Unleashing the potential of data

Managing data as a shared resource

Report prepared for the Chancellery of the Prime Minister

Blanka Wawrzyniak

Marta Musidłowska

Jan J. Zygmuntowski

Warsaw, August 2022

Fundusze Europejskie
Polska Cyfrowa

Rzeczpospolita Polska

Unia Europejska
Europejski Fundusz Rozwoju Regionalnego

Introductory note:

This publication is the subject of a contract commissioned by the Chancellery of the Prime Minister. It includes a comprehensive discussion of the topic of data sharing with the distinction of four challenges identified by the authors, i.e. data sharing within in the model of: (1) personal data intermediaries; (2) virtual shared repositories of non-personal data ; (3) public shared repositories of personal data; and (4) the management of highly sensitive data.

We recommend quoting:

Musidłowska M., Wawrzyniak B., Zygmuntowski J.J.: “Unleashing the potential of data. Managing data as a shared resource.” – the report based on Robert Kroplewski’s project commissioned by the Chancellery of the Prime Minister.

Written for:

The Digitization of the Chancellery of the Prime Minister

Authors:

Blanka Wawrzyniak, Marta Musidłowska, Jan J. Zygmuntowski

Project initiator:

Robert Kroplewski R.pr., Plenipotentiary of the Minister of Digitization for the Information Society

Contact:

Blanka Wawrzyniak, Leader of the Digital Economy Research Program, blanka.wawrzyniak@instrat.pl

Cover design, illustration on the cover, composition:

Anna Olczak

Project Manager:

Robert Kroplewski R.pr., Plenipotentiary of the Minister of Digital Affairs for the Information Society

Any mistakes are ours. Customary disclaimers apply.

Table of contents

Terms and acronyms

(Compiled for the purposes of this publication)

Artificial intelligence: A machine system that is able to influence the environment by producing output data (predictions, recommendations or decisions) for a given set of purposes. It uses information from machines and/or humans to (i) perceive real or virtual environments; (ii) create abstractions of this perception in the form of models, through analysis, in an automated manner (e.g. using machine learning) or manually; and (iii) use model-based inference to formulate outcome options (OECD, 2019).

B2G: Business to Government: sharing private (business) data with public administration

Big data: Otherwise, big data analytics; these are collections of information of high volume, high variability or high diversity, which require new forms of processing to support decision-making, discover new phenomena and optimize processes (Borowik M., Maśniak L., Kroplewski R., Romaniec H., 2018)

CJEU: Court of Justice of the European Union

Data capsules: Private data silos used by individuals or organizations to collect data on those entities or data generated from devices used by those individuals / organizations.

Data intermediary: A trusted third party to the sharing of data, acting as a mediator between those who want to make their data accessible and those who want to use it; manages data appropriately and assists users in making informed choices about consent to the processing of their data (Janssen H., Singh J., 2022).

Data repository: A place to store and organize documents to share.

Data sharing: A set of practices, technologies, cultural elements and legal frameworks designed to foster the sharing of the value of data by various actors. Sharing can include various ways of processing data, such as providing access and entrusting (under the GDPR), but also sharing access to data (without transfer).

Database: A structured set of systematized information (data), collected according to specific rules; normally stored in a computer system in electronic form. For example, data in the most common types of databases currently used is placed in rows and columns of a series of tables, which streamlines data processing and querying, makes it easier to access data, and allows to manage, control, modify, update, and organize data.

DICOM: The standard defining the format and method of exchange of image data between imaging devices (CT, MRT, digital angiographs or digital X-ray machines) and units used for analysis and secondary processing of data (diagnostic description stations) or archiving systems (infoRadiology)

Digital footprint: A unique set of traceable activities, statements and digital communications of a person that appear on the Internet or in digital devices. Digital traces can be classified as passive or active. The former are records of the user's activity on websites and information stored in cookies. The latter are often created intentionally by the user in order to share information through websites or social media. Although the term usually refers to a person, a digital footprint can also be left by a company, organization, or corporation.

Digital sovereignty: The ability of a state, international organization and each user individually to enforce their rights and influence digital platforms and technology companies in accordance with their own social and developmental needs, for the autonomous shaping of opportunities (Wawrzyniak, Zygmuntowski and Lamański, 2020).

ENISA: EU's Agency for Cybersecurity

EU: European Union

GAFA: The largest American companies in the IT industry: Google, Amazon, Facebook, Apple

GDPR: 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"), OJ L 119, 4.5.2016, pp. 1-88.

Health data: According to the guidelines of the European Data Protection Board (EDPB), these data include:

- information collected by the healthcare provider in the patient's medical records;
- information which has become health data as a result of referring it to other data, which has revealed a medical condition or a health hazard;
- information provided in "self-check" questionnaires by data subjects as part of answering questions about their health (descriptions of symptoms);
- information that has become health data due to the way it is used in a specific context (e.g. information regarding a recent trip or presence in a region affected by CoViD-19).

HL7: The standard for the digital exchange of information in medical environments. The protocols described in this standard apply to the application layer (seventh) of the OSI model. One of them is a communication protocol for exchanging medical data that defines the application-level messages used by several major hospital systems. The main functions of the system include the exchange of messages about data access, data retrieval, data transmission, control and retrieval of results and clinical observations.

Innovation debt: Unjustified costs incurred by the country's economy or by an economic bloc due to the lack of adequate investments in one's own innovations (Borowik M., Maśniak L., Kroplewski R., Romaniec H., 2018). With regard to data sharing, innovation debt means that Polish entrepreneurs lose potential profits or incur additional costs resulting from the lack of access to artificial intelligence tools and innovative solutions. A typical method of reducing the innovation debt is "leapfrogging", i.e. a sudden modernization of an industry lagging behind by investing directly in the best technologies, bypassing transitional and older ones.

IT: Information technology: an industry dealing with the applications of computing technologies

NCAS: National Center for Agricultural Support

NGO: Non-governmental organization

Poor interoperability: The ability of a system or product to interact with other systems or products only within one sector, most often public.

Public shared data repository: A public or private-public (hybrid) entity that manages personal data from various sources of public importance (e.g. health data) with appropriate privacy safeguards. This institution grants various entities (both public and private) the ability to access the collected data on certain terms.

Shared data repository: A trusted institution sharing data in public and common interest. It can take the form of a space for both personal and non-personal data (including those of a special nature). "Virtual shared data repository" means a space for data sharing based on cooperation within a federation of associated entities; provides for the building of distributed data repositories in which members provide access to data on jointly recognized technical, organizational and legal principles (Kroplewski R., 2020). The concept is based on the virtual data repository but the term "shared data repository" emphasizes the common nature of access to data, while the "virtuality" of the shared data repository emphasizes its digital form and a federated model of data exchange using distributed repositories (as opposed to a central repository in which common data is stored). The aim is to enable business intelligence (BI) activities, in particular analytics, most often based on industrial, agricultural and business data.

SMEs: Small and medium-sized enterprises

Strong interoperability: The ability of a system or product to fully interact with other systems or products of a cross-sectoral and widespread nature, including, for example, the transfer of data from the private to the public sector and vice versa

Virtual data repository: A form of cooperation within the organization of associated entities; a pattern of behavior involving the sharing of data within a business federation in a trusted digital environment (data spaces) based on reciprocity, within the adopted standard logic of data availability.

1. Introduction

With the emergence of new technological solutions and increasing access to the Internet, not only the global economy but, above all, digital social and economic awareness was gradually transformed. The first Internet users believed that access to free and open content – while maintaining anonymity – could facilitate the process of democratization of knowledge (Mayer-Schönberger, Ramge, 2022). Despite the illusory free-of-charge nature of Internet services, in reality, it was people using them who contributed to the expansion of information resources on the Internet, leaving behind themselves and their activities an invaluable, digital footprint.

In 2020, both GAFA and other technological giants (Microsoft, Tencent and Alibaba) owned more than 300 subsidiaries located in tax havens, used to obtain more favorable taxation (Wawrzyniak B., Iwanowski D., 2021). Moreover, these companies claimed for many years that they were not affected by any national regulations because it was impossible to clearly determine where their capital was actually located and where the value from the processed data was generated. Consequently, 80% of corporate wealth was in the hands of just 10% of the world's enterprises (Foroohar, 2019).

At the same time, the purposes for which these companies used their data collections went far beyond the usual basis for processing information related to, among other things, improving suggestions for users or search results. Numerous infringements of competition protection laws are also known, including algorithmic discrimination against price comparison websites other than Google Shopping in the search engine, or influencing freedom of expression on the Internet by biased blocking of selected political content. There were also situations in which data was used to profile users in a way reinforcing harmful stereotypes and inequalities or radicalizing certain social groups, which led to the destabilization of public order (Zygmuntowski, 2020a).

The result of this state of affairs was a paradigm shift in thinking towards increasing the protection of privacy on the Internet. Although the protection of personal data is an attribute of personal rights, due to the digital format of the information contained therein and the possibility of monetization, data began to be treated as protected by property rights, which is not based on the established law. The public unveiling of the Cambridge Analytica scandal coincided with the entry into force of the General Data Protection Regulation (GDPR) in 2016, which to this day is an important starting point for the discussion on the possibility of "free flow of personal data". The main purpose of the GDPR was to regain the individual's control over the data that concerns them by obliging platforms to disclose information about how the data is processed and to ensure that this information is made accessible in a transparent and understandable way. Requiring platform users to obtain consent for the processing of their personal data has contributed to an improvement in their awareness of their digital rights and to deepening their resentment towards the largest technology companies. As the report of the Polish Economic Institute shows, average users expect monetary compensation in exchange for extensive access to their data and for watching personalized advertisements on digital platforms (Polish Economic Institute, 2020).

A breakthrough step towards changing the way we think about data, as having only personal properties and subject only to the disposal of an individual, was the submission of a complaint by Maximilian Schrems to the Irish Data Protection Authority regarding the rules allowing the transfer of personal data from the EU to the United States. As a result, the CJEU invalidated the Privacy Shield, while ruling that further data transfers under this decision are prohibited. Despite acting on his own behalf, the problem highlighted by Max Schrems was in fact about European data understood as a common good deserving at least as sublime protection as the GDPR provides.

Attributing only economic characteristics to data seems to inherently inhibit the socially beneficial use of the data. It is this approach that makes companies overlook the many opportunities in which the data they collect and store could also create valuable public assets (Swant, 2019). The emergence and dynamic development of technological tools based on artificial intelligence systems has particularly highlighted the need to move away from a commodity approach to data in favor of its reuse in order to ensure overall social growth and the creation of a shared and public value.

Thus, although some refer to data as "new oil" (e.g. The Economist), in fact, data has fundamentally different characteristics. Raw data, for example, can be compared to "air" as a resource existing in the natural environment (e.g. in the field of data generated by smart cities, smart devices (IoT) or autonomous transport). Complete freedom of access to data should apply to public data or data from the human environment. Data held by companies, especially small and medium-sized enterprises, or data on the public health, due to its specificity, should be usable based on approaches developed for specific sectors (Borowik M., Maśniak L., Kroplewski R., Romaniec H., 2018). That is the purpose of this report.

Digital data is not competitive in consumption because its resources cannot be exhausted through multiple exploitation (Zygmuntowski, 2020a). It is a result of activity of humans (personal data) or human-operated devices (non-personal data). Responsible data management must therefore take into account people's rights, giving them the right to protect privacy or grant access to, and share, data but also take into account the possibility of multiple uses of information for various purposes. There is data which, for various reasons, is confidential, constituting a trade secret or data concerning the health of patients, which require special protection. A trade secret may consist of information of economic value which, as a whole or in a particular context, is not generally known to persons normally engaged in this type of information or is not easily accessible to such persons, provided that the rightholder has taken appropriate steps to keep it confidential. This data may have personal nature (e.g. customer lists) and non-personal one (e.g. ways to improve a product). In the case of the former, the principles of the GDPR will apply – the level of protection of individual personal data will depend on whether the processed data constitute "sensitive data" or not. In the case of the latter category, i.e. non-personal data, the provisions of the GDPR do not apply; however this data may be covered by regulations on unfair competition.

The potential of data should therefore be used not only by technology giants but also (or above all) by administrations, public service providers, research centers, authors in the fields of culture, art and innovation, NGOs and non-technological SMEs (i.e. using technologies only subsidiarily, in connection with the conduct of other types of commercial activities). Ignoring the power of digital information, these entities would deprive themselves of access to new intelligent solutions and, thus, risk falling into technological debt and the resulting losses.

The most forward-looking strategy for modern economies is to manage data as a common resource considered as an infrastructure ("a means for many actors for many purposes"), rather than a mere commodity. Maximizing the potential of data is offers Poland an opportunity to develop its own high-tech solutions, modernize public services and improve quality of decision-making in all sectors. To achieve this, we need to unleashing the potential of data by effectively sharing it within trusted spaces, institutions and technologies set up for this purpose. Although data-sharing institutions are a completely new concept, their establishment may be compared to the once innovative idea of setting up banks or cooperatives without which the functioning of the modern economy is difficult to imagine.

2. The state of the data-based economy in Poland

The value of data in Poland is estimated at EUR 6.2 billion. In the next 3 years this value may double (Bożykowski et al., 2019). This data turns out to be significant for economic growth: the overall data-related productivity in Poland is about 92%, well above the European average. This means that the economic benefit of increasing the use of data in Poland will be particularly measurable. At the same time, the intensity of the use of data, including non-personal data, in the Polish economy has turned out to be significant because 46% of the Polish GDP depends on the cross-border flow of non-personal data (Koloch G., Grobelna K., Zakrzewska-Szlichtyng K., Kamiński B., Kaszyński D., 2017).

