Dr Marcin Z. Zieliński

# Non-personal data sharing initiatives. A comparative approach

## Table of contents

# 1 Key findings of the raport

- The basic substance of all data sharing schemes constituted public sector data. This seems as the only reasonable approach to give the scheme the initial impetus to grow.
- Data sharing schemes, which would not allow private entities to contribute their data (eg. Australia) could have difficulties with being adopted as the preferable solution for distributing data;
- Most national approaches to create a nation-wide data sharing scheme involved a central body which would approve datasets to be included in the scheme, review data-sharing requests and resolve disputes between data custodians and processors; Such approach could be considered safe, as the state eventually retains control and data and its processing. However, as data resources grow, this can lead to a slowdown in the overall system. It also makes it difficult to set up transnational data sharing initiatives, eg. in the framework of cooperation between member states of the European Union.
- India considered implementing a distributed system of specialized data custodians and the application of blockchain technology to verify the various elements of the data exchange system;
- Another interesting concept developed in India was to give the data subjects the possibility to control if their personal data could be anonymized by processors. This level of data ownership is unknown even in highly advanced legal data protection regimes such as the EUs' GDRP.
- The data trust model, considered in India Poland and most extensively in the United Kingdom, could constitute an attractive alternative to regulation. However, this solution carries a strong risk of major market players such as Google and Facebook taking control over the technical and ethical standards of data sharing and processing.
- As shown by the UK approach, even with the data trust model in place, data sharing and processing would still require a legal structure in the form of a contractual model, a equitable trust model, corporate structure or a central regulator. Establishing such legal structure would be difficult and in most cases require intensive legislative intervention.
- The Polish concept of including unprocessed raw data as part of the public domain, thus excluding it from legal protection and making it freely available to everyone, could significantly contribute to the popularity of data sharing schemes. It is however unlikely that such solution would gain support from the biggest data producers. Maybe this is one of the reasons the concept wasn’t eventually implemented.

# 2 Australia – first implementation of a scheme for sharing ‚*non-sensitive data*‘.

## 2.1 History

Australia was one of the first countries to pick up on the idea of developing a comprehensive framework for non-personal data. Already in 2009 the Australian government established a special body, called the Government 2.0 Taskforce, chaired by Nicholas Gruen, the objective of which was to examine solutions to create a more open, accountable, responsive, and efficient government, by using the new technologies already available or emerging at that time. One of the Taskforces' key recommendations was making public sector information open by default, as opposed to the standard closed data repositories available only to the data holder or protected behind a paywall. The Taskforce also recommended establishing a central portal (data.gov.au) that would enable access to and discovery of the data and skills necessary in preparing government information to be released as open public sector information (Engage: Getting On With Government 2.0, 2009, p. 7).

The portal was eventually launched in July 2013 (Bellwood, 2016), running on CKAN, one of the leading open-source data management systems, powering data hubs and data portals, used among others by the governments of Singapore, Canada, the United States, the Helsinki Region and the City of Berlin (https://ckan.org/government).

In December 2015 the Australian Government released the Public Data Policy Statement, according to which data held by the Australian Government was perceived as a strategic national resource. According to this document, all non-sensitive publicly created datasets should be organized with open access as the default solution, in the intention to effectively manage this resource. The aim of the Statement was to lay the foundation for harnessing the value of data, which was perceived as contingent for preserving the country's capacity to remain competitive in the global digital economy (Australia, Public Policy Statement, 2015, p. 1).

Besides making the non-sensitive data open by default, the Strategy identified further conditions crucial for establishing a truly functional and valuable data-sharing ecosystem. Those conditions included:

- making data available with free, easy-to-use, high quality and reliable Application Programming Interfaces (APIs);
- making high-value datasets available for use by the public, industry and academia, in a manner that is enduring and frequently updated using high-quality standards;
- ensuring that non-sensitive publicly funded research data is made open for use and reuse;
- enabling charging fees only for particular, specialized data services;

- building partnerships with the public, private and research sectors;
- enabling secure sharing of data between Australian Government entities;
- engaging by the Australian Governement openly with the States and Territories to share and integrate data;
- preserving the highest standards of security and privacy for the individual, national security, and commercial confidentiality;
- ensuring all new systems support discoverability, interoperability, data and information accessibility, and cost-effective access to facilitate access to data (Australia, Public Policy Statement, 2015, p. 2).

Furthermore, the Strategy identified some minimal technical requirements for the data, which should be upheld to enable sharing by different entities. Those standards include among others:

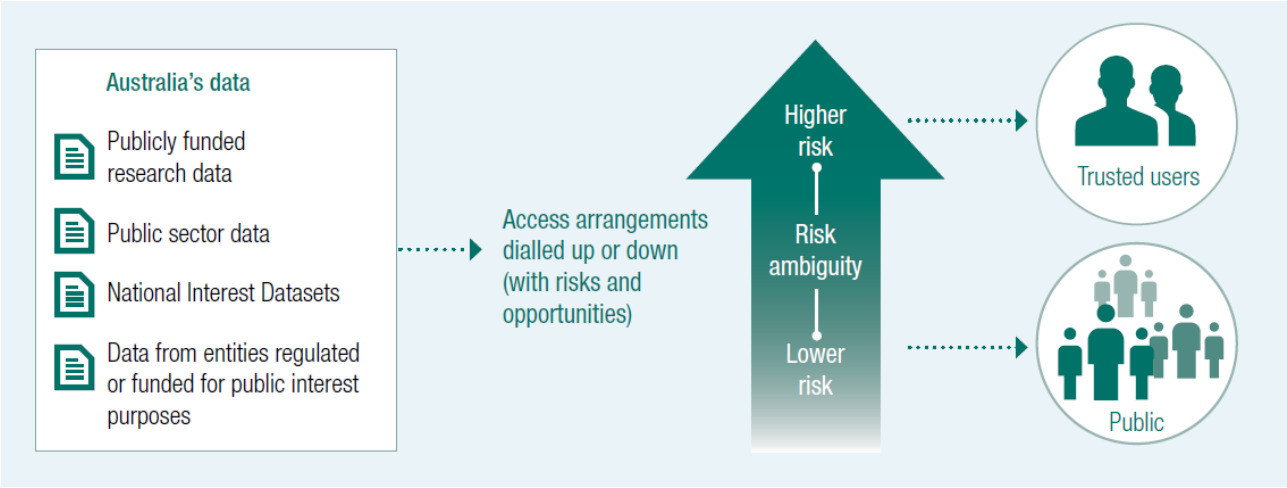
- making data available in a machine-readable, spatially-enabled format;
- employing high quality, easy to use and freely available API access;
- using descriptive metadata, to be able to easily sort out particular datasets;
- using agreed open standards;
- keeping data up to date in an automated way, and
- making data available under a Creative Commons By Attribution license or - if necessary - under another open license (Australia, Public Policy Statement, 2015, p. 2).

Two years later, in 2017 the Productivity Commission proposed a model data-sharing framework that would allow access arrangements to be adjusted according to the different risks associated with different types of data, particular uses and various use environments (Data Availability and Use, 2017, p. 187). The idea behind the project was to shift from the concept of releasing data on request for particular projects, toward creating a coordinated ecosystem of data openly available for processing, subject to specified conditions (Data Availability and Use, 2017, p. 42). According to the project's initial premise, all non-sensitive publicly funded datasets should be made openly available by respective Australian governments (Recommendation 6.1., Data Availability and Use, 2017, p. 52).

