Government
St Andrew's House
Edinburgh
EH1 3DG

ISBN: 978-1-80525-778-3 (web only)

Published by The Scottish Government, August 2023

Produced for The Scottish Government by APS Group Scotland, 21 Tennant Street, Edinburgh EH6 5NA
PPDAS1238722 (08/23)