Noticing the potential of data for the development of the Polish economy gave the foundations for the adoption of various policies, recommendations and regulatory changes in the field of data reuse for individual technologies. Already in 2013 it became clear that there was a need for opening access to public information for businesses and citizens by making data and documents available for re-use (Ministry of Economy, 2013).

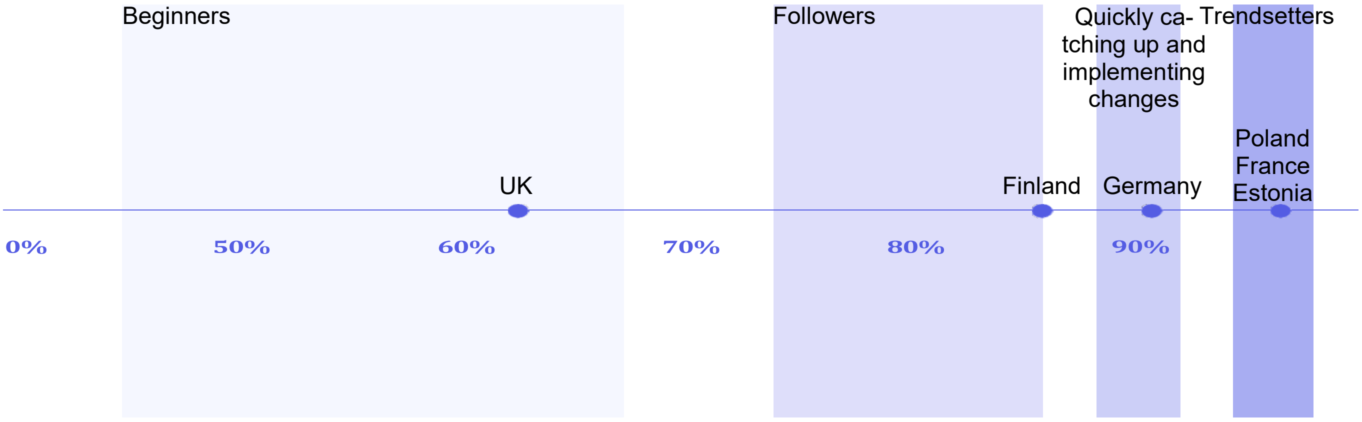
In the following years, Polish strategic documents increasingly pointed to the need to direct regulatory activities and implement innovative changes in the field of digitization and access and effective use of collected data. Starting from the program of opening public data, an expression of the right to information in line with technological progress (Ministry of Digitization, 2016), through the Strategy for Responsible Development, the primary goal of which was to ensure sustainable economic growth based on knowledge, data and organizational excellence (Strategy for Responsible Development, 2017), Polish recommendations in an increasingly comprehensive way sought to shape a uniform approach to the information technology.

When it comes to non-personal data, the heading for the Polish strategy in this area was given by the report "Industry Plus – data-based economy" from 2018. It defines 5 key pillars for economic development based on the use of data: access to data, advanced processing skills, digitization of industry, communication and trust of participants and processes within it (Borowik M., Maśniak L., Kroplewski R., Romaniec H., 2018). The document indicates that the key task in the near future is the digitization of industry, involving both the digitization of objects used in industry, as well as the data based on business processes or the reuse and connection of data streams within and between sectors.

After numerous discussions at the EU level, the "Policy for the development of artificial intelligence in Poland" (Council of Ministers, 2020) became another important step emphasizing the potential of data for the development of various sectors of the state. The document shows how important the stage of acquisition, collection, analysis and conscious use of data is for the development of artificial intelligence in areas such as society, innovative companies, science, education, international cooperation and the public sector.

Despite the successes of the digitization of public services in recent years (e.g. mObywatel, OInline Patient's Account), Poland is at the end of the Digital Economy and Society Index ranking prepared by the European Commission in 2021 (European Commission, 2021). This is due to many different factors that influenced the final results of the ranking. Despite the increased rate of online access to public services for citizens, the level of basic digital skills of Poles is below the European average (44% vs. 56%). Despite this, according to the Open Data Maturity report from 2021, measuring the degree of progress in the field of opening public data in individual European countries, Poland ranks 4th, reaching a data maturity level estimated at 95% (Van Hesteren et al., 2021), just behind France, Ireland and Spain.

Fig. 1: Maturity of open data in the EU



However, the success of Polish projects for the digitization of public administration and health care should not obscure their imperfections and barriers to be faced. The Online Patient's Account and the associated Electronic health records continue to provide only poor interoperability. The obligation to apply uniform standards was imposed only on projects related to the development of the Electronic Platform for Collecting, Analyzing and Sharing Digital Resources about Medical Events and the platform for making medical records available to private health care institutions (Ministry of Health, 2018). These projects are not about the exchange of patient data between the public and private sectors. Compared to other EU countries, Poland and Romania shared the last place in terms of interoperability of medical data contained in the patients' electronic health records (Empirica, 2022).

The success of the Open Data program gives hope for the success of the data sharing program – both personal and non-personal, coming from the public sector, businesses and ordinary citizens. Poland still has a chance to be among the countries promoting the egalitarian, socialized nature of databases and ensuring their availability to various social and economic actors. At the same time, the low level of trust cannot be underestimated, as clearly demonstrated by the public response to the implementation of the Integrated Analytical Platform and the Electronic health records. To modernize the existing digital information management model, it is necessary to identify existing barriers and prioritize and develop guidelines and standards for data sharing in trusted spaces.

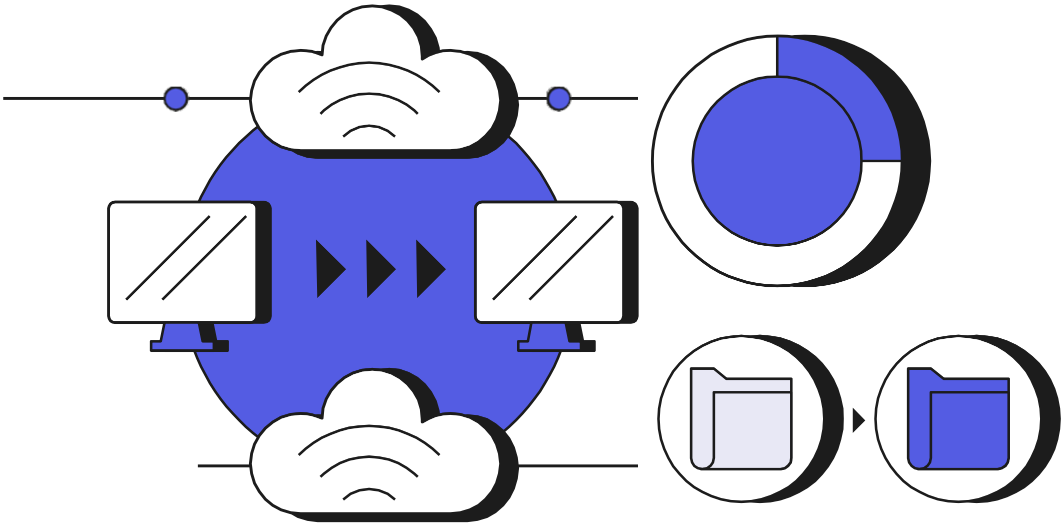
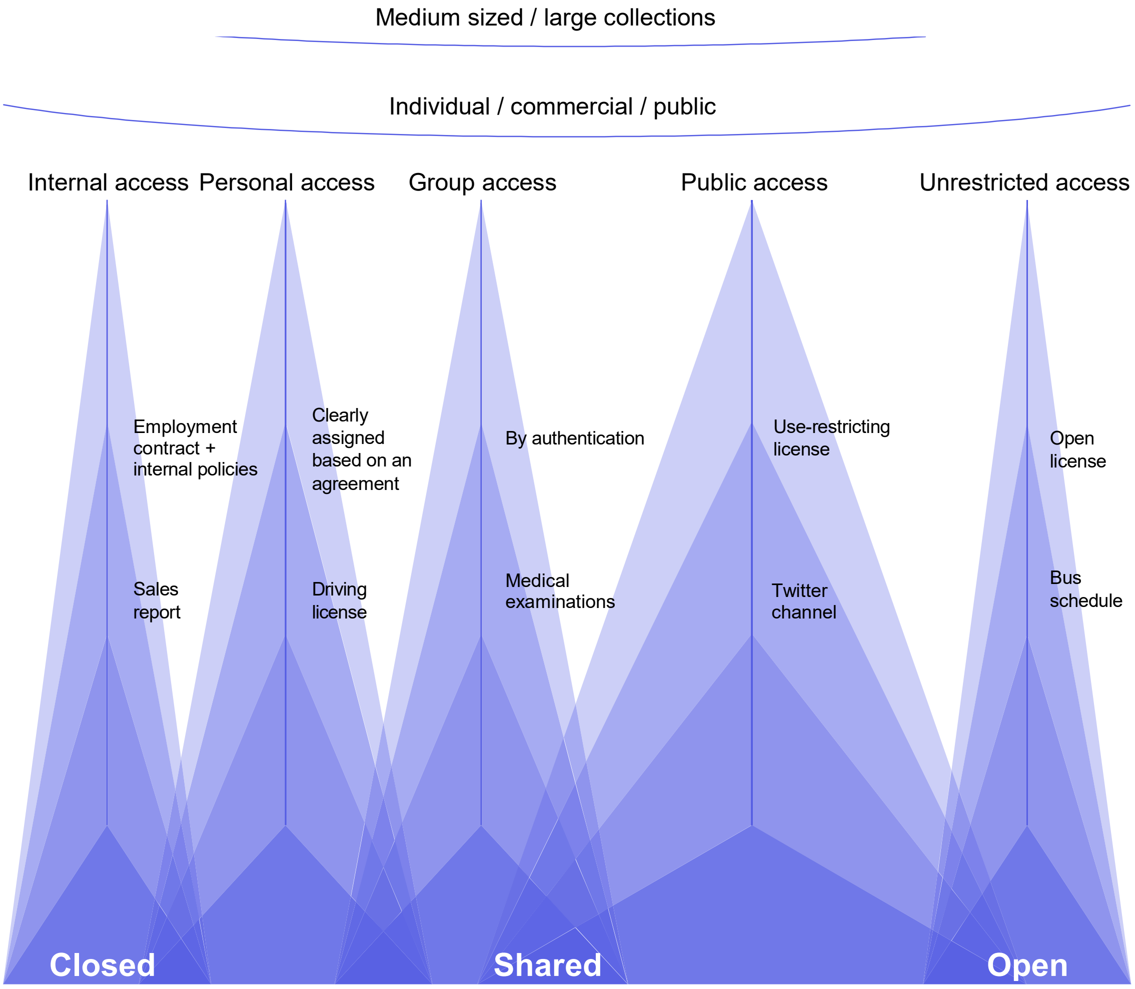


Fig. 2: Cross-section of data openness



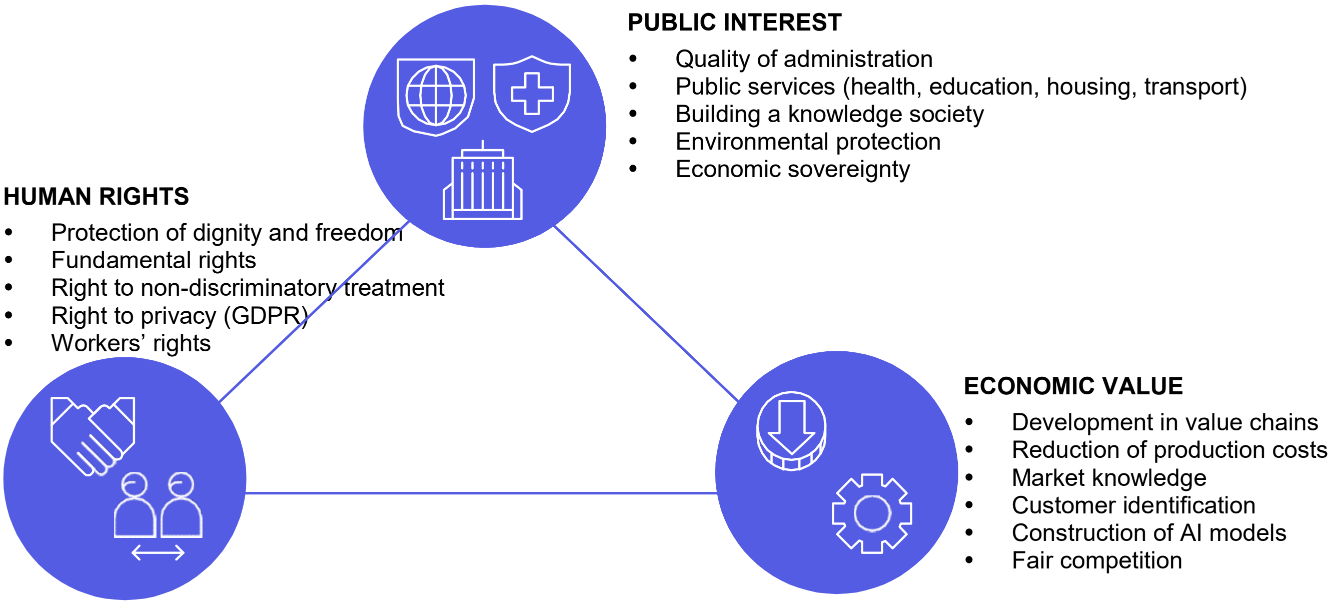
Źródło: The Open Data Institute’s (ODI) Data Spectrum

3. Benefits of data sharing

Due to the importance of privacy protection in the discussion about data, so far it has been assigned value on an individual basis, through the prism of proper enforcement of the rights of data subjects. Therefore, according to some, it is possible to talk about “ownership of personal data”, allowing the formation of a market for its commercial exchange. However, this approach leads to data being locked in private silos and the potential of aggregation and further (re)use is lost. To maximize the benefits, data sharing must not only avoid the trap of commodification of data, but also the trap of individual ownership.