Instead of referring to non-personal data, the framework employed the wider notion of 'non-sensitive data', which encompassed both anonymized data that did not identify an individual or breach privacy or security requirements (Australia, Public Policy Statement, 2015, p. 1). Non-senstive data could include anonymized (once personal) data, as well as data which was never personal, such as environmental, weather or spacial data. Hence the grounds for exclusion from the framework of data deemed as sensitive could be either its potential for identifying individuals, or its importance for public safety.

The data subjected to release within the framework of the central data.gov.au portal was to be structured in special datasets, called National Interest Datasets (NIDs). The framework provided for a complex procedure for nominating, assessing and designating datasets as NID;s. After their creation, datasets were to be nominated as potential NIDs to a special public body called the National Data Custodian (NDC). This organ would assess the public interest merits of a particular 'nominee' dataset, present it for consideration to the Australian Government, and refer the thus selected nominations to a public scrutiny process performed by a special parliamentary committee. After positive evaluation the dataset would be designated by the NDC as a National Interest Dataset (Data Availability and Use, 2017, p. 58).

One of the key foundations of the framework was the so called scalable risk based approach. According to its objectives, designated NIDs containing non-sensitive data should be made available for immediate release to the open public. If however a NID included data on individuals (personal data), it would be excluded from open access release. It could however still be made available to selected trusted users, that is ones that were accredited by the relevant Accredited Release Authority (Data Availability and Use, 2017, p. 59).




Picture: Australia, A scalable risk based approach for making NIDs available, available at: https://www.pc.gov.au/media-speeches/articles/pc-news/pc-news-august-2017/data-access

## 2.2 The Data Availability and Use framework

### 2.2.1 General outline

Eventually the bold concept of a central data-sharing platform devised in 2017, including both public data and databases held by private entities, has not found a full implementation. The data-sharing scheme was implemented in 2022 in a somewhat different form, as anticipated by the Data Availability and Use framework.

### 2.2.2 The Data Availability and Transparency Act of 2022

The Data Act, a comprehensive legal framework to authorise the sharing of public sector data for related purposes, was adopted on 31st March 2022. It established a data sharing scheme under which public bodies were authorised to share their public sector data with accredited users who were authorised to collect and use the data, in a controlled way.

The law identifies a number of data sharing purposes, which justify the collection and use of data by the accredited users. Those purposes are:

- delivery of government services, such as providing information and services, paying and determining eligibility for a payment entitlement or benefit;
- implementing government policy and programs;
- research and development (Australia, DATA Act 2022, section 15.1-1A).

Interestingly, some purposes were explicitly excluded as admissible data sharing purposes. The list of precluded applications of the data made available under the sharing scheme includes among others:

- law enforcement (detecting, investigating, prosecuting or punishing offenses, contraventions or other detrimental practices), and
- purposes related to national security, unless such applications only relate in a general way to the above purposes, or do not involve any particular person (Australia, DATA Act 2022, section 15.2-4).

Accordingly state security public entities, such as the Australian Security Intelligence Organisation and the Australian Federal Police are so called 'excluded entities' and may not participate in the sharing scheme (Australia, DATA Act 2022, section 11.3).

The reasoning behind those exclusions was most probably preventing the possibility of abuse, through the use of data by the State, or private entities acting under the authority of the State, for purposes perceived by the public as unduly interfering with the right to privacy, thus addressing widespread concerns that a central data repository could be used by the state to spy on its own citizens.

The law also identifies particular sharing principles, determining the conditions under which data sharing was admissible:

- the shared data may be used only for projects that can be reasonably expected to serve the public interest and only if the data processor observes appropriate ethical standards (the project principle);
- access to data may only be provided to appropriate persons, that is individuals who have attributes, qualifications, affiliations, or expertise appropriate for the access. The National

Data Custodian, as the sharing entity has to verify the data collector, among others with respect to his or her experience with projects involving the sharing of public sector data, capacity to handle such data securely, and previous data breaches (the project principle).

The list of so-called 'accredited users', that is entities that are authorised to process the data includes Commonwealth, State and Territory bodies and Governments, as well as Australian universities. Non-Australian nationals may access the data only if affiliated at an Australian university, having the status of an 'accredited university' (Reverse Explanatory Memorandum, 2022, p. 60).

At the current state of this legislation, private law entities, individuals and unincorporated bodies (such as partnerships and trusts) are precluded from participating in the DATA Scheme. According to the authors, these preclusions are only temporary and intended to provide an opportunity for the Scheme to establish and mature (Reverse Explanatory Memorandum, 2022, p. 3). This would mean that, in due time, the Scheme is intended to be offered as an open platform for all entities, both public and private.

### 2.2.3 The Data Sharing Agreement

According to the Scheme, the data sharing agreement, regulating the conditions of data sharing, must specify all data sharing purposes of a project, and the accredited user is prohibited from using the output for any other purpose, even if it is not a 'precluded purpose' (Reverse Explanatory Memorandum, 2022, p. 35). Penalties are provided for sharing data for purposes other than provided for in the agreement (Reverse Explanatory Memorandum, 2022, p. 45).

## 3 India - subject's consent to data anonymisation and possibility to share private sector data.

A similar concept of a non-personal data sharing framework as in Australia is being prepared for implementation in India. The project is still in the conceptual stage and cannot wait for implementation, which makes it possible to doubt the existence of sufficient political will to do so. However, this does not change the fact that it is based on a number of interesting assumptions and is thus worth analyzing as part of this study

In September 2019 the Ministry of Electronics and Information Technology established the Committee of Experts on Non-Personal Data Governance Framework. A characteristic feature of the Committee was its composition. Representatives of the government side constituted minority in the body, while a dominant role was played by representatives of the commercial and local self-government sectors. As Chairman of the Body was nominated Mr. Kris Gopalakrishnan, a successful entrepreneur in the IT sector, founder of Inforys, a major global digital services and consulting company. The Committee consisted of 8 Members, representing public and private institutions active

in various branches of the data industry, such as commercial services, governement and self-governement administration, as well as NGO's.

The goal of this body was to study various issues relating to non-personal data and to make specific suggestions for consideration of the Central Government on regulation of non-personal data.

### 3.1 Earlier developements

The decision to establish the Committee was preceded by several initiatives, emphasizing the need to regulate the sharing of NPD. In August 2017, TRAI, the Telecom Regulatory Authority of India published a document called ‘Consultation Paper on Privacy, Security and Ownership of the Data in the Telecom Sector’. One of the questions raised in this document was the notion of establishing a 'data sandbox', where TRAI-regulated companies could create anonymized data sets which then could be used for the development of improved services (India, TRAI 2017, p. 24).

In 2018, another study was published, by NITI AAyog, a government founded think tank. The ‘National Strategy for Artificial Intelligence’ suggested that the concentration of data in the hands of a few global players, such as Facebook or Google, could constitute a significant entry barrier for startups (India, NITI AI Strategy 2018, p. 72). The authors of the Strategy also noted, that currently in India data is shared in an informal market, without public supervision. There are also no mechanisms to ensure that the actual data owners have given appropriate permissions before data custodians share the data about them.

NITI AAyog indicated that an effective consent management model could make it possible to establish a voluntary system of compensation of data subjects for sharing data regarding them (India, NITI AI Strategy 2018, p. 74). In such a model, data subjects could make an informed decision if they wish the data collected about them to be shared and on what conditions this could occur. While addressing the importance of providing for an exchange platform for data in order to build AI products, the Strategy also recognized the flaws of establishing a centralized data marketplace. Instead, the solution proposed was to create a distributed data exchange system verified using blockchain technology. Several features were also indicated, which, if provided for, should encourage data providers to share data, such as traceability; access controls; compliance with local and international regulations, and a robust price discovery mechanism for data (India, NITI AI Strategy 2018, p. 79).