Rather, data management presents itself as a trilemma that must be resolved in a trusted secure space of cooperation – both in the context of infrastructure and procedures of institutions and legal solutions.

Fig. 3: The trilemma of data management



PUBLIC INTEREST

- Quality of administration
- Public services (health, education, housing, transport)
- Building a knowledge society
- Environmental protection
- Economic sovereignty

HUMAN RIGHTS

- Protection of dignity and freedom
- Fundamental rights
- Right to non-discriminatory treatment
- Right to privacy (GDPR)
- Workers' rights

ECONOMIC VALUE

- Development in value chains
- Reduction of production costs
- Market knowledge
- Customer identification
- Construction of AI models
- Fair competition

Mutual trust of participants in the data sharing process plays a key role. Large-scale cooperation, both cross-sectoral and within a given industry, allows to shape a conscious decision-making process in the area of further more sustainable development. Only honest and transparent stakeholder cooperation is able to eliminate excessive digital dominance based on limited access to information. The concentration of information power is beneficial for the few, because it inhibits innovation and hinders access to benefits for society and, consequently, for each of us (Mayer-Schönberger, Ramge, 2022).

Below are examples of the benefits of data reuse by each sector:

State administration and local governments:

- Reducing the cost of public services
- Implementation of smart city and communities solutions (e.g. optimizing energy consumption through smart urban tariffs efficient management of urban transport, intelligent combination of urban and rural services)
- Integrated knowledge of real estate market and housing needs
- Forecasting infrastructure consumption and necessary investments
- Local development and the creation of new businesses, products and services based on the known habits of residents
- Modernizing public services with data-driven technologies
- Improving the quality of healthcare, medical devices and better detection of diseases (e.g. development of AI in medicine)
- Personalized education

Entrepreneurs and business:

- Increasing the availability of trade data useful for forecasting trends
- Identification of competitive advantages
- Development of new products and services, including high-tech ones
- Building pricing strategies and market analysis
- Optimization of the customer service process
- Reducing the costs of running a business by optimizing logistics
- Generation of system products in complex value chains

Industry and agriculture:

- Implementation of innovations in industry (e.g. Inventing new ways of producing goods; Improvement of machine mechanisms)
- Reduction of industrial production costs due to optimization
- Improving energy efficiency (anticipating electricity demand and the formation of balancing sources, forecasting energy losses in networks, optimized planning of investments in the energy sector)
- Raising the technological potential of food crops in Poland
- Improving resource efficiency, efficiency and environmental sustainability
- Providing rural communities with better living conditions
- Improving the relationship between the consumer and the various actors in the value chain

Science, NGOs, civil society:

- Improving quality of public debate and decision-making processes for the common good
- Increasing social participation (e.g. creating IT tools to involve the public in processes taking place at the local level)
- Increasing social control over data
- Supervising quality and reliability of public administration, business and other data sharing
- Medical use of “wearables” (e.g. watches warning of an epileptic seizure)
- Conducting scientific research, both by scientists and citizen science

3. Panorama of international practices

4.1. Activities at the EU level

In the context of the importance of data, until recently, EU regulations concerned only the individual protection of the rights of the entity to which the digital information relates. But although one of the guiding principles of the GDPR remains the principle of limiting the use of data to one specific purpose, the regulation offers derogations from it in certain cases. It expressly allows for the “further processing for archiving purposes in the public interest, for scientific or historical research purposes or for statistical purposes” (Alemanno, 2018). Both this principle and the previously insufficiently used right to data portability (data portability, art. 20 of the GDPR), are the foundations of the data sharing ecosystem.

This paradigm shift stems from the regulator’s recognition of the EU’s potential for data thanks to the efforts of the D9+ member states including Poland. In the strategy of Feb. 19, 2020, the European Commission declared that its aim was to create a common European data space within the single market where it can be used – in accordance with the applicable rules and regardless of the physical place of storage in the EU (the European data strategy). As a result, the EU seeks to open up public data, encourage private actors to share data, and increase availability of proven cloud services. Both soft actions (raising awareness or offering incentives) and creating an appropriate regulatory framework (Data Governance Act, Data Act) are intended to support the achievement of the fundamental, perhaps the most ambitious, goal of the EU’s digitization policy (European Commission, 2020a).

The first act of the strategy is the Data Governance Act which complements the Directive on open data and on the re-use of public sector’s information. The aim of the Act is to broaden the scale of this phenomenon and improve the conditions for data sharing. According to the Commission’s proposal, the main provisions of the regulation cover four areas:

- Making public sector’s data available for re-use in situations where the data is covered by the rights of others;
- Sharing data between companies in exchange for remuneration in any form;
- Enabling the use of personal data with the help of a “personal data sharing intermediary” to assist individuals in exercising their rights under the GDPR;
- Enabling the use of data for altruistic reasons.

The proposed regulation also places particular emphasis on increasing trust in data intermediaries. These entities, referred to in the regulation as data sharing service providers, are tasked with ensuring a secure and trustworthy environment for data sharing by data holders. This possibility applies to both legal and natural persons. Services intended for legal entities will consist in creating dedicated spaces in which entities will be able not only to place their data but also to use any data contained there. For natural persons, the EU regulator has provided for facilitations in the provision of data to their potential future users. For this purpose, appropriate applications will be developed to allow access to data spaces / portfolios through which access to data is granted to entities wishing to use it. These measures are intended to make it easier for individuals to exercise their rights under the GDPR by consciously deciding who to share their data with (Małobęcka-Szwast, 2021). More about the institution of data intermediaries as one of the methods of sharing, the benefits of this model and the identified barriers, can be read in the next part of the report.

Another proposed act regulating the rules for sharing data is the Data Act. It concerns, among other things, data generated by users using IoT (Internet of Things) devices and the possibility of extracting the data for further use or transfer to other entities. Data processing service providers and operators of the spaces in which data will be located will also be obliged to comply with the rules on the strong interoperability of data recording formats (Małobęcka-Szwast, 2021). At the same time, the project unfavorably restricts the access of the public sector to data of private entities only to cases requiring extraordinary intervention (e.g. in situations of natural disaster) and narrows the scope of use of the data by a third party to whom the user may transfer it.

The first sectoral project out of the planned nearly 10 is the project to create a common European health data space (European Commission, 2022). Its main objective is to encourage the use of health data for research, policy-making and law-making purposes, based on a trusted governance framework and data protection principles. The draft provides for the establishment of a “data access body” at the national level, whose task would be to grant permissions for access to data. The project also draws attention to safety and accountability issues in the context of the use of AI in the field of health, while promoting a proactive attitude of citizens in controlling their health data.

One of the main objectives of the project is to introduce requirements for infrastructure and for strong interoperability covering the entire area of the EU (Bertuzzi & Fortuna, 2022). The design is consistent with the aforementioned Data Governance Act and the Data Act, which are also at the design stage. According to the project, patients are to be the main entities deciding on limiting access to their data or making it available free of charge. Unlike before, the data collected will come from various sources: health records, public registers, social, administrative, genetic and genomic data, clinical trials, research questionnaires and biomedical data (e.g. biobanks). This data may be used, inter alia, in the activities of public authorities in the performance of their tasks, for research and development and scientific purposes, or for the creation of new solutions in the public interest. In addition, according to the draft, individuals will have free access to a minimum set of “basic” health data, including vaccinations, electronic prescriptions, photos, laboratory test results, discharge reports and more. Ensuring adequate security of electronic health records is also extremely important: they will have to meet strictly defined technical requirements, including those related to strong interoperability.

4.2. Activities at the national level

Due to the complicated process of implementing EU policies, some European countries have already started to take steps to share data and build dedicated open spaces. Data sharing ecosystems already exist or are being implemented in countries such as Estonia, UK, Germany, Finland, France, and Japan.

Here are some of the solutions:

Estonia – *XRoad*

- Access to data is possible only by means of identification cards for authentication and digital signatures.
- Solution for the public sector – entities outside XRoad do not have access to the data collected there.
- Obligation of healthcare professionals to transmit data to the health information system accessible only to licensed workers.
- The ability to use most of the pan-European data held in the main databases in Estonia.

- High security features, such as overlays on individual patient data, which reveal only necessary information.

UK – *Open Data Institute*

- A non-governmental research institute working with actors from various sectors to create pilots for secure and ethical data ecosystems.
- Cooperation with both private and public entities.
- Growing interest in exchanges between enterprises and government bodies (B2G).
- Purpose: supporting actors from different sectors in building open trustworthy data ecosystems in their organizations.
- Conducting studies together with the UK Office for Artificial Intelligence and Innovation to assess the potential of using one of the data trust models based on 3 pilot programs on urban data, food data and international wildlife trafficking.
- Research on different national legislative regimes and approaches to data sharing models – concludes that debates arising in all jurisdictions on the development of data trusts indicate the importance of relying on local circumstances to meet new governance challenges, adapt structures to the relevant context and purpose, and control tensions between individual and collective interests.

Germany – *Daten-Treuhänder*

- Planned intermediary for automotive industry and insurers.
- Data made available for projects on improving road traffic safety and construction and repair of machinery.
- Access for the public sector, insurers, technical inspection unions and service centers.

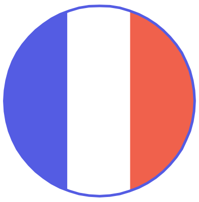
Germany – *Bundesdruckerei*

- A state-owned enterprise producing documents and devices for their verification, also offering intermediary services in the transfer of data.
- Authorizing access to data, combining data into larger sets, monitoring data sets and their quality, data analysis.
- Securing data by pseudonymizing or anonymizing it.
- CenTrust users: Robert Koch Institute (e.g. COVID-19 vaccination database), health sector agencies.

Finland – *Finlandia*

- Finnish Authority for Permissions to Share Social and Health Data.
- Public authority managing access to various datasets on health and social affairs.
- Storing data from both private and public healthcare providers.
- Granting access to data.
- Paid data sharing system.
- Tariffs depending on the amount of data and the purpose of its use.
- Opt-out system for citizens (only 200 people used this option since the start of Findat's operation).
- Access granted to different actors to conduct research.
- The result of the research for which Findat's data was needed should be made publicly available.
- Introduction of the function of data controllers (people who watch over and ensure completeness and proper implementation).

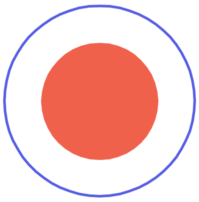
France – *Health Data Hub*



- Central data access point.
- Collection of various types of health data, including those related to the reimbursement of health insurance, regardless of the subject and type of medical service.
- Securing data through pseudonymization.
- Use of data only for public interest purposes.

- Prior consent of the National Commission for Data Protection and Freedoms (CNIL) is required.

Japan – *Smart Data Platform with Trust*



- Improving data exchange between Asian companies.
- Project supported by the Ministry of Economy, Trade and Industry (METI).
- Information for companies on stocks of products and parts, potential supply disruptions, etc.
- Companies may decide what data they intend to share.
- Depending on the nature of the specific data (e.g. constituting a trade secret or copyrighted), the possibility of sharing data under different conditions.
- Development of data exchange infrastructure at regional level.

5. Data sharing models – research workshops

This publication presents the basic data management models identified during the workshops on the "Unleashing of the potential of data". The whole cycle of work consisted of four workshop blocks lasting 40 hours in total. Each of the modules comprised a theoretical and a research parts (discussion; problem-solving). The first three meetings concerned the following, in this order: personal data intermediaries; virtual shared repositories of non-personal business data; public shared repositories of personal data. The fourth block was devoted to the analysis of ways of managing data depending on the degree of sensitivity.

Fig. 4: Management obajectives and methods

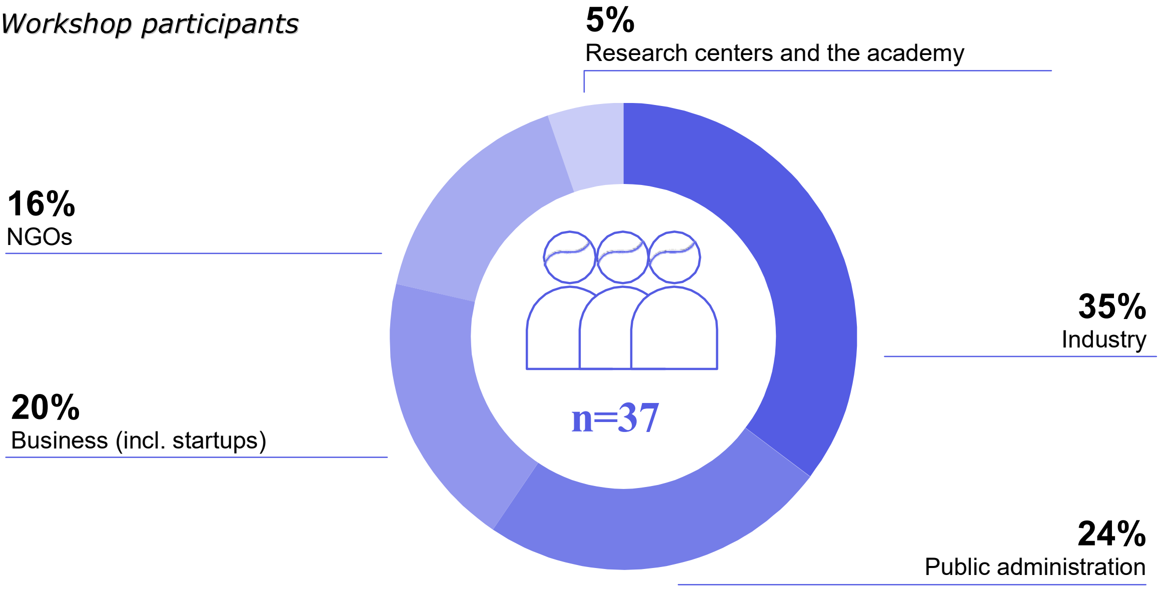
Value allocation / management purposes

Stakeholder control / management methods		Individual / rights-based	Institutional / trust-based
	Private / profit	Personal data capsules	Virtual shared data repositories
	Public / common good	Personal data intermediaries	Public shared data repositories

Division developed during the workshops on the „Unleashing of the potential of data"; inspired by Zygmuntowski J.J., Zoboli L., Nemitz P.F. (2021). Embedding European values in data governance: a case for public data commons. Internet Policy Review, 10(3). https://doi.org/10.14763/2021.3.1572

The project was attended by representatives of state administration, NGOs, industry, business, as well as representatives of scientific and research centers and academic communities, including those working in leading institutions involved in the sharing of data from abroad.

Fig. 5: Workshop participants



We used the following research methods during the workshops and subsequent analysis:

- **SLEPT analysis** (Socio-cultural, Legal, Economic, Political, Technological): A method of general segmentation of the macro-environment and identification of trends, barriers and other variables aimed at analyzing the market in many contexts (not only the competitive position). Used as an introduction to feasibility studies. Most often used in the PEST variant, here taking into account legal and political factors separately (at the level of government strategies in Poland and in the EU).
- **Critical Success Factors**: Elements of a company's / organization's operation necessary for its success, based on which the level of success and goals are often set, for example KPIs.
- **Delphi method**: A heuristic method (developing creative thinking) consisting in using the knowledge, experience and opinions of experts in a series of research questions, where in subsequent rounds the results of the previous one are treated as input data. Thanks to this, there is a feedback loop correcting deviations from the main forecast.

In this report we highlight the benefits of data sharing, as well as the barriers resulting from the SLEPT analysis, which can potentially inhibit data sharing in a given model. Specific solutions and factors were also proposed, which, according to the respondents, are crucial for the creation and proper functioning of secure data exchange spaces.

The publication was written based on the material collected during the workshops, which is why we will find in it parts devoted to each of the data spaces discussed during the cycle. While maintaining a comprehensive approach to the issue of data sharing, the authors' attention focused primarily on the shared repository of industrial and agricultural data and on the virtual shared repository of health data. The workshop has shown that it is these spaces that have the best odds for coming to existence in the Polish public space, contributing to the improvement of quality of life of all citizens and reducing the innovation debt.

5.1 Personal data intermediaries

Personal data intermediaries represent a promising concept for the use of data while respecting privacy principles. These institutions can be created for purposes such as more effective implementation of personal data protection or more effective encouragement of data sharing throughout the value chain. By taking action in accordance with the interests of Internet users and putting their needs first, intermediaries have a chance to become an alternative model to the one imposed by the largest digital platforms. While digital giants are accused of collecting a huge amount of data used primarily for commercial purposes, data intermediaries would act primarily for the benefit of their users.

The main assumption of this model is to provide a secure and trustworthy environment in which legal entities or individuals (data holders) can share their data. Personal data intermediaries are also responsible for assisting data subjects in making informed choices about their consent to the processing of their data, enabling individuals to collect data voluntarily for mutual benefit. The introduction of an independent intermediary between data subjects and data collectors also allows for more effective negotiation of the terms of use of data in accordance with the requirements of a secure environment for their exchange. This is due to the greater bargaining power resulting from the aggregation of data in the hands of one entity (Data Trust Initiative, 2021). As it is known, data gains value only in mass. While a single withdrawal of consent to their processing will not adversely affect the platform's business model, the loss of more records would be a real threat to the service (Paszcza, 2022).

The need to give users more control over their data resources has been recognized by the EU regulator. However, the rules provided for in the EU legal acts (the Data Governance Act and the Data Act) take into account the low flexibility in terms of the freedom of activity of intermediaries. According to the draft regulation on European data governance, data sharing service providers are only tasked with acting as intermediaries in transactions and must not use the data provided for any other purpose. Thus, the inability to profit from data management calls into question the existence of intermediaries in the real world. An attractive business model is crucial for the emergence of new entities on the market and for the development of competitive high-quality services.

FOR WHAT DATA TYPES?

- Commercial data
- Payment data
- Data from smart devices
- Location data
- Social data
- IP addresses

WHAT FORMS?

Data trusts

Institutions that manage a person's data on their behalf in the manner of the "trust fund" present in Anglo-Saxon law. Relationships are established based on repetitive frameworks occurring under contract law (Mehta, Dawande and Mu, 2022). The entrusting entity transfers data to the trustee and the data can then be used by a third party, i.e. the beneficiary. Data trusts are divided into those that store data and those that manage individual and collective data access rights. This model can be compared to libraries or collections that allow digital access to content intended to serve a specific community and protect resources from unauthorized access (Artyushina, 2021). In terms of responsibility for data, this concept is also compared to real-world professions burdened with professional secrecy (e.g. legal, medical). Data managers gain access to personal, potentially sensitive, information but at the same time are legally obliged to act in the best interests of the beneficiaries of their services (Artyushina, 2021). Thus, it is evident that, in the context of trust, the fiduciary duty is associated with a high degree of impartiality, prudence, transparency and loyalty (Delacroix and Lawrence, 2019). If the trust law is applied to the data trust, the trustee is legally obliged to be loyal and diligent to the beneficiary. On the other hand, compliance with these commitments requires the data trust to be independent, which may prevent the institution from becoming a for-profit firm (Mehta, Dawande and Mu, 2022).

An example of this form of data sharing is the non-profit PlaceFund, which acts as a geospatial data trust, promoting the use of data to correct issues related to land ownership, unsustainable land use, and climate change. The PlaceFund's ambition is to create a trusted space for geospatial data that will be processed sustainably and then shared with local communities.

Data cooperatives

This proposed institution is described specifically in the Data Governance Act as one of the data sharing services. By ensuring supervision and transparency, data cooperatives are to enable individuals not only to exercise data rights but also to participate in the management of the cooperative, in accordance with the principle of 1 person – 1 vote. Their services would make it possible to empower individuals in their relationships with platforms and to support users in making informed choices about consent to the use of their data (European Commission, 2021). These institutions would also have the task of improving the conditions offered to data subjects and resolving disputes concerning several data subjects within the group (Bayamlioglu, 2021).

Examples of existing cooperatives include the Driver's Seat, a cooperative that aggregates data related to the use of smartphones by drivers active in the gig-economy area. In turn, the Swiss Midata or the Spanish Salus.coop collect health data and allow cooperatives to decide for which tests they want to transfer them and on what conditions.

The use of data intermediation service providers has a potential to address various socio-economic issues and, in addition, to improve the position of the individual in relation to digital entities. First of all, by aggregating individual data, an institution such as a data cooperative strengthens its bargaining power and thus can obtain more favorable conditions for data collection (Mehta, Dawande and Mu, 2022). However, in addition to the benefits of entrusting digital resources to cooperatives, or more broadly, personal data intermediaries, barriers and doubts related to this model can also be observed.

BENEFITS

Exercise of control over data by the user

The assumption of using the services of personal data intermediaries is to increase the agency and decision-making power of individuals in terms of what happens to their “digital footprints”. Data subjects may also have control over quality and quantity of data they share (Mehta, Dawande and Mookerjee 2021).

High data security rating

The overarching purpose of the intermediation provided by the cooperative is to provide a safe and trustworthy environment in which legal entities or natural persons (data holders) will be able to share their data. Data trusts, on the other hand, are characterized by a high security rating due to the obligations resulting from their “fiduciary” nature.

Greater bargaining power

An individual’s personal data is not of great value in itself (Pentland and Hardjono, 2020). For this reason, the negotiating power of a single user is small, which is sometimes used by platforms using the “take it or leave it” policy. Internet users, not having the opportunity to interfere with the platform’s regulations, are often forced to accept the conditions set by the website, although they are not always favorable. The introduction of the institution of a personal data intermediary is an opportunity to change this paradigm. Assuming that data is gaining in value in mass, it can be expected that trusts or cooperatives with larger data resources will be in a better position to dictate terms to platforms and demand more sustainable data processing.

Relieving data subjects

Instead of engaging in relationships with individual entities, data users can enter into a single agreement with a cooperative that governs access and terms of use of the data (Mehta, Dawande, and Mookerjee, 2021). The data trust, on the other hand, relieves the entity of the need to make the most important decisions regarding its personal data, while ensuring that all operations are in accordance with the requirements of privacy and security.

BARRIERS

Uncertainty of the business model (economic barrier)

As already mentioned above, in accordance with the assumptions of the EU’s draft Data Governance Act, service providers should not use shared data for purposes other than the intermediation itself. Therefore, they cannot profit from the data, for example by selling it to other entities (Council of the EU and the European Council, 2021). Moreover, the business model of intermediaries is to guarantee absence of inappropriate incentives for individuals to share more data for processing than is in their own interest (European Commission, 2020b). Thus, personal data intermediaries have limited opportunities to monetize the digital information provided to them. This was pointed out by the participants of the workshops. The concept of how data intermediation institutions could earn money calls into question the effectiveness of this data sharing model.

Lack of awareness of the benefits of data sharing (socio-cultural barrier)

While Poland ranked fourth among the 27 EU Member States in terms of data openness (Van Hesteren et al., 2021), it ranked 24th in the Digital Economy and Society Index (DESI) in 2021 (European Commission, 2021). This poor result is due, among other things, to the insufficient level of digitization of the nation. On the other hand, the lack of adequate knowledge in the field of technology is manifested by public distrust and unawareness in the field of innovative concepts and solutions, such as data sharing.

Non-uniform standards (technological barrier)

Both private entities and public administration institutions in Poland use different formats and various standards of data exchange, which hinders effective sharing.

Lack of a proper government strategy for data sharing (political / strategic barrier)

Despite the high level of openness of public data, the data sharing between entities operating in both public and private spaces is limited. At the same time, the government has no relevant strategy, such as the one for Open Data.

PROPOSED SOLUTIONS

1. Incubation of an attractive business model

In practice, data intermediaries can operate in various forms: from private entities to non-profit organizations and public institutions. Their structure, motivation and management depend on different conditions and goals written in the organization’s statute or company strategy. Therefore, the business model of the service provider should be tailored to the specific case. In the case of entities operating in a non-profit form, it would be a good idea to launch a program of small grants for data intermediaries (public benefit organizations and cooperatives) to build a basic data sharing infrastructure, maintain its proper operation and formulate rules for using the service. Private entities, on the other hand, should be able to obtain financial benefits for their intermediation services, for example by charging fees for joining the data ecosystem they have built (Janssen and Singh, 2022).

2. Raising awareness of data sharing

Workshop participants pointed to the insufficient knowledge of the public about the benefits of data sharing. Awareness of the advantages of transferring data to independent proven intermediaries is crucial to overcome the concern of users and their reluctance to new institutions. Therefore, it is necessary to place greater emphasis on high-quality education in the area of digitization (primary schools, general secondary schools), as well as activities promoting the shared nature of data (e.g. campaigns promoting subsidies for employee training in companies).

3. Development of a uniform standard for data exchange

A proposal to separate from the government administration an entity that could deal with setting standards and requirements and certifying entities intending to provide intermediary services in terms of compliance of their operation with top-down standards.

4. Planning an effective strategy

The low rate of “digitization” of our country translates into difficulties in implementing concepts such as data sharing. It would therefore be advisable to look at strategies for opening up public data in order to draw inspiration and identify good practices that could also apply to the sharing of personal data. We are talking about paying attention to the means (capital, human, organizational, information) that were used to open public data, and then using this knowledge to outline the directions of activities in the field of building data sharing institutions.

5.2. Virtual shared data repositories

The use of business intelligence by entrepreneurs is becoming more and more common due to the possibilities offered by AI tools in terms of improving business performance, reducing costs and improving quality of products. However, building accurate predictive models and creating artificial intelligence using machine learning requires access to huge data sets. While in the case of large corporations this is usually not a problem, SMEs, start-ups, local producers, farmers continue to lag behind, thus falling into innovation debt. This is why it is so important to create conditions for data sharing between the above-mentioned entities. One of the solutions may be to create a space for business data discussed by us during research workshops.

The virtual shared data repository provides a form of cooperation in which economic actors share access to data (via API, blockchain technology) to create collective value (Data Collaboratives, 2021). The new name adopted in the publication for the former virtual data repository results from the fact that the workshop hosts noticed the loss of relevance of the division into personal and non-personal data, and the need to focus on their functions and purposes of use. The term “shared data repository” is intended to emphasize the common nature of access to data, while the “virtuality” of the shared data repository is to emphasize its digital form and the federated model of data exchange using distributed repositories (as opposed to a central repository where common data is stored).

The main assumption of this concept is to encourage enterprises (large companies, SMEs, start-ups) to establish partnerships in order to create a certain pattern of data exchange, based on the free flow of digital information within the federation of stakeholders. Communication and collaboration could take place through interconnected platforms with a uniform standard for shared data. Then, by exploiting the positive network effects, it would be possible to create more complex products and links together.

Such systems can be found, for example, in the case of the already existing form of cooperation between entrepreneurs – the International Data Spaces (formerly: Industrial Data Space). Another interesting example of a form of sharing agricultural data is the Ethiopian Commodity Exchange: an exchange that brings together farmers and provides them with data they need to improve quality of agricultural products and optimize crops (Verhulst, Young and Srinivasan 2022).

FOR WHAT DATA TYPES?