A critical role for the success of the future marketplace solution as a universal platform for sharing of non-personal data, not limited to the private sector, plays the involvement of Public Sector Undertakings (PSUs). As one of the major challenges in this regard was indicated in the problem of encouraging and educating such PSUs on the basic standards for data sharing and the essential know-

how, such as: how to provide clean datasets, balance data to sufficiently represent infrequent occurrences, hide test data for evaluation of solutions while sharing training data, define appropriate specs, decide on appropriate evaluation methods, etc. Without sufficient involvement of those public entities, the future platform could have difficulties in achieving a certain critical mass of users, to trigger network effects, which would guarantee its further growth and development (India, NITI AI Strategy 2018, p. 82).

It was therefore proposed to establish a government-endorsed committee of experts, researchers, AI developers and regulators to create the standards for a data marketplace and propose a way to implement them (India, NITI AI Strategy 2018, p. 79). The Kris Gopalakrishnan Committee seems to embody this suggestion.

### 3.2 Key assumptions of the Framework

Eventually, after slightly more than a year of preparations, meetings with representatives of various sectors of business, and hearings of experts, the Committee presented its Report, dated 16th December 2020 (India, Report 2020). The report focused on creating a framework allowing for the sharing of non-personal data originating from government entities, such as anonymized data of land records or vehicle registration; from communal entities, such as datasets collected by municipal corporations or public electric utilities; as well as from private entities, such as datasets collected or generated through proprietary insights, algorithms or knowledge of private companies.

The Framework's main focus is on safeguarding the privacy of data subjects, called 'data principals', which may be individuals, communities of people, as well as companies. This is achieved by limiting the risk of re-identification of such principals, as a result of processing anonymized and/or genuinely non-personal datasets compiled from different sources. The very choice of the term 'data principal' (as opposed to, e.g., data subject) shows, that the Committee's primary objective was to create a mechanism to allow for data sharing, only on the condition that such activity doesn't negatively influence the individual's privacy.

The Committee explicitly declared that in the case of personal data, the rights to process are exercised by the data principal, despite the fact that the Framework only allows the processing personal data after anonymization (India, Report 2020, 7.2.i). As a result, an additional consent refusal mechanism was recommended. At the time of collecting personal data, the collector should inform the individuals about the possibility of subjecting this data to anonymization, offering them a mechanism to opt out of this form of processing. In the case of refusal to anonymization the collector would be prohibited from anonymization of the data for the purposes of making it available within the Framework (India, Report 2020, 5.4.iii).

The Framework thus implements the privacy by design principle. The mechanism for opting out of the anonymization process appears to be modeled on the standard for enabling opt-out of data processing for marketing purposes, which is used in the case of browser cookies within the European Union. There, too, data collectors, fulfilling their statutory obligation to inform about the use of cookies, are obliged to enable website visitors to refuse their consent to the processing of data in this way. The Report also implies the possibility of treating as data principals not only natural persons but also legal entities. However, it remains unclear from how private law entities could express their consent to data anonymization, which by definition can only refer to personal data. Also the "best interests" principle only speaks of the prevention of harm to individuals and communities, without referring to legal persons, such as companies (Bailey 2020, p. 8).

According to the Framework, data custodians will be obliged to organize their procedures for collecting, storing, and processing data in a manner that is focused on the best interest of the data principal, giving him the opportunity to express his opinion in this regard. It should be noted, that the best interest protection requirement, expressed in such generally applicable form, exceeds even the standard set forth in the personal data protection regulations. Even the GDPR, establishing a relatively data-subject friendly processing framework, does not provide for such obligation of the data processor.

To further strengthen the protection of privacy the Committee proposed using a risk-based approach to identify different categories of data with particular privacy risks involved, such as:

- non-personal data which is derived from sensitive personal data (such as health, caste or tribe) which bears a risk of re-identification, and
- data which bears risk of collective harm to a group, and
- data related to national security or strategic interests.

To represent the interests of various data principals, a special category of intermediaries was introduced, called 'data trustees'. Those entities will have the right to recommend transparency and reporting obligations to the regulator, who in turn may adopt respective regulations addressed at data custodians. Trustees are selected with the specific characteristics of the particular data in mind. For example, the Indian Ministry of Health would be acting as a data trustee with respect to the health data of the citizens.

The report seems to start from the idealistic assumption that public entities act solely in the interests of individual citizens. It should be noted however, that in practice, each the decisions of these actors may be influenced by other factors, like the pubic good or budgetary interests. It may also happen that the data custodian will have to balance the interests of a certain person or community with the

conflicting interests of another person or community. The report does not provide guidance on how such conflicts should be resolved (Bailey 2020, p. 10).

Responsible for implementing the entire Framework will be a special expert body, called Non-Personal Data Authority. This authority will have the competence to issue guidelines with respect to data sharing and risks associated with non-personal data. The possibility to request access to the data shared within the Framework will be open to any entity, provided it can demonstrate at least one of the legitimate purposes:

- sovereignty (such as national security or legal requirements);
- public interest (eg. policy making or better delivery of services), or
- economics (eg. creating level playing field for all market players).

In particular, the inclusion of the latter purpose in the Framework makes it an attractive data-sharing platform for businesses, which may encourage them to contribute their data.

The data custodian will have the possibility to refuse a request to access his data. In such situation, the request will be forwarded to the Non-Personal Data Authority for evaluation. It may decide that the social, public or economic benefits associated with the particular data-sharing request justify granting access to the data and thus overrule the data custodians' objection. There are however constitutional doubts, regarding the competence of the central government to incorporate a new regulator, equipped with the competence to govern the entire area of non-personal data on all levels of the public administration (Bailey 2020, p. 12).

As a rule data sharing within the Framework is to be possible free of any charges, not only with respect of public and community data, but also regarding raw and/or factual private data. In the case however of more sophisticated private data-sets, the creation of which was associated with higher costs, there is the possibility to charge a fair, reasonable and non-discriminatory fee. Nevertheless, it is not clear from the Report what is meant by raw and/or factual data, which can make it difficult in practice to determine whether data processing can be carried out free of charge or not. It is unclear for example if anonymized data, the creation of which usually requires substantial effort, may at all be subject to processing without remuneration (Bailey 2020, p. 7).

In return for the possibility to process the shared data, the processing entity will be obliged to share the meta-data it produced based on this data. This is believed to be a significant innovation-encouraging factor. Such a mechanism should also ensure constant growth of the national data repository. It may however also be a hindering factor, because data holders might decide that, allowing others, in particular their competitors, to use the data produced as the result of processing data available within the Framework, constitutes to great a detriment to the company's business

secrets, not justifying the benefits of access to additional data.

The Framework is to be implemented by a special law adopted by the central governement. In the oppinion of the Commitee only such national law can guarantee that the framework will be coherently and effectively enforced at a national scale (India, Report 2020, 7.13.i, p. 21-22). The Report doesn't further elaborate on the assumptions of the proposed legislation, nor does it include any draft provisions. Commentators note however that the Central Government cannot pass a law that affects, data generated by institutions governed by the individual states, such as land or lower-instance judicial records. It also doesn't have the authority to assign competencies and obligations to state bodies. Thus the proposed law runs the risk of being in violation of the constitutional division of powers between the central and local governments (Bailey 2020, p. 11-12).

### 3.3 Assesment of the proposed framework

While the Report is praised for its bold approach to the notion of creating a 'level playing field' for Indian companies to access NPD on an equal basis, it is also criticized for its lack of consideration of other circumstances hindering the development of data economy in India. In particular, its authors are accused that, while designing a platform for cooperation and data exchange by government, local government, and private entities, they fail to address the problem of changing state policies and cumbersome bureaucracy, which can be apparent barriers to the operation of even the best-designed data exchange system (Bailey 2020, p. 3).