- Industrial data (internal production or service lines of enterprises, data produced by machines, data related to the maintenance of machinery, supply chain data)
- Business data (including the number of visits and the time spent on the website, big data analysis, logistics data)
- Agricultural data (sensor data, data from tractors, meteorological data)
- Logistics and transport data
- Financial data (B2B / C2B transactions)

WHAT FORMS?

Trusted entity

It assumes the selection or creation of an intermediary entity (potentially controlled by the sharing parties) that takes care of technical standards and terms of cooperation, but does not store data. The creation and promotion of trusted entities could also take place through enhanced cooperation within the EU; NATO; the countries of the Three Seas Initiative; The Visegrad Group or the Weimar Triangle. Then, it would be possible to establish trusted entities not only between Polish entrepreneurs but also between foreign entities (Council of Ministers, 2020). The advantage of internationalization would be the ability to combine much more data to generate value, reduce costs of future transactions and unify data exchange between entrepreneurs in the supranational arena. Market standards may be imposed or innovatively introduced not only by regulatory authorities but also by intermediaries including international organizations (Baron et al., 2019).

In the case of the industrial and agricultural sectors, cooperation between the parties is suggested by setting up data cooperatives: voluntary associations of an unlimited number of entities that, in the interest of their members, would carry out joint activities for the sharing of data from many farms. Alternatively, the role of a trusted entity can be played by a public agency, supporting the creation of a virtual shared data repository. The National Center for Agricultural Support such an entity. As a state executive agency, is located “close to the government” (which could facilitate cooperation and supervision over the implemented project). On the other hand, it would be able to stay in touch with farmers and promote the implementation of certain solutions at the local level (through its local branches, one in each province). The third model may be a contractual consortium with uniform rules and a standard for sharing data within the consortium, with the principle of reciprocity between members as regards the logic of data accessibility.

Contractual cooperation

A form consisting in creating links between business partners by combining their resources. It could consist in the parties’ commitment based on private law contracts (forms of cooperation; cooperation of several enterprises with their leader at the forefront), as well as partnership agreements of a public-private nature (implemented under the Act on Public-Private Partnership of Dec. 19, 2008). A solution similar to the legally known patent pool, i.e. a model in which inventors make their patents available for a predetermined price, and a given project may be purchased by entities that are party to the contract. In the case of data pooling, stakeholders from different sectors would conclude licensing agreements entitling them to use the “pool” to share resources with one another within the scope of their choice. Despite the similarity to patent pools, in the case of data-pooling we would be dealing with much more complex forms of cooperation. They would require not only access licensing contracts but also additional agreements for the processing technology needed to combine transferred data (Schneider, G., 2020).

For industrial data, data could be combined by building industry and cross-industry ecosystems on an online platform. The model would be based on the cooperation of a limited group of companies and limited access of entrepreneurs to closed and secure digital environments. The pooling of data is already in use and is of particular importance in agriculture, which is why this model is worth considering when creating pilots for shared repositories of agricultural data. It is pointed out that thanks to the pooling of data on agricultural fields, machines and weather, it is possible to use the so-called smart-farming tools, which then translates into more efficient use of resources and higher yields (Schubert and Harari Dayan, 2020).

BENEFITS

Leveling up opportunities

Access to shared repositories of agricultural data will strengthen the position of individual farmers and agri-food enterprises which mostly belong to the SME sector and do not have access to extensive digital resources and analytical tools. However, these companies should be confident that they can actually benefit from data sharing, avoiding the risks posed by the largest companies on the market (Nagel and Lycklama, 2021).

Increased availability of various data for machine learning

Advanced applications play a fundamental role in business processes and critical industries, including agriculture. Decision-making algorithms, Digital Farming systems, remote sensing, precise application of fertilizers based on soil and yield analyses, real-time tracking of yield and assessment of the fertilization effect allow not only to take care of the level of productivity and the condition of the entire enterprise but also to monitor the condition of individual plants (Council of Ministers, 2020). Federated analyses of distributed data make it possible to share results provided by AI tools without the need to share the original data. Thus, it is possible to guarantee security of data from individual farms and to ensure a balance between privacy, autonomy, protection of intellectual property (Big Data Value Association, 2019) and freedom.

Building new data-driven business models

The most innovative data-driven business models demonstrate a wide range of opportunities to build economic value – from directly monetizing data to building subscription-based platform access services.

Increasing productivity for the economy as a whole and improving competition

As has already been shown, the repeated use of data creates a lot of opportunities. However, so far, the most valuable digital information is limited by secrecy policies of the largest platforms – despite the fact that most of the data is not their product but only represent a series of phenomena occurring in the Internet space. The obligation to share data introduced for the most socially sensitive sectors could therefore increase production efficiency among smaller companies, which until now could not compete with the largest suppliers, which have huge data resources locked in private silos. In order to restore competitiveness, it is necessary to reduce the cost of accessing data by those who have not had such an opportunity so far.

Improved resource efficiency, increased productivity and sustainable solutions

Artificial intelligence tools for agriculture based on large data sets from sensors allow the use of solutions conducive to more sustainable agriculture. As experts point out, thanks to the so-called “smart-farming” agricultural producers could increase food production by as much as 70% by 2050 while reducing production costs and environmental exploitation (Nayyar A., Puri V., 2016). Agricultural data could also be used by state research institutions to obtain more accurate information about the state of the country's economy and would allow for a better understanding of local agricultural practices, which would then allow for more effective planning of the national agricultural policy (GPAI, 2021).

Based on the workshop conducted with representatives of the industrial and agricultural sectors

BARRIERS

Concerns about data sharing (socio-cultural barrier)

Workshop participants pointed to the fear of information leaks emerging among entrepreneurs, as well as improper use of data by external entities. In the case of cloud solutions, farmers have concerns primarily about security of their data due to the belief that it is better to store data locally in their own computers.

Reluctance to change common especially among the older generation of farmers (socio-cultural barrier)

It is a well known fact that rural areas are strongly attached to tradition, which results in a skeptical approach to new ways of managing farms and using innovative solutions.

Distrust of new technologies (socio-cultural)

Agriculture is an area characterized by low saturation with technology compared to other sectors of the economy. According to research conducted by the University of Agriculture in Krakow, even among young farmers who see the potential in modernizing farms there is a visible conservatism against the implementation of large-scale innovations (Kiełbasa B., Puchała J., 2015).

Internet white spots and lack of proper infrastructure (technical)

More than half of people (55%) who have never used the Internet live in rural areas (Bartol A., Herbst J., Pierścińska A., 2021). Among all rural residents, the elderly are particularly vulnerable to digital exclusion. Thus, the demographic structure of farmers (a small number of young people) and the location of farms in rural areas may result in difficulties in reaching for new digital solutions. Even if farmers are willing to use state-of-the-art technologies, they face problems related to insufficient broadband coverage of the country, dispersion of databases, lack of appropriate tools to support them, or insufficient number of sensors.

Lack of strong interoperability between systems and data exchange standards (technical)

Workshop participants pointed to different ways of collecting data, different data aggregation formats used by machines and the lack of interoperability between systems.

Concerns about the loss of competitiveness; reliance on trade secrets (socio-cultural / legal)

It is evident that there is still a belief among entrepreneurs that making their data accessible to other entities operating in the same or another sector may result in a loss of competitiveness and a deterioration of their position on the market.

5.3. Public shared data repositories

The model of public shared data repositories provides an alternative to corporate data silos, ensuring fair cooperation between the actors involved and conscious communization of the value generated. Digital data is, firstly, a record of reality (a digital copy) but in the information environment it gains natural features of common goods – it can be extremely simply replicated, while its sustainable use must be taken care of. The main goal of the public shared data repository is to create an ecosystem of trust for pro-social and pro-innovative use of data based on the principles of co-management and the established hierarchy of values (Zygmuntowski, 2020a). Therefore, it is an institution of a kind of data bank or infrastructure for the information society, combining government initiatives with bottom-up participation in the joint creation of value.

The value of aggregating this data is primarily used by a specific community – local, regional or pan-European. With the ability to access data previously held only by the largest companies, both SMEs and public administration, in addition to ordinary citizens, will get an opportunity to improve their internal processes through more in-depth data analysis, while ensuring protection of the public interest. Important aspects of the functioning of public shared data repositories are the co-management model, inclusivity and participation. A great example of such action is the model of the British National Institute for Health and Care Excellence (NICE), a non-governmental public institution that sets health guidelines with the participation of councils open to stakeholders.

FOR WHAT DATA TYPES?

- Personal data of public importance (e.g. health data)
- Data produced in the public service system
- Social data (e.g. from social media)
- Administrative data
- Local government data

WHAT FORMS?

Co-managed public institution

The operator of the repository, infrastructure and standards is a public agency from a given sector, for example the eHealth Center for health data. The operator also takes care of data quality, promotes its further sharing and grants access permits (similar to the European concept of data access body). Co-management (e.g. in the form of a supervisory board) guarantees that access to data, technical conditions and directions of development of the shared data repository are determined by a social representation and the risk of abuse by the state is minimized. The introduction of a single strong institution that safeguards both the public interest and the rights of data subjects empowers citizens, potentially enabling them to exercise their rights not in the form of breach reports (ex post), but directly in the form of a public service (Zygmuntowski et al., 2021).

Co-managed scientific center

Research institutions, private and public universities as well as R&D centers already manage scientific repositories and have extensive competences in the field of data management. They also enjoy significant public trust and, in the context of certain types of sensitive data (e.g. medical data), have access rights greater than non-scientific business enterprises. Therefore, the operator of the shared data repository may be a scientific center, but also assuming that the supervision over the data will be exercised by a non-scientific group representing other stakeholders, including the public administration itself.

STRUCTURE DESCRIPTION

The basic principle of shared repositories of personal data should be transparency of the intentions of the entity wishing to use the collected data for its own research. Their results should be made publicly accessible, or under other rules that protect the public interest, and the protection of the interests of individuals requires that an assessment of the effect of the algorithm be carried out on a case-by-case basis. Entities that would not comply with the established ethical principles and reciprocity rules at an initial stage would not be granted access. What is also important, in order to ensure an adequate level of data security, it is postulated that, in accordance with the “move algorithm to data” concept, external entities should transfer their algorithm to the shared data repository, avoiding transferring data outside the trusted and secure infrastructure (Hardjono and Pentland, 2019). This algorithm, after being sent through an appropriate interface, would perform calculations directly on the data collected in the shared data repository, later obtaining the results alone.

Due to the personal nature of data collected in shared data repositories, it is necessary to ensure an adequate level of privacy protection and data security. There are many technical methods to ensure an appropriate level of confidentiality for data, including homomorphic encryption, consisting in performing calculations on encrypted content, or differential privacy allowing to store information about groups in a data set in such a way that it is not necessary to disclose the data of a single person (Zygmuntowski, 2020a). In terms of security, choosing the right cloud infrastructure provider is also an important aspect. US-based companies are subject to US surveillance laws that allow government agencies (e.g. FBI) for virtually unrestricted access to non-US citizen data in certain specific cases (Konarski, 2020). In order to avoid being subject to these regulations, it is recommended to choose European service providers whose headquarters and data centers are located in the EEA. In the absence of such possibility, it is important to establish appropriate overarching rules for free flow of data between different entities subject to different legal regimes, establishing a well-functioning international metasystem for assessing conformity of both the internal regulations of a given supplier and the national regulations with established rules.

Maintaining a complex technical architecture requires adequate financial resources. Due to the neutral nature of the project, the chosen business model should be as self-sufficient as possible. A shared data repository can be supported primarily by license fees paid for access to its data via APIs if it concerns data of a higher order than raw or structured data. It should also introduce criteria for differentiating amounts of fees, which would be lower for entities operating non-commercially, conducting scientific research and social initiatives, and correspondingly higher for those who use data for the development of their own business (Zygmuntowski, 2020a). Due to the broadly understood public nature of shared data repositories, relying on grants from public institutions or NGOs also seems to be a good solution.

BENEFITS

Strengthening trust in data sharing

Involving all stakeholders in the process of managing shared data repositories could gradually restore confidence in public initiatives aimed at creating Common Social Value.

Improving the predictive quality of ai systems – the more data the better; positive external effects of data aggregation

According to the assumptions of the “Policy for the development of artificial intelligence in Poland”, it is estimated that the development of artificial intelligence in Poland will contribute up to 2.65 pp to the economy’s GDP each year. Over the next eight years, AI will allow for the automation of approx. 49% of working time in Poland while generating better paid jobs in strategic sectors (GovTech Polska, 2020). However, in order to be able to properly operate in the areas identified as the most important, a key factor in the development of artificial intelligence is to provide it with as much qualitative data as possible, which will be complete, properly described and labeled (Kawalec, 2021). The more information about a particular phenomenon we feed to the algorithm, the better the prediction generated by the system will be. The institution of the shared data repository, due to its advantage of aggregating large amounts of data, offers an excellent potential for use by AI models.

Positive network effects resulting from cooperation between different actors

The outbreak of the coronavirus pandemic in 2020 has made the technological transformation postponed by many entities indispensable to survive the crisis. More than half of Polish companies accelerated their digital transformation during the pandemic but, of all possible technological tools, big data analysis tools and predictive systems were the least popular (Ernst & Young, 2021). This could be due to too high costs of such solutions or to incompleteness of data sets for proper use by artificial intelligence models. Collecting data from various entities could encourage companies, universities and the public sector to combine their resources (financial, intellectual and organizational) and cooperate within specific sectors of the economy to find the most effective solutions.

Project example: shared repository of health data

The largest national health database. Data collected as part of it is used to improve quality of the health care system, design modern telemedicine solutions and conduct groundbreaking scientific research. The National Health Fund, the State Sanitary Inspectorate, the Center for Healthcare Information Systems, private medical networks, registers of medical entities (medical facilities, hospitals, etc.), insurance and reinsurance companies, pharmacies, as well as smart devices and health and medical applications are the sources of data going to the shared data repository. Due to the infrastructure and experience in collecting data from different systems and securely managing existing resources, the eHealth Center (CeZ) could be the shared data repository operator.