The report also lacks an explanation why, in the first place, its authors believe that insufficient access to NPD poses a significant obstacle for Indian companies, impossible to overcome without particular actions from the legislator (Bailey 2020, p. 4). Given the weak condition of the central government in India, it is believed that merely the adoption of a new law and establishment of a new regulator, without further actions promoting and supporting NPD sharing, may not be enough to ensure the success of the framework (Bailey 2020, p. 5).

Eventually, in a situation in which NPD sharing could be carried out under the Framework without sufficient oversight by the regulator, backed by strong government support for his actions, could pose a particular threat to the privacy of data principals (Bailey 2020, p. 4). Faced with processing data practices inconsistent with the Framework, they would be deprived of effective protection. Without proper support from state agencies, data subjects would have great difficulty asserting their rights in connection with the processing of their data, particularly against private entities not subject to public information laws. Thus, a system that was initially designed around the idea of protecting privacy, if the central oversight system failed, could prove to be a significant threat to this core value.

## 4 Japan – data sharing system designed to tackle the countrys' particular

## problems

### 4.1 data-sharing as means for disaster management

The Japanese Basic Act on the Advancement of Public and Private Sector Data Utilization of 2016 is a general document of a rather programmatic nature focused on the declared goal of comprehensively and effectively developing an environment for the utilization and application of public-sector and private-sector data (Japan, Basic Act 2016). The 13-page document describes the general outline of a system by which a variety of data processed through computer networks can be used to seek solutions to problems facing this highly technologically developed country, such as natural disasters or a declining birth rate and aging population.

The declared purpose of this act is to determine the responsibilities of the State, local public entities, and companies by providing for basic principles with respect to the advancement of the appropriate and effective use of public and private sector data (Japan, Basic Act 2016, Article 1). Unlike the frameworks developed in Australia or drafted in India, the Japanese law does not actually refer to any technical solutions to be applied in achieving the regulation's goals. Only the use of technologies related to artificial intelligence, the Internet of Things and cloud computing are vaguely mentioned (Japan, Basic Act 2016, Article 3 (8)).

The Act also strongly emphasizes the need to include as many private sector entities as possible, encouraging "companies" to actively endeavour to advance public and private sector data utilization in relation to their own business activities. This is to be conducted on a voluntary basis (Japan, Basic Act 2016, Article 6).

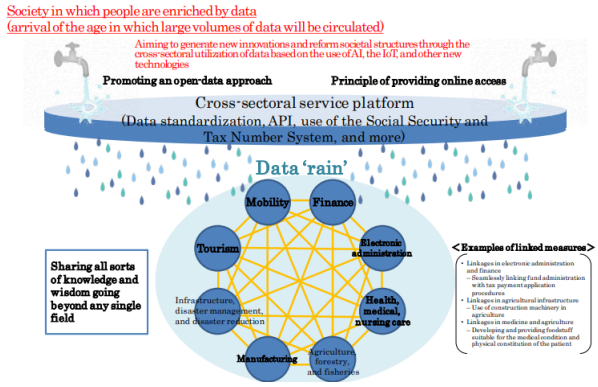
The next step toward implementing the goals envisioned in the Act was the preparation of a ‘Basic Plan for the Advancement of Public and Private Sector Data Utilization’ This document was to cover the following items:

- the basic policy on measures for the advancement of public and private sector data utilization;
- matters concerning public and private sector data utilization by the national government administrative organs;
- matters concerning the promotion of public and private sector data utilization at local public entities and companies;
- measures to be implemented as a priority by the government in relation to public and private sector data utilization;, and
- matters necessary to comprehensively and effectively promote measures for the advancement of public and private sector data utilization (Japan, Basic Act 2016, Article 8).

Such a document was actually developed and published in 2017 under the somewhat lofty title:

"Declaration to Be the World's Most Advanced IT Nation Basic Plan for the Advancement of Public and Private Sector Data Utilization" (Japan, Basic Plan 2017).

The nature-oriented assumptions behind the data sharing concept stem most likely from the country's specific geographical location, strongly affecting its culture. For centuries Japanese people have been adapting to life in a highly tectonically active archipelago. For example, the pool of data available for sharing, which is to be accumulated as a result of the planned measures is depicted as a 'Data rain' feeding life into the various data processing projects, as if they were crops in the field.




(Picture: Society in which people are enriched by data (arrival of the age in which large volumes of data will be circulated); Japan, Basic Plan 2017, p. 26.

The Basic Plan continues the technology-neutral narrative familiar from the Basic Act, avoiding endorsing specific technical solutions. Instead, the document emphasizes that '…) *it is important that we ensure that we are capable of flexibly accommodating environmental changes by constantly upgrading and assuming an agile stance based on an understanding that the environment in which IT operates can undergo occurrences and changes that cannot be imagined at the moment*.'. (Japan, Basic Plan 2017, p. 4).

Rather than proposing solutions, the Basic Plan focuses on pointing out the most important problems Japan is facing, which the document's authors promise to solve by expanding universal data accessibility. Besides foreseeing and coping with natural disasters, among these problems is in particular the progressing aging of the population. Combating this phenomenon requires both an increase in the productivity of the working-age population and solutions that allow for the elderly to once again become active members of a productive society (Japan, Basic Plan 2017, p. 5).

# 5 United Kingdom - education and data trusts to raise confidence in data sharing and empower data subjects to consent to processing their data.

## 5.1 Introduction

In the post-brexit United Kingdom, the Central Digital & Data Office has prepared a strategy to improve data use in government and to emphasize the importance of sharing data (UK, National Data

Strategy 2020). The impetus for the project came from the experiences of the pandemic, when the potential inherent in data and the ability to share it among various private and public entities allowed the British society to adapt to the new threat, effectively preventing the spread of the virus and ensuring the continuity of the economy while allowing people to remain connected.

The National Data Strategy builds on several previous initiatives regarding the development of the data industry. Those are among others:

- 'Industrial Strategy: building a Britain fit for the future' (UK, Industrial Strategy 2017), later replaced by the 'Build Back Better: our plan for growth' (UK, Build Back Better 2021),
- the independent report: 'Growing the artificial intelligence industry in the UK' (UK, Growing the artificial intelligence industry 2017);
- the Policy paper. AI Sector Deal (UK, AI Sector Deal 2019), and
- the UK Research and Development Roadmap (UK Research and Development Roadmap 2020).

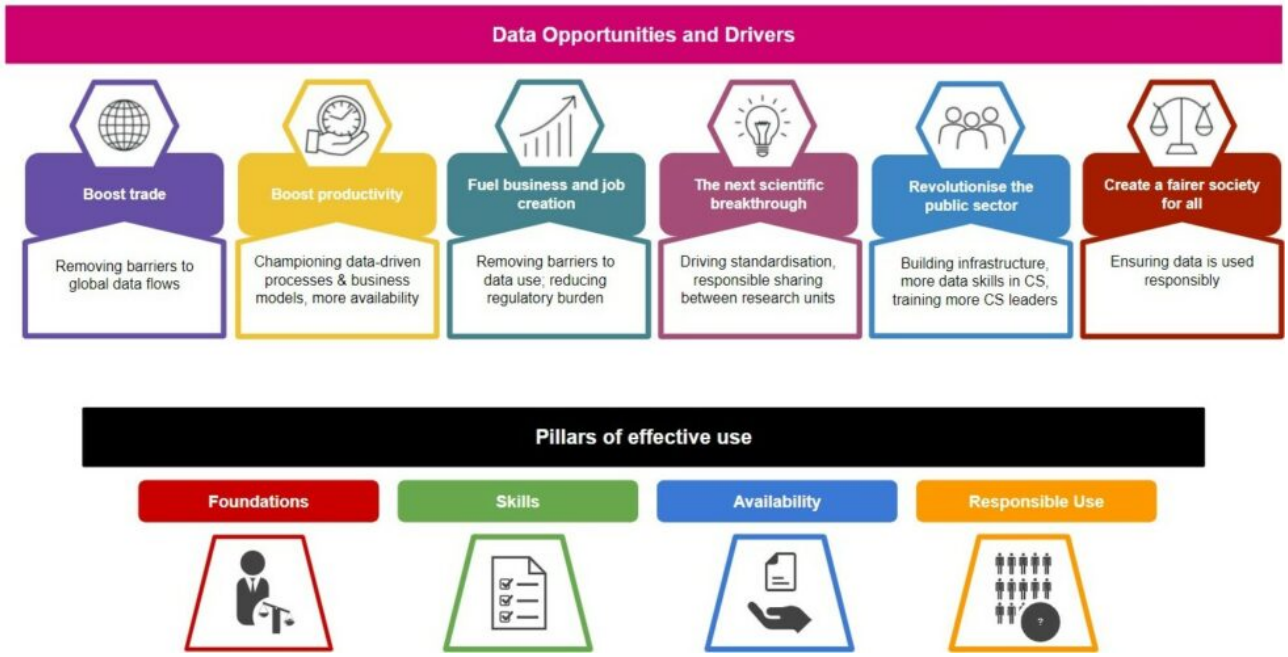
What emerges from the National Data Strategy is an attempt to capitalize on the United Kingdom's unique position as a country drawing on both the legal tradition of the EU, which until recently UK was still part of, as well as being open to global trends in data sharing. Not surprisingly as the starting point for a national data-sharing system, the drafters have designated the resources at the disposal of the government. Interestingly enough, in contrast to the traditional bifurcation into public and private entities, the Strategy distinguishes an additional category in the form of so-called "third sectors". This group includes charities, social enterprises and voluntary groups, engaged in not-for profit activity aiming to create social rather than material wealth (UK charity sector 2022). As typical data access restrictions, which this project intends to remove is the practice of so called 'hoarding of data', usually done by a few global players, enjoying supreme market power. Other problems include incompatible data formats, unclear access rights or failure to make good use of the data already possessed.

The Strategy is build around four core pillars, which were designed to enable the best use of data. Those are:

- data foundations: standards for recording, organizing and storing the data, ensuring that it is findable, accessible, interoperable and reusable;
- data skills: providing an education system, which will enable people to acquire and continuously develop data skills necessary to take advantage of the framework.
- data availability: designing the system in such a way, as to make the data appropriately accessible, mobile and re-usable, by encouraging organisations of the public, private, and third sectors to engage in better coordination of their efforts to collect, access and share data.

- data responsibility: ensuring that data is used in a way that is lawful, secure, fair, ethical, sustainable and accountable, while also supporting innovation and research.

There are several actions aligned with the above pillars, which are supposed to create opportunities for improving trade, boosting productivity, create new jobs, improve scientific research, foster innovative public services, and develop society




(Picture: The National Data Strategy pillars of effective data use and opportunities, https://dcmsblog.uk/2022/04/national-data-strategy-update-2022/ )

The actions foreseen in the Strategy include among others 'securing a pro-growth and trusted data regime', in particular by removing unnecessary burdens, which could hinder an average-size company from accessing and using the data, as well as supporting responsible innovation. As the authors of the Strategy put it: 'The UK's data regime will support vibrant competition and innovation, building trust and maintaining high data protection standards without creating unnecessary barriers to data use' (UK, National Data Strategy 2020).

The basic tenets of the UK's pro-growth legal data regime are an attempt to reconcile the increased freedom of data use with the requirements arising from regulations protecting data and data subjects. Additional guidance and co-regulatory tools are to be provided in particular to small and medium-sized businesses, which might have the most difficulties with taking advantage of the new data regime.

### 5.2 The Centre for Data Ethics and Innovation

The Centre for Data Ethics and Innovation (CDEI), was set up in 2018 as an advisory institution to the Government regarding the use of data-driven technologies and AI. It is described as world's first

body of its kind, whose main goal is to develop best practice guidance for ethical and innovative uses of data. The Centre partners with public and private sector organizations searching for solutions to specific barriers to responsible innovation at an operational level. The developed tools and methodologies are then replicated in other organizations. For example in CDEI is cooperating with Bristol City Council, Police Scotland and the Ministry of Defence, helping to develop an ethical data governance framework to support achieving their data-driven objective (Bancroft, 2020).

Among the actions undertaken by the Centre are: identifying steps to ensure that law, regulation, and guidance are up to date with developments in data-driven and AI-based technologies; issuing recommendations to the Government on ways to support safe and ethical innovation in data and AI, as well as providing expert advice and support to regulators on the implications of the uses of data and AI and areas of potential harm (UK, Consultation outcome 2018). One of the interesting projects, the CDEI is involved in, is the introduction of a responsible data access work program. It aims to distinguish and promote innovative approaches to tackling some of the barriers to data-driven innovations, such as data fragmentation, poor data quality, and insufficient data availability (Durkee, 2022).

CDEI has published the 'Privacy-enhancing technologies: Adoption guide', an interactive site dedicated to helping organizations consider how the adoption of privacy-enhancing technologies (or PETs) could unlock opportunities for data-driven innovation, whilst protecting the privacy and confidentiality of sensitive data. The guide includes the presentation of traditional and emerging PETs, highlighting their qualities and limitations, as well as an interactive question-based flowchart, enabling decision-makers to determine which PETs may be useful for their projects (CDEI Pets Adoption Guide, 2021).

Another interesting document prepared by the CDEI are the ‘Good practices for sharing and processing data’. In this very general document, the Center recommends such measures as standard data sharing agreements, appropriate cybersecurity standards (eg. the ISO27001, or the Cyber Essential Certification of the National Cyber Security Centre's), employing Application Programming Interfaces (or APIs) for making data available and access controls and auditing infrastructure (CDEI Good practices).

Although CDEI is a government body, its critics note the overrepresentation of industry-friendly experts in its composition, starting with the chairman, Roger Taylor, co-founder of health data processing company Dr Foster. Several other members of the Centre heave, either worked for, or participated in PR projects organized by key market players, such as Google or Facebook (Orlowski, 2018)

### 5.3 Data trusts

Not surprisingly, given the industry-friendly composition of the CDEI, instead of searching for ways to regulate the sharing of data, a concept of data trusts has been pointed out as a way to increase the confidence of data holders to share their resources to ensure fair and equitable data sharing between organisations in the private sector, and between the private and public sectors. The Report describes them as: '*Mechanisms where parties have defined rights and responsibilities with respect to shared data – in order to protect sensitive data, facilitate access to data, and ensure accountability'.* (AI Sector Deal 2019, p. 8). Within these trusts, individuals can provide access to their data to trusted third parties and authorize them to act on their behalf (UK, CDEI Independent Report 2020).

Data trusts can constitute a way to make the individual data trustees more active stakeholders, allowing them to benefit more directly from the data sharing and processing (Decode 2018, p. 12). It should however also be borne in mind that overconfidence in ethical and technological standards set by the 'industry', dominated by hudge gobal players, like Google or Facebook, carries the risk of creating illusory protection mechanisms and progressively increasing the scope of processed data. Also keeping in mind that non-personal data can, in combination with other data, lead to the identification of an individual, it would be necessary to ensure in the development of the concept of data trusts the widest possible participation of institutions, organizations and experts acting as guardians of privacy.