Based on the workshop with representatives of the health care sector

BARRIERS

No unified (technical) standards

Regulation of the Council of Ministers of 12 April 2012 on the National Interoperability Framework, minimum requirements for public registers and the exchange of information in electronic form and minimum requirements for electronic systems (hereinafter referred to as the “NIF Regulation”) contains specific requirements for maintaining appropriate standards only for public sector entities. Although 68% of medical institutions have IT solutions that allow them to keep records in electronic form, 93% of respondents prefer paper as the most popular form of exchanging information. What is more, almost 70% of institutions do not enter such documents into the system (e-Health Center, 2021).

Multitude of systems used in the health care sector (technical)

Although, compared to 2018, the budget for the digitization of the healthcare sector has almost doubled, the Ministry of Health does not have information or control over which IT systems are chosen by public health care institutions. However, they have a duty to ensure at least poor interoperability with the central digital architecture of the health care system (Minister of Health, 2022). There is no regulation regarding the harmonization of standards for health data and for supporting the creation of the so-called strong interoperability within the entire healthcare sector, i.e. both public and private institutions. Due to the lack of compatibility of systems, patients are often forced to independently transfer data about their health from private to public institutions and vice versa, thus generating higher costs of diagnostics and treatment.

Different data collection cultures in various institutions (socio-cultural / technical)

The lack of adequate regulations regarding standards and describing specific cases often results in different ways of collecting, labeling and structuring data in healthcare facilities. An example of this state of affairs can be the different blood group determination methods used by hospitals, which lead to numerous confusions.

Inability to ensure complete secrecy of personal information* (technical)

Due to the specificity of data pseudonymization, i.e. the technique of encrypting data using separately held passwords, it is impossible to ensure one hundred percent security. If someone has access to the keys that allow “decoding” of this type of data, there will always be a risk and a certain degree of probability that this data will become fully accessible to unauthorized persons. In addition, as health data are often composed of many different pieces of information, it is difficult to assess in the absence of which of them further identification of the data subject is no longer possible and which, when compared with other data at hand, make it possible to identify the data subject.

Distrust of sharing health data (socio-cultural)

According to the report of the Polish Economic Institute, as a rule, Poles are not willing to share their data. Less than half of the respondents (45.2%) would be ready to share data on their health habits for the needs of a public prevention program (Grzeszak J., Łukasik K., Święcicki I., 2021). This state of affairs is primarily due to the fear of processing data in bad faith and the possibility of using the results of processing against the data subject. The associated fear of surveillance by public authorities was particularly evident in the context of the public’s reaction to the pregnancy register project. In accordance with the amendment to the Regulation of the Minister of Health of June 26, 2020, on the detailed scope of data of a medical event processed in the information system and the method and dates of transferring these data to the Medical Information System, all entities providing medical services will be obliged to enter data on pregnant patients to the medical information system. Although the reasons for the changes were legitimate (e.g. prescribing drugs to pregnant women, using doctors’ records out of turn, saving lives), due to the political context associated with getting pregnant in Poland, many people were inclined to presume that, in fact, the regulation is aimed at exercising control over citizens (PAP, 2021).

Lack of mutual trust of actors involved in the data sharing process (socio-cultural)

According to the latest Edelman Trust Barometer research measuring the level of citizens' trust in particular sectors, the government and the media are behind business and NGOs (Edelman Trust Barometer, 2022). Moreover, despite attempts at cooperation between these entities, they also do not trust one another.

Lack of understanding of the relationship between the "public interest" and the "private interest" and the resulting difficulty in determining the individual benefit of data sharing (socio-cultural)

The public interest, both in law and in the general public perception, is usually associated with the restriction of the fundamental rights and freedoms of individuals in favor of the good of a particular community (e.g. security, health or public order). In addition, many people associate the public interest with the government and the possibility of unwanted surveillance by public authorities. In view of such a panorama of associations, the question arises among many people about what is the benefit for an individual citizen of sharing their data with a wide range of unknown entities. As research conducted by the Polish Economic Institute shows, Poles are most likely to share data with their loved ones and nurses (Grzeszak J., Łukasik K., Święcicki I., 2021).

No clear definition of scientific objectives (legal)

Although art. 9(2)(j) of the GDPR refers to "scientific purposes" as one of the grounds for the processing of special categories of data, it does not contain a legal definition of these categories. According to recital 159 of the GDPR, "scientific purposes" include:

- technological development and demonstration;
- basic research;
- applied research;
- privately funded research.

As indicated in the recital, the expression "for the purpose of scientific research" should also cover research carried out in the public interest in the field of public health, also allowing private establishments. In this context, the GDPR also indicates the need to create a European research area. However, the recitals contained in the GDPR are only a certain direction of interpretation and cannot be the legal basis for possible justifications for data collection.

In view of the fact that member states have a certain regulatory freedom, Polish legislation has adopted a broader concept referring to R&D works and scientific activities including, inter alia, scientific R&D works (the Law on Higher Education and Science). The inability to determine how the Polish definition refers to the one contained in the GDPR makes it much more difficult to determine which of the set goals belong to the permitted scientific purposes.

In accordance with art. 26(4) of the Law of Nov. 6, 2008, on the rights of the patient and the Patient Ombudsman, medical documentation may also be made accessible to a university or research institute for use for scientific purposes, without revealing the name or other data enabling the identification of the person to whom the documentation relates. This means that other entities, such as private companies and organizations not included in the list, may not use the possibility of restricting the application of the GDPR on the same terms as universities and research institutes (and thus bypassing art. 15 of the GDPR concerning the right of data subjects to access their data) (Najbuk P., Pachocki J., Kruczyk-Gonciarz A., Kaźmierczyk P., Lorent R., 2020). However, private operators may benefit from art. 9(2)(j) of the GDPR and conduct research activities, without omitting other provisions.

No legal definition of health data and its relationship to medical data (legal)

In addition to data that is clearly related to health (e.g. information from Electronic Medical Records), there is a number of data that, in combination with others, can form the basis for drawing conclusions about someone's health condition. This difficulty in demarcation applies in particular to fitbit bands and sports applications which can measure not only a person's fitness but also their heart rate or sleep quality. The lack of an unambiguous definition, both in the EU and in the Polish law, means that many data types that are on the border of "ordinary" data and those types of a special nature can be automatically treated with less caution than it would be in the case of sensitive data.

Closing data in databases of private companies (legal)

Large technology companies (such as Amazon, Google or Apple) enter the medical marked and succeed there, offering their users convenient solutions for measuring everyday activities and drawing conclusions about their health on this basis. However, these companies do it for their own commercial purposes, not wanting to share data with local governments or other entities working for the public good.

5.4. Methods for managing highly sensitive data

5.4.1. Non-personal data

Profiting from digital data collections by one entrepreneur does not exclude their usefulness for another entity (Paszcza, 2022). However, despite the possibility of multiple uses of data, its transfer and re-use still faces many barriers.

Non-personal data is understood as electronic information types other than those named in the GDPR – that is, they do not involve information about an identified or identifiable natural person. They are often not man-generated, although they can be collected, processed and used by humans. Examples include digital information produced by machines (derived from various types of sensors and sensors) and electronic products (e.g. anonymized BigData collections), as well as meteorological and natural data. It happens that access to them is fully open, as in the case of data on the volume of traffic on roads. Other times data may be controlled by a specific organization or entrepreneur such as a software developer or machine owner. However, with regard to the issue of accessibility of non-personal data, it is pointed out that if its source is the (natural) environment or if it has been anonymized (e.g. data on consumers), they should not be restricted and monopolized (Borowik M., Maśniak L., Kroplewski R., Romaniec H., 2017). Data from internal production lines, although often owned by enterprises, could be used wider through industrial re-use due to their stimulating impact on the entire economic ecosystem (Borowik M., Maśniak L., Kroplewski R., Romaniec H., 2017).

It happens, however, that entrepreneurs refuse access to data sets they aggregate, most often citing issues of privacy (trade secrets) and intellectual property rights. When it comes to trade secrets, in the case of non-personal data we can talk about information coming from websites, devices or machine sensors, taking the form of, for example, text, numbers or images, structured or unstructured. However, exercising the right to protect this data requires adequate protection of the collections against access by unauthorized persons and having an economic value or belonging to a set or compilation that represents such value.* Despite the fact that some of the digital information contained in the company's resources actually meets the above-mentioned conditions, many entrepreneurs abuse the rights regarding trade secrets resulting from the act on combating unfair competition, thus inhibiting the development of not only their own company, but also the entire economy.

* In Polish law, issues related to trade secrets are regulated primarily in the Act of April 16, 1993, on combating unfair competition. It follows from art. of the Act that a trade secret is understood as technical, technological, organizational information of an enterprise or other information having an economic value, which as a whole or in a specific combination of its elements is not commonly known to persons usually dealing with this type of information or is not easily accessible.

As far as intellectual property rights are concerned, some data in the digital economy, such as digital goods (music works, e-books and software), are actually protected by copyrights. In relation to these data, their creators are entitled to exclusive property rights. However, much digital information, especially machine-generated, does not meet the requirements for copyright protection (Kerber, 2016). The situation is similar regarding the sui generis right to databases: while it is granted to database producers, the rights resulting from it are often abused by data controllers. The sui generis protection already mentioned is granted only if a substantial investment has been made in connection with a database for the purposes necessary for the acquisition, verification or presentation of the database (art. 7(1) of Directive 96/9). Database protection thus protects the investment made to collect and organize already existing digital information, not the mere production or collection of data (Kerber, 2016). In the context of Big Data, it is worth noting that while large collections can be (indirectly) protected both by copyrights and under the sui generis regime, this protection will never cover the content of databases, i.e. data "as such" (Żyrek, 2022).

In summary, data from internal systems of companies, as a rule, will be subject to protection by these enterprises. However, due to their high economic potential and the possibility of developing value chains through mutual exchange of information, the concept of mutual access to data should be promoted and implemented on a large scale (Borowik M., Maśniak L., Kroplewski R., Romaniec H., 2017). It is also necessary to look at the existing barriers to the exchange of data between different economic actors. The abuse by entrepreneurs of the right to the protection of databases and intellectual property rights most often results from low awareness of the benefits of data sharing in business and fears of losing the company’s competitiveness. However, there are also voices of business representatives who point at uncertainty in the security of shared data and fear of their possible leakage and improper use. Building an IT system that allows for effective data sharing while ensuring maximum security and control is difficult due to various legal and technical statuses of databases (Dymek, Komnata, Kotulski, 2011). However, this is not an impossible task. Thanks to various types of technical solutions, it is possible to mitigate risks associated with the exchange of data, as well as to ensure confidentiality of data and competitiveness of individual entities.

Federated model of virtual shared repository of business data

The virtual shared data repository model analyzed in this report would be based on a federated structure which in itself would constitute a kind of security. The system designed in this way allows for mutual sharing of data, but only to the extent resulting from the needs of the shared data repository, while maintaining all security requirements and without the need to disclose details regarding the construction of individual databases. In addition, it makes it possible to ensure supervision over data exchange processes at the internal and external levels (at the level of the shared data repository). This concept can be compared to the secure transfer of confidential data between law enforcement authorities, which is based on the inquiry-answer mode. In this mode, thanks to the use of unified inquiry forms, the inquiring authority does not need to know the database model or structures of the questioned body (Dymek, Komnata, Kotulski, 2011). It could be similar in the case of virtual business shared data repositories: an entrepreneur would provide the entity operating the shared data repository with only standardized raw data, without having to disclose the results of data analysis carried out for the entrepreneur’s company.

Federated learning

Federated learning emerges as a new paradigm of collaboration and partnership between companies. It allows companies to share their collections in a “closed system” and allows them to build a common efficient machine learning model without the need for actual data exchange. Federated learning provides the ability to “train” the algorithm while storing data at the device level of individual entities. The entrepreneur keeps their private local data on corporate devices, thanks to which they minimize their concerns about security. Thus, due to the fact that, in the case of federated learning, all data required to train the model remains under strict supervision of the organization, this model can be used in numerous sectors, such as healthcare, industry and e-commerce. This model can be a technical layer of cooperation within virtual shared data repositories.

Edge computing

Edge computing is an approach based on the dispersion resulting from the ability to process data directly on smart devices, such as mobile phones and networks. Data processing takes place locally: the “edge gateway” processes data from the device and, only after it is structured, the gateway sends relevant data to the server for further storage. Thanks to this solution it is possible to process data as close as possible to the “source”, which allows to accelerate the response of the device and reduce delays in operation (Velotio Technologies, 2019). It should be noted, however, that due to the multitude of individual devices connected to the server, it is necessary to take special precautions. In order for companies to guarantee full data security, they should ensure that every element of data stored on the company’s devices is encrypted. It’s also a good idea to always make sure that connectivity uses multi-factor authentication and SSL/TLS certificates or similar corporate-level security solutions.

Edge computing could be used for shared repositories of industrial and agricultural data, as it includes a wide range of digital tools such as networked sensors (they remain connected to a central server and transmit information in real time). Edge computing could also greatly benefit the machine learning used for many business tools. It is proposed that algorithm training should be performed in the cloud and, next, ready models should be implemented on edge devices in order to forecast phenomena.

Blockchain technology

Blockchain and distributed ledger technologies (DLT) enable organizations to verify stakeholders and transparently check data access permissions through the use of cryptographic methods (Bechtel, Buchholz, 2022). Users of this solution can view the history of operations stored in the ledger (e.g. consents and access instances), however, access to data requires appropriate permissions. Any attempts to manipulate or carry out an unauthorized operation are immediately detected and blocked (the system does not allow them to be included in the blockchain). The only attack vector may be registry manipulation but this requires control over 51% of the ledger confirmation nodes, which is much more difficult than taking control of a single server. What is more, blockchain features high resistance to IT infrastructure failures due to the fact that data is saved simultaneously in the memory of not one but thousands of servers. When it comes to using this technological solution in business, a good example is the transport and logistics industry in which, thanks to the system, it is possible to keep a register of drivers, share trips and store information about the history of vehicles. It is also possible to control shipments in real time and exchange information between all participants in the supply chain while ensuring security and transparency.

The blockchain technology could therefore gain particular importance in the context of designing a virtual shared data repository for business data. Thanks to the blockchain it would be easier for companies to manage data, certificates, digital products created together and, at the same time, the fear of losing control over data would be reduced.

5.4.2. Sensitive data

Sensitive data require particularly enhanced protection due to the often confidential and intimate nature of information about the data subject. Its processing can be a rather serious interference in the private sphere of human life, at the same time constituting a basis for discrimination against such a person (Fajgielski, 2021).

The GDPR enumerates the categories of data considered to be special categories of data:

- data on racial/ethnic origin;
- data on political views;
- data on religious and philosophical beliefs;
- data on trade union membership;
- genetic data;
- biometric data processed for the purpose of unambiguously identifying a natural person;
- health data;
- data on sexuality or sexual orientation.

Unlike under the regime of the pre-GDPR directive on the protection of personal data, member states may no longer extend the catalogue of these data types or add other types. However, for some data types they may adopt exceptional regulations protecting individual interests of individuals. This applies, for example, to the obligation of banking secrecy expressed in the Banking Law, covering information on banking activities obtained while negotiating, signing and performing a data processing agreement (the Banking Law).

As a rule, the processing of sensitive data is prohibited unless one of the nine conditions listed in the GDPR applies. The following are among those that may be relevant to data sharing models:

- Explicit consent of the data subject;
- The necessity of processing for the fulfillment of obligations and the exercise of specific rights by the controller or the data subject in the areas of labor law, social security and social protection;
- The necessity of processing to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent to the processing;
- The necessity of processing for an important public interest, provided that the processing is proportionate to the objective pursued, does not undermine the essence of the right to data protection and provides for appropriate and specific measures to protect the fundamental rights and interests of the data subject;
- The necessity of processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, provided that the processing is proportionate to the purpose pursued, does not affect the essence of the right to data protection and provides for appropriate specific measures to protect the fundamental rights and interests of the data subject.

An important basis in the context of the model for processing special categories of data, especially in the initial phase of the creation of its infrastructure, is the requirement of the explicit consent of the data subject. It should not be implicit, which means that the person concerned should clearly express their consent to the processing of specific personal data for the purposes named by the controller.

The controller should also use only such data that is appropriate for a given purpose and limited to what is necessary, choosing those sets that are needed for processing (principle of data minimization). In the model of sharing special categories of data it will be extremely important to clearly define the planned projects and clear reasons why data is collected – even before the data subject's consent. Where it is desirable to extend the purposes, it will also be necessary to inform the data subjects in advance.

An interesting solution is the Finnish opt-out regulation in which the sharing of data for common purposes takes place by default and users can opt out of further sharing of their data, if necessary. Due to the role of the public interest in the development of tools for improving health care, such a solution seems to be the most effective.

Despite the requirements and challenges imposed by the GDPR, it is possible to provide such an architecture of the data sharing model to ensure adequate privacy protection and data security already at the design stage (privacy-by-design). To this end, however, conditions should be created to ensure adequate internal security of the system (consisting in appropriate data secrecy) and external security, regarding the resilience of the system to entities that do not meet their own standards set by a given data sharing model, and the principles resulting from the EU law.

Private cloud

Sensitive data requires higher security standards to be ensured both in terms of its collection and subsequent use. Due to its extremely confidential nature, it is important to have adequate control over the system and the ability to choose infrastructure and security elements. For this reason, it is recommended to store sensitive data in a private cloud providing the ability to restrict access to its resources.

According to research conducted by VMware, the main reason for organizations to choose a private cloud is the ability to control data (VMware, 2017). Thanks to the private cloud, its resources can only be used by authorized users who either belong to the organization managing the collected resources or have received permission to use the data in a specific way. Despite the lack of general accessibility, the private cloud has several features similar to those of the public cloud, including flexibility and scalability (OVH, 2018). It also has several special advantages of its own, including a low entry threshold and no capital expenditure. It is a good solution for organizations wishing to use the computing power of a cloud service provider, while being able to develop their own proprietary solutions.

When choosing an infrastructure, we should also take into account certificates held by the provider and the services it offers, to make sure that there are an information security management system (ISMS) and a privacy management system (e.g. OVH) in place able do deal with risks, vulnerabilities and challenges to business continuity.

Centralized and distributed data architectures

Despite the common view that a centralized database system and a distributed system are two disjoint ways of organizing a data asset management architecture, in reality they can complement each other. On the one hand, the centralized system consists in storing all data within a single database, making it easier to design and manage, and in the event of any system failure, it is easier to restore its state based on its backup while maintaining the fullest possible consistency. On the other hand, having one large data center for a large geographical area is vulnerable to communication link failures. For this reason, a more natural solution in this situation would be to create a distributed system, which is additionally distinguished by greater processing power – extremely important in the context of processing large data sets (Wojciechowski, 1998).

Building a hybrid architecture with elements taken from the centralized and the distributed models is a recommended solution, especially in a situation where data storage on edge devices may pose a risk to privacy protection or is significantly difficult (Zygmuntowski, 2021a). Thanks to this, data can be collected on one or more servers belonging to the organization, guarding the adopted conditions for its collection and processing.

MIT open algorithms (OPAL)

The processing of sensitive data is associated with the occurrence of increased risk not only due the lack of adequate security, but also in the event of data leakage and its falling into the wrong hands. In the case of granting access to a database, external entities can copy data to their servers, as a consequence of which data may be used uncontrollably in bad faith or in contradiction with the established principles of a specific data sharing model.

To prevent this situation, we propose to move away from data sharing understood as the transfer of data to another entity's database, in favor of the "move algorithm to data" concept developed by the Massachusetts Institute of Technology. This concept involves transferring an algorithm belonging to an external entity to the edge devices of data sharing models (e.g. shared repository of health data) – in such a way that sensitive data never gets out of the secure environment of the organization's repository. It is possible to adopt additional safeguards to prevent data from flowing out from another site by using homomorphic encryption, which consists in operating on encrypted data without having to "declassify" it (Hardjono and Pentland, 2019).

One of the assumptions adopted by this concept is also the issue of verifying the transmitted algorithms – to make sure that they are "free" from prejudice, discrimination or privacy breaches. The results returned to the owner of the algorithm in the form of aggregated responses must be accurate enough not to allow the recipient to carry out correlation attacks that could lead to the re-identification of individuals. If, on the other hand, the algorithm seeks to obtain answers specific to the data subject, the operations carried out through the algorithm may only be carried out after obtaining the confirmed and fully informed consent of the subject (Hardjono, Pentland, 2019).

Blockchain technology

Although blockchain technology is used for decentralized data collection, due to the general availability of the ledger and its contents, it is not a good solution for storing sensitive data. However, due to its transparency and inability to interfere with data inputs, blockchain could serve as a register of permissions granted both to external entities transferring their algorithms to data sharing models and to persons who have given their consents to the processing of their data. This would ensure public control over the access granted and the effective transparency of any data operation.

Data protection engineering

In view of the development of new ways of storing data and the emergence of new threats to data collections, ENISA has published recommendations on data protection engineering, specifying the technical and organizational processes that, already at the design stage and at the default phase of creating a data protection structure, will allow to ensure an adequate level of security, meeting the most important objectives of infrastructure development, which include integrity, confidentiality, accessibility, ability to intervene and lack of relationships (ENISA, 2022).

Particularly noteworthy are privacy enhancing technologies which take the following as the main principles of application of individual technical solutions:

- Preserve the truth: The goal of privacy engineering is to preserve the veracity of data while reducing its identifiability. This can be achieved, for example, by reducing granularity of the data (e.g. from date of birth to age). Then, the data is still accurate but in a "minimal way", suitable for the purpose in question.

  The recommended method in this respect is data encryption consisting in the transformation of explicit and open information into a cryptogram, i.e. encrypted text, that can be "declassified" using separately held keys (Bitdefender, 2022). This technique can be considered to preserve the veracity of the data set, since encryption applied in the opposite direction fully restores the original data without introducing any uncertainty into the process (ENISA, 2022).
- Maintain readability: The data is stored in a format that "makes sense" to the controller, without revealing the actual attributes of the data subjects.

  One of the methods used to maintain legibility may be the so-called "differential privacy", which involves adding noise to actual data, which does not have much of an impact on its overall usability (Kaczmarek, 2022). This does not change the overall impression and characteristics of the data set, but ensures its confidentiality.

Other methods to ensure adequate data security include:

- Homomorphic encryption: allows to perform calculations on data without the need for decryption;
- Synthetic data: creating data that resemble real data (e.g. distribution of aggregate values) but are not related to any existing natural person; the production of this data is aimed at manipulating the possibility of re-identification of natural persons.

Data leakage / loss prevention

Due to the fact that data types of various degrees of sensitivity are stored within a single database, it is in practice impossible to independently trace and detect those that require special protection. The solution to accelerate this process are data loss prevention software systems that use a set of tools and processes to ensure that sensitive data is not lost, misused or made accessible to unauthorized users.

Data loss prevention (DLP) software classifies data, distinguishing, among other things, confidential and critical business data, identifying violations of policies defined by a given organization or those resulting from generally applicable regulations (e.g. HIPAA, PCI-DSS or GDPR). For systems containing sensitive data, the DLP system can identify, classify and flag it appropriately, as well as monitor activities and events associated with the data. In addition, reporting capabilities provide detailed information required for compliance audits. Once breaches are identified, the system enforces countermeasures in the form of alerts, encryption and other protective actions to prevent end users from accidentally or maliciously sharing data, which could pose a threat to the organization (De Groot, 2020).

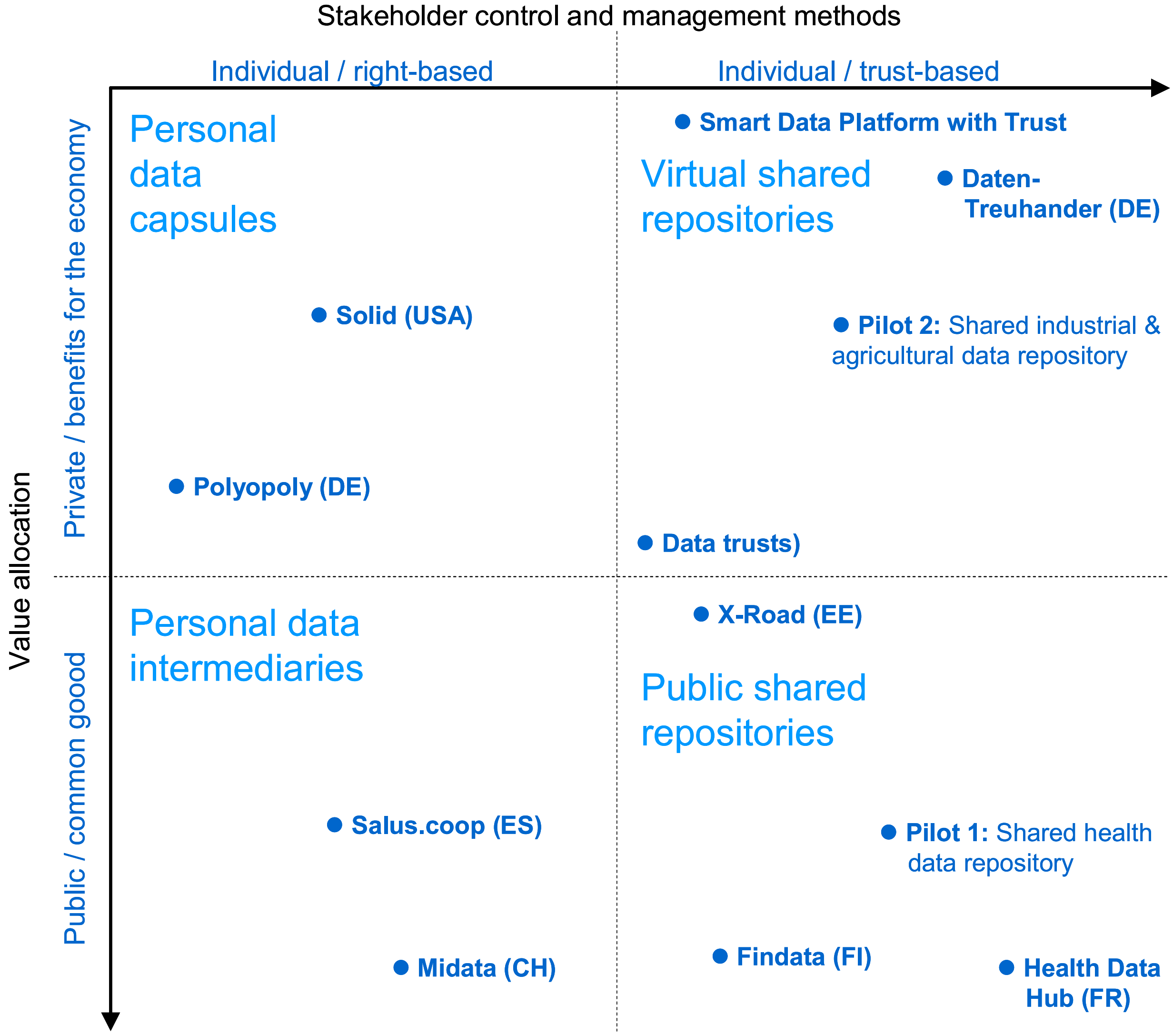
Compliance of cloud service providers

The real challenge is also to ensure resilience to external threats, i.e. to strengthen the digital sovereignty of Poland on the international arena. This necessity results from discrepancies in the level of protection of personal data provided by EU regulations and regulations of third countries. This applies in particular to the US law, the unclear scope of which was the reason for Maximilian Schrems' complaint to the Irish Data Protection Authority. In the aftermath of the claim, it has been established that the transfer of data to the US based on the previously applicable “Privacy Shield” is prohibited. The CJEU justified this decision by citing the regulations underlying the functioning of intelligence surveillance programs. They remain unrestricted as to the ability to interfere with the privacy rights of non-US citizens. In addition, in some cases, the right to challenge court decisions that may grant certain authorities the power to conduct surveillance of non-US persons has not been established.