The data trust concept, might be an interesting alternative to other data sharing vehicles, however currently it is still at infancy stage. One of the institutions actively researching the risks and opportunities associated with this concept is the Open Data Institute, a UK based non-profit private company founded by Sir Tim Berners-Lee and Sir Nigel Shadbolt in 2012. The ODI's mission is to connect, equip and inspire people around the world to innovate with data. It initiated three data trust pilot programs, regarding:

- urban space, in cooperation with the Greater London Authority and the Royal Borough of Greenwich;
- international illegal wildlife trade, in cooperation with the Wildlabs Tech Hub, and
- global food waste, in cooperation with food and drink manufacturers.

Among the conclusions drawn from these pilots are that each data sharing project requires a separate approach and that an easily repeatable approach to building a data trust is yet to be elaborated. One of the key obstacles to making data trusts simple is the fact that data – and its use by different stakeholders for different purposes – is very context-sensitive, hence the dynamics of each particular processing purpose has to be taken into account (ODI, Data Trusts 2019, p. 13).

The same also applies to the legal structure of each data trust. In particular the GLA and RBG urban space project showed, that it may not be possible to recommend any single form of legal structure which could instantly be applied to form a data trust. A contractual model turned out to be difficult to implement in a multi-party environment. An equitable trust model with a standard set of rules or a central public regulator model, would require an intensive and time-consuming legislative intervention (ODI, BPI Solicitors 2019 p. 6). This lack of a suitable legal regime made it even difficult to suggest a set of templates from which future data trust projects could choose (ODI, Data Trusts 2019, p. 13). The only viable solution identified was the corporate or organisational model, where the data trustees would licence their datasets to a third party organisation, in the form of a limited company (LC), a community interest corporation (CIC) or a partnership (LLP), with a separate legal identity, capable of holding assets and having legal rights assigned (ODI, BPI Solicitors 2019 p. 6).

# 6 Poland – a bold assumption to treat raw non-personal data as part of the public domain?

## 6.1 Introduction

In its efforts to establish conditions for the development of artificial intelligence Poland presented an interesting approach to the issue of processing non-personal data. At the core of this concept was the overwhelmingly logical assumption that the creation of an economy based on the widespread use of AI solutions would not be possible without ensuring an adequate supply of data, based on which the algorithms could generate its results. This involved, in particular raw and unprocessed data, which could be processed for other purposes than the one in which they were collected.

To ensure such universal access to data, initially, the assumption was made that unprocessed machine-generated data should be treated as a common good and thus part of the public domain. Such raw data should therefore be excluded from the protection provided by the intellectual property rights regime (Poland, Polityka 2019, p. 54). The rule would apply to non-personal data generated through human activity, public data or data from the human environment.

As a fundamental development barrier to innovation is treated the phenomenon of so-called data siloing, i.e. closing the data by creating closed ecosystems built by both global digital corporations, as well as other large companies that do not operate in the technology sector, but nevertheless collect large quantities of data (Poland, Polityka 2019, p. 54). Access to this data, and to the markets that emerge through its processing, is limited among other things by contractual terms, business models or interoperability standards. The dominant innovations developed and implemented by these digital leaders block the possibility of new alternatives emerging (Poland, Przemysł +, 2018, p. 19-20).

The concept, developed under the auspices of the Polish Ministry of Digitalisation, involved the

creation of an ecosystem in which competitive advantage does not come from access to data, with the exclusion of other market participants, but from the specific data processing skills and innovation of a particular entity, allowing it to use the data in a better, more effective and attractive way (Poland, Przemysł + 2018, p. 40). The authors also recognized the potential lying in unprocessed data and the fundamental lack of competitiveness between processing it and using a product that was originally created based on that raw data. For example, one of the suggestions expressed in the 2019 Policy was to exclude raw data and unstructured data from the scope of corporate secrecy. Creating opportunities for businesses to share raw data within virtual data repositories (data trusts) was seen as a way to ensure collaboration and increase competitiveness in the marketplace (Poland, Polityka 2019, p. 55). A public procurement system was to be used as a tool to implement the demand for acquiring raw data and the rights to use it for data exchange. A condition for obtaining a contract for publicly funded work was, that the raw data be made available within a central virtual repository irrespective of the obligation to deliver the final product (Poland, Polityka 2019, p. 76).

Unfortunately, most of the bold and progressive demands for the planned data exchange framework were abandoned in the course of further work on the project. In particular, in the final version of the Policy, published in 2020, the demand to automatically include raw non-personal data as parts of the the public domain was not explicitly expressed. Still, among the medium-term goals (to be achieved by 2027), were indicated among others an update of the law to ensure access to data, including sensitive data (e.g., medical data), and the creation of suitable regulatory conditions governing the operation of trusted spaces for sharing such data.

This regulatory framework should be open to accomodate both non-personal and personal data, taking into account, of course, the protection of privacy of data subjects (Poland, Polityka 2020, p. 29). Among the short-term goals (to be achieved by the end of 2023) was the creation of incentives for sharing access to data on a reciprocal basis by entities collecting significant amounts of data of various types (Poland, Polityka 2020, p. 35). The Polish data repository is to be included in the network of similar databases created in other EU countries, such as Belgium, Bulgaria, Czech Republic, Finland, Denmark, Ireland, Latvia, Lithuania, Luxembourg, Netherlands, Slovakia, Slovenia, Sweden, and the United Kingdom (Poland, Industry+ 2018, p. 40-41). These activities were also to be carried out in accordance with plans and initiatives for the development of artificial intelligence at the level of the European Union (Poland, Polityka 2019, p. 37).

### 6.2 Public Data Opening Program

The starting point for the creation of the Polish data trust were the resources gathered by the Public Data Opening Program, which has been running since 2016. The program includes the creation of a single data portal, dane.gov.pl, which collects public sector data accumulated by state authorities.

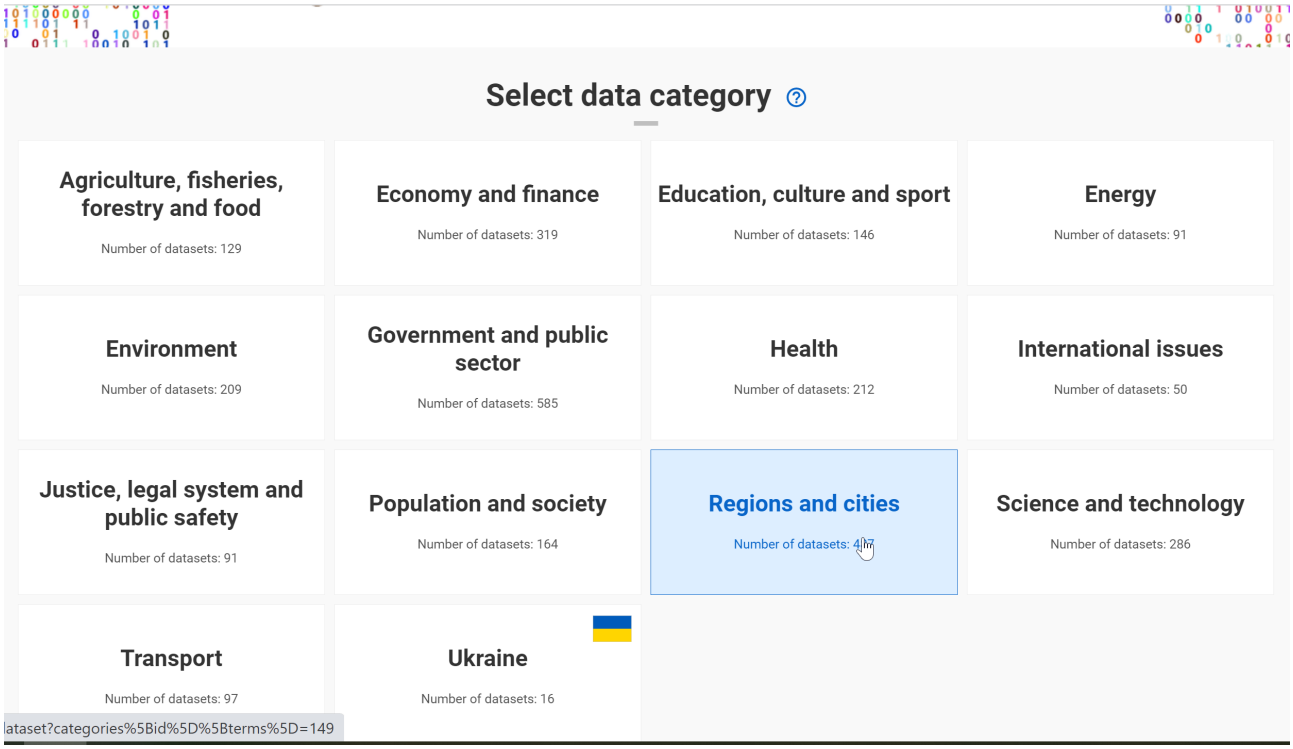
The principles of collecting and accessing data in the Portal were formulated in accordance with the so-called pillars of openness. Thus, the data are to be available to anyone, for any purpose, in original form, saved in commonly used formats, and also complete, up-to-date, and machine-readable. No registration or identity verification is required to access the data. There are also no licensing restrictions on the use of the data or the results of their processing. Access to data should be provided in the highest possible degree of openness and aggregation (Poland, Raport 2021, p. 3).

### 6.3 A network of Data Openness Plenipotentiaries

The Public Data Opening Program and the dane.gov.pl Portal are administered by the Minister for Digitalisation. Additionally, in all ministries, the Prime Minister's Office, the Central Statistical Office and the Central Office of Geodesy and Cartography, plenipotentiaries for data openness have been appointed. This network of plenipotentiaries collaborates to add new resources to the Portal, determine optimal ways to ensure data quality and usability, and share experiences, discuss emerging issues, and consult on training needs (Poland, Raport 2021, p. 8).

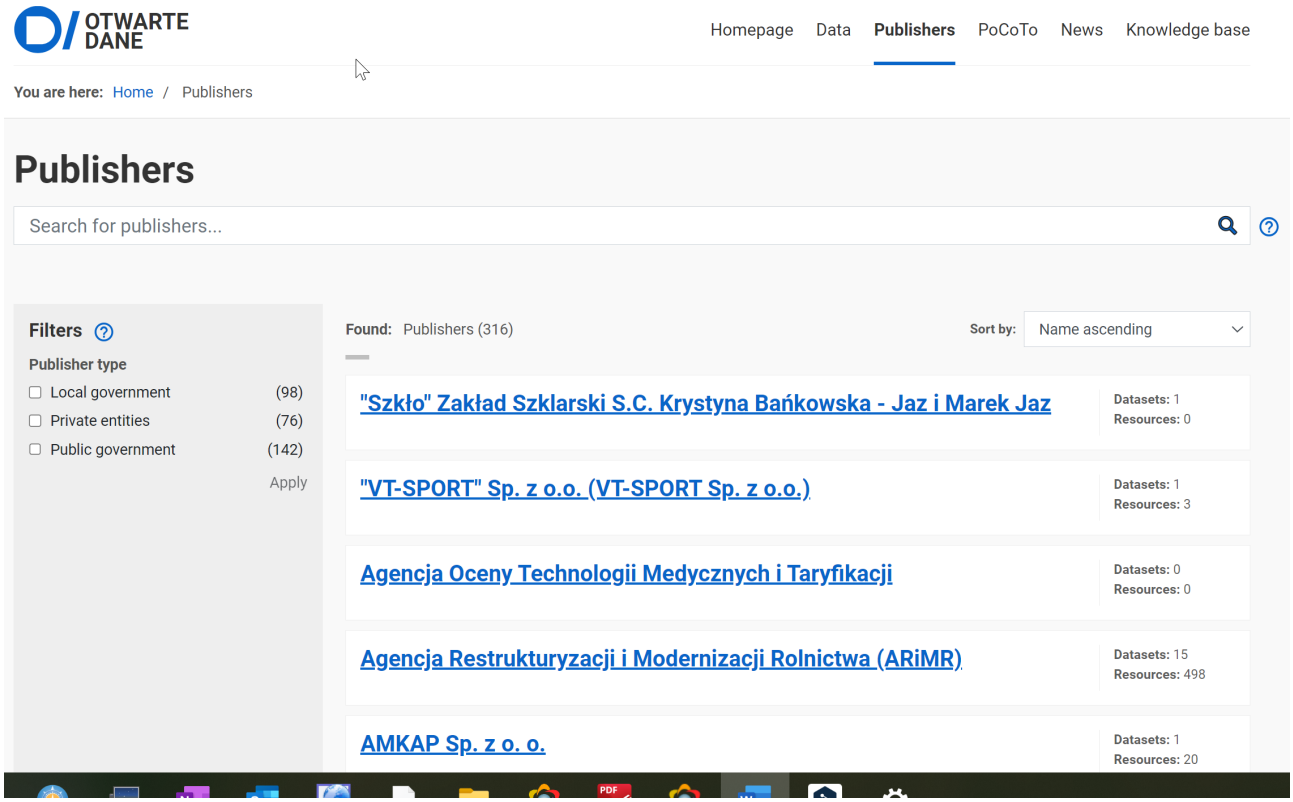
### 6.4 The dane.gov.pl Portal

The portal has been operating since 2014 as the Central Repository of Public Information but has undergone a strong transformation within the framework of the Public Data Opening Program. Additions include among others the ability to submit new data to be made available on the portal, a tool to manage user comments, and modules to support data openness plenipotentiaries from the various public agencies (Poland, Raport 2021, p. 10).



(Picture: Portal dane.gov.pl, https://dane.gov.pl/en).

Currently, the Portal features data provided by 316 publishers, including 76 private entities that make their data available on a voluntary basis. It includes 2829 individual data sets, available for processing via 575 API interfaces




(Picture: The dane.gov.pl Portal, https://dane.gov.pl/en/institution).

## 6.5 Data Openness Standards

In order to streamline the process of opening data, a series of data openness standards have been developed. They were divided into legal, security, technical and API (application programming interface) dimensions. These standards are generally addressed to public entities publishing data on the Portal, but private entities are strongly encouraged to apply them as well. They constitute a set of guidelines, the observance of which will enable the released dataset to be useful to the processor, regardless of the purpose for which the processing is carried out.

Among the legal barriers to making data openly available, licensing restrictions that prohibit a certain use of the data are mentioned first. Sharing data subject to licensing restrictions violates one of the pillars of openness of data. However, since making data available under a license is a legally safer solution than posting it on the Portal without any guidance on the permissibility of its use, the Standards provide for an exemplary listing of fields of exploitation in which the use of the dataset should be permitted to guarantee its use in a way that meets the requirements of openness. The proposed range of permissible fields of exploitation is broad, but nevertheless a far cry from the

full openness that would characterize the use of data in the public domain. For example, the Standard explicitly allows the possibility of authorizing the dissemination of a work on a specific website with a specific name (Poland, Legal Standards 2020, p. 7).

Another interesting element of the Legal Standards is the recommendation that a public administration body, outsourcing a task involving data processing, should contractually ensure its right to the raw data based on which the task is executed. The legal title to such raw data should include, in particular, the possibility of making it available for further processing within the Portal (Poland, Legal Standards 2020, p. 10). These data should be made available in an unprocessed form, not in the form of analyses, summaries, abbreviations or summaries, nor without aggregating or modifying them, so that it is possible to combine data from different sources, depending on the concept of the portal user. Such model would also allow the verification of the final effects of processing (Poland, Legal standard 2020, p. 25). Making source data public has also been identified as one of the pillars of data openness (Poland, Legal standard 2020, p. 41).

### 6.6 Summary

The Polish concept of a virtual repository of data, originating from both public (government and local) and private entities, and available to everyone regardless of the purposes of processing, is a good example of a planned and systematically implemented project. However, there is some dissatisfaction with the abandonment of the bold proposal to treat raw, unprocessed data as a resource exempt from copyright protection and thus entirely exempt from restrictions on further use. Implementing the principle in the Polish framework that would treat raw data as part of the public domain could simplify reuse and significantly improve the legal certainty associated with using the framework. It could also help to solidify this concept as one of the global legal standards for data openness.

# Bibliography


Australia, DATA Act 2022 - Data Availability and Transparency Act no No. 11, 2022, C2022A00011, available at: https://www.legislation.gov.au/Details/C2022A00011/Download

Australia, Public Policy Statement, 2015 - Australian Government Public Policy Statement, 7th December 2015, available at: https://www.finance.gov.au/sites/default/files/2022-10/aust_govt_public_data_policy_statement.pdf

Bailey R., S. Parsheera, R. Sane, Comments on the 'Report by the Committee of Experts on Non-Personal Data Governance Framework'. National Institute of Public Finance & Policy (NIPFP). 13th September 2020, https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3724184_code3108238.pdf?abstractid=3724184&mirid=1

Bancroft, 2020; J. Bancroft, Testing new approaches to responsible innovation. Centre for Data Ethics and Innovation Blog. 17 December 2020, https://cdei.blog.gov.uk/2020/12/17/testing-new-approaches-to-responsible-innovation/

Bellwood T., Working on the new data.gov.au, 6 September 2016, https://blog.data.gov.au/news-media/blog/working-new-datagovau

CDEI Good practices; Centre of Data Ethics and Innovation. Good practices for sharing and processing data, https://cdeiuk.github.io/pets-adoption-guide/good-practice/

Data Availability and Use, 2017 - Australia, Productivity Commissions Inquiry Report, no 82, 31 March 2017

Decode 2018 - Decode. Reclaiming the Smart City. Personal data, trust and the new commons July 2018, https://decodeproject.eu/publications/reclaiming-smart-city-personal-data-trust-and-new-commons.html

Durkee, 2022; M. Durkee, Introducing our responsible data access work programme. Centre for Data Ethics and Innovation Blog. 13 June 2022. https://cdei.blog.gov.uk/2022/06/13/introducing-our-responsible-data-access-programme/

Engage: Getting On With Government 2.0. Government 2.0 Taskforce, December 2009, http://hdl.handle.net/2123/6542

India, NITI AI Strategy 2018, NITI Aayog, National Strategy For Artificial Intelligence, June 2018, http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

India, Report 2020 - Revised Report by the Committee of Experts on Non-Personal Data Governance Framework, https://ourgovdotin.files.wordpress.com/2020/12/revised-report-kris-gopalakrishnan-committee-report-on-non-personal-data-governance-framework.pdf

India, TRAI 2017 - Telecom Regulatory Authority of India, Consultation paper No: 09/2017 on Privacy, Security and Ownership of the Data in the Telecom Sector, New Dehli, 9.08.2017, https://trai.gov.in/sites/default/files/Consultation_Paper%20_on_Privacy_Security_ownership_of_data_09082017.pdf

Japan, Basic Plan 2017, Declaration to Be the World's Most Advanced IT Nation. Basic Plan for the Advancement of Public and Private Sector Data Utilization, May 30th, 2017, https://japan.kantei.go.jp/policy/it/2017/20170530_full.pdf

Japan, Data Utilisation Act, 2016, Basic Act on the Advancement of Public and Private Sector Data Utilization, Act No. 103 of December 14, 2016, https://www.japaneselawtranslation.go.jp/en/laws/view/2975/en

ODI, BPI Solicitors 2019 - ODI, BPI Solicitors, Data Trust for the Royal Borough of Greenwich and Greater London Authority, 2019, https://www.theodi.org/wp-content/uploads/2019/04/BPE_PITCH_GREENWICH_GLA_A4-FINAL.pdf

ODI, Data Trusts 2019 - Open Data Institute, Data Trusts. Lessons from three pilots, April 2019, https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit#heading=h.3fngvdcfo2cs

PETs Adoption Guide 2021; Centre of Data Ethics and Innovation. Pets Adoption Guide. 2021, https://cdeiuk.github.io/pets-adoption-guide/what-are-pets

Poland, Industry+ 2018; Przemysł + Gospodarka oparta o dane [Industry + A data-driven economy], Ministry of Digitalisation, 19.01.2018, https://www.gov.pl/attachment/7cb1aede-e692-40c5-aea3-000fc0ad433d

Poland, Legal standard 2020; Standardy Otwartości Danych. Standard Prawny [Data Openness Standards. Legal

Standard], 2020, https://dane.gov.pl/media/ckeditor/2020/05/29/standard-prawny.pdf

Poland, Polityka 2019; Polityka Rozwoju Sztucznej Inteligencji w Polsce na lata 2019 – 2027. Godna zaufania sztuczna inteligencja autonomia i konkurencja +PL. Projekt dla konsultacji społecznych.[Policy for the Development of Artificial Intelligence in Poland for 2019 - 2027. Trustworthy artificial intelligence autonomy and competition +PL. Draft for public consultation.] 20 August 2019, https://www.gov.pl/attachment/0aa51cd5-b934-4bcb-8660-bfecb20ea2a9

Poland, Raport 2021; Raport nt. rezultatów wdrażania programu otwierania danych publicznych. Okres sprawozdawczy: 01.01-31.12.2020 r. [Report on the results of the implementation of the public data opening program. Reporting period: 01.01-31.12.2020] Ministry of Digitisation, 19.07.2021, https://www.gov.pl/attachment/f53a4f9d-2acf-44e4-9357-e7efc07e5d65

UK, AI Sector Deal 2019; Policy paper. AI Sector Deal, 21 May 2019, https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal

UK, Build Back Better 2021; HM Treasury, Build Back Better - our plan for growth. March 2021, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/969275/PfG_Final_print_Plan_for_Growth_Print.pdf

UK, CDEI Independent report 2020; Independent report. Addressing trust in public sector data use. 20 July 2020, https://www.gov.uk/government/publications/cdei-publishes-its-first-report-on-public-sector-data-sharing/addressing-trust-in-public-sector-data-use#key-findings

UK, Charity sector 2022; J. Smith, Overview of the UK charity sector, November, 2022, https://www.prospects.ac.uk/jobs-and-work-experience/job-sectors/charity-and-voluntary-work/overview-of-the-uk-charity-sector

UK, Consultation outcome 2018 - Department for Digital, Culture, Media and Sport, Consultation outcome. Centre for Data Ethics and Innovation Consultation. 20 November 2018, https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation

UK, Growing the artificial intelligence industry 2017; D. W. Hall, J. Pesenti, Growing the Artificial Intelligence Industry in the UK, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

UK, National Data Strategy 2020; Policy paper. National Data Strategy, 9 December 2020, https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy#about-the-national-data-strategy

UK Research and Development Roadmap 2020; UK Research and Development Roadmap 2020, https://www.gov.uk/government/publications/uk-research-and-development-roadmap