Although, after the CJEU's judgment in the Schrems II case, it is possible to transfer data to the USA based on standard contractual clauses, decisions are still issued that the use of US-based providers is inconsistent with the GDPR's requirement to meet the same conditions for the protection of personal data. Due to the reluctance of the US authorities to change their regulations, it is necessary to consider the choice of European IT solutions so as to ensure that citizens' data, both in terms of proper encryption and storage in the infrastructure, are not in any way threatened. Another option may be to set standards necessary for the cloud provider to meet, including, among others, requirements for having appropriate certificates or giving the privilege of maintaining keys to encrypted data to the institution using the cloud (Datailsynet, 2022).

6. Pilot conclusions and recommendations

Based on the workshops and expert research, we propose to implement pilots of two data sharing models: the virtual shared repository of industrial and agricultural data and the public shared repository of health data. The role of the state in this respect should be to fully finance such innovations and coordinate activities, inter alia, by establishing a legal framework and rules of their operation, going beyond the pilots themselves, taking into account the legal and contractual layer, technical standards and organizational processes.


Stakeholder control and management methods
Individual / right-based
Individual / trust-based
Value allocation
Private / benefits for the economy
Public / common good
Personal data capsules
Solid (USA)
Polyopoly (DE)
Virtual shared repositories
Smart Data Platform with Trust
Daten-Treuhander (DE)
Pilot 2: Shared industrial & agricultural data repository
Data trusts)
Personal data intermediaries
Salus.coop (ES)
Midata (CH)
X-Road (EE)
Public shared repositories
Pilot 1: Shared health data repository
Findata (FI)
Health Data Hub (FR)

Source: Diagram based on the research workshops on the „Unleashing of the potential of data"

6.1. Pilot of the virtual shared repository of industrial and agricultural data

New technologies are playing an increasingly important role in global agriculture and industry. The efficiency of production and the performance of the economy are influenced by standard factors (e.g. skills of employees), technological developments and the intensity of data use (Koloch G., Grobelna K., Zakrzewska-Szlichtyng K., Kamiński B., Kaszyński D., 2017). Countries caring about competitiveness of their agri-food, logistics or industrial sectors on the global market make significant investments in the so-called "smart farming" and in the IoT. In order not to lag behind and not to expose Polish farms to falling into technological debt, it is necessary to develop new institutions in this area. This is particularly important in the context of leveling up opportunities of Polish farmers on the EU arena. Despite productivity growth in recent years, it is still lower than in the leading countries of Western Europe (and is close to the level of France and Germany in the 1970s) (Miniszewski M., 2021). It is possible to increase productivity by implementing innovations in the form of new technologies including, among others, big data, precision agriculture, plant breeding technologies.

→ See page 62

It is worth noting that although Polish industry is characterized by a relatively low intensity of data use compared to other European countries, data has a high share in the productivity of the economy (even compared to European leaders) (Ministry of Development, 2020). Thus, "delaying or even abandoning actions aimed at developing socio-economic conditions conducive to development of a highly data-driven economy will result, already in the medium term, in a significant reduction of achievable economic benefits" (Koloch G., Grobelna K., Zakrzewska-Szlichtyng K., Kamiński B., Kaszyński D., 2017).

During the workshop we identified 10 key success factors that determine the main challenges to be solved during the pilot. Attention is also drawn to factors that, in the course of the workshops, significantly gained or lost on importance in the eyes of the stakeholders.

Fig. 6: 10 Key success factors for a virtual shared repository of industrial and agricultural data

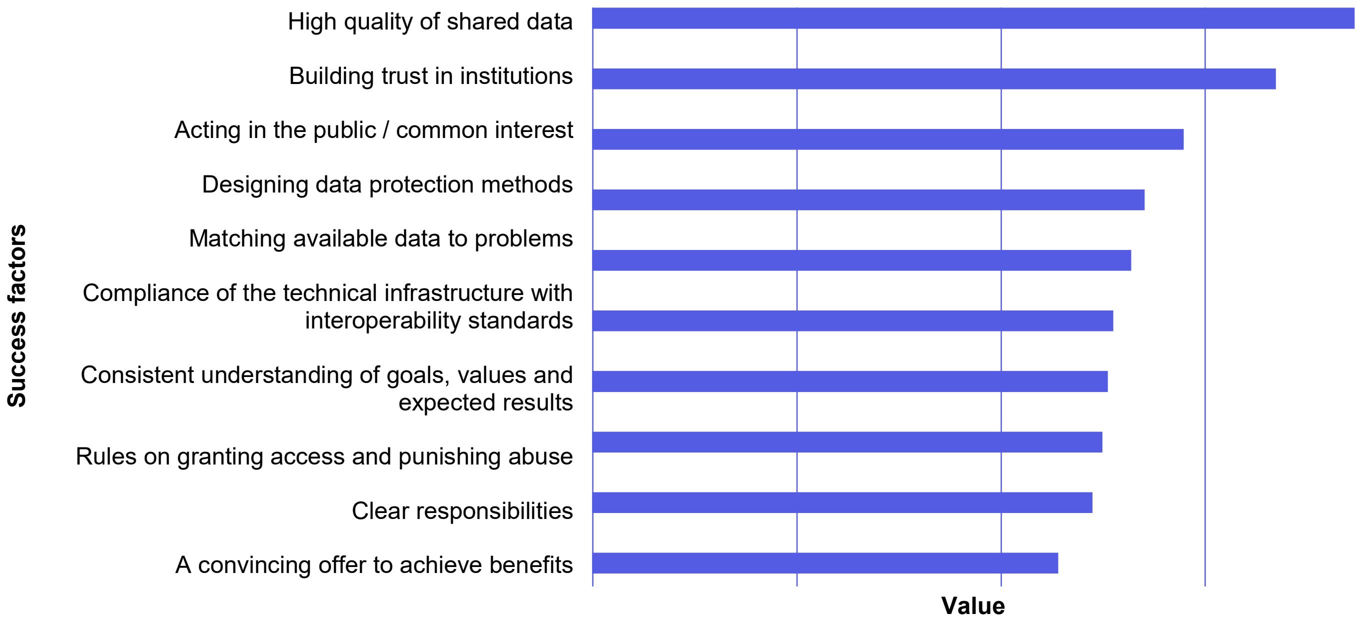
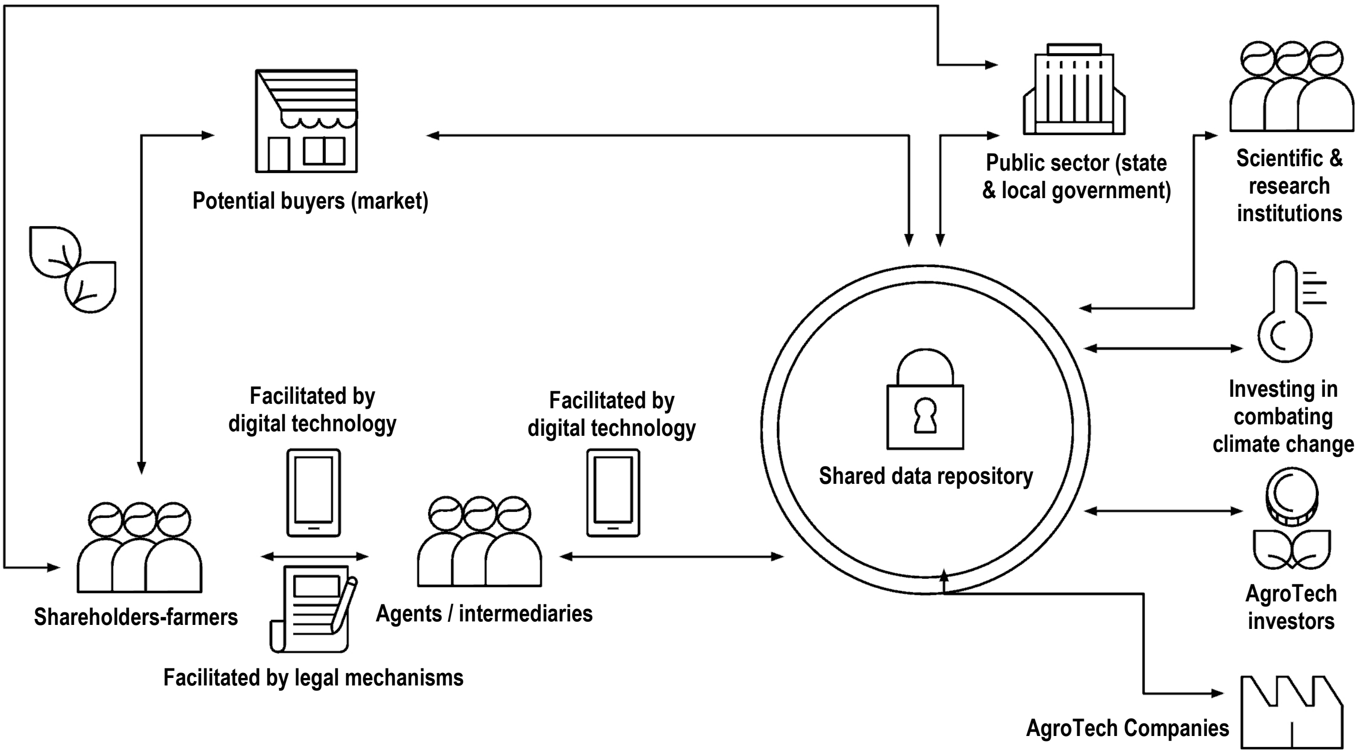


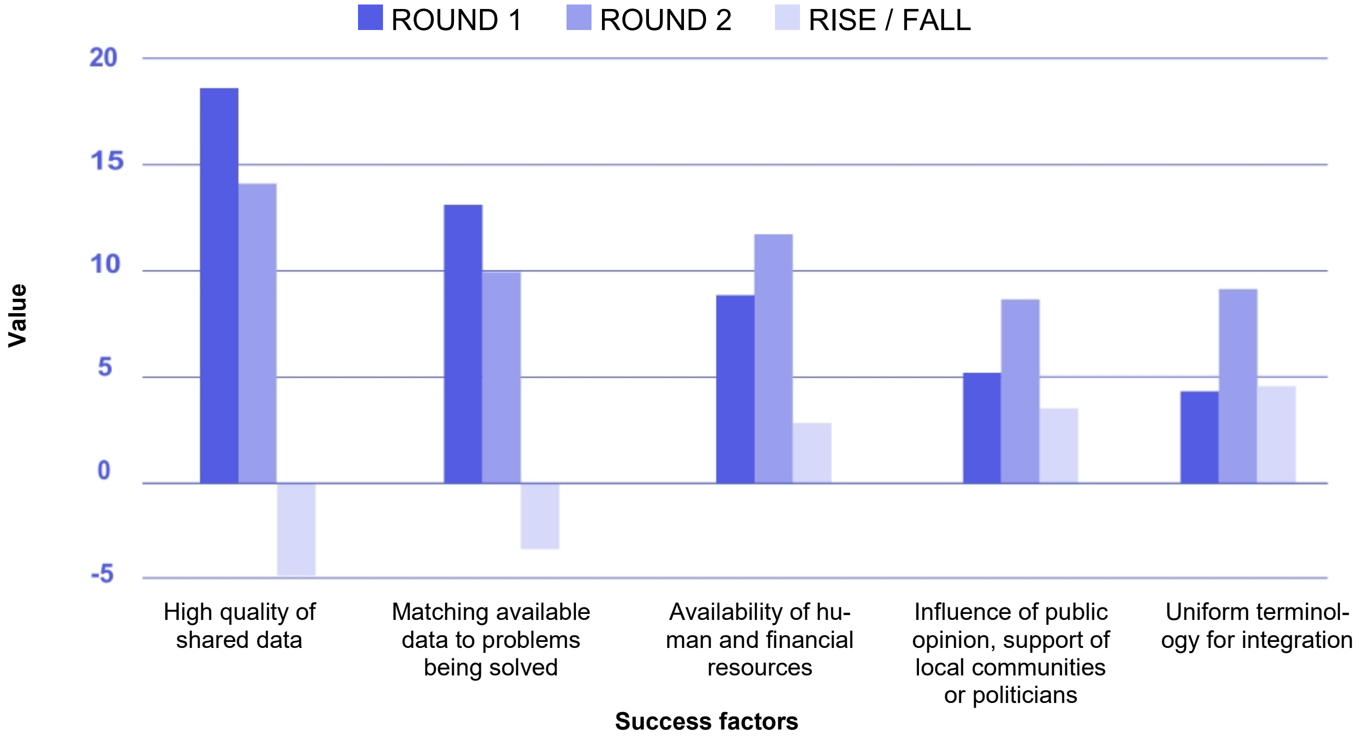
Fig. 7: An example of a virtual data repository model



Potential buyers (market)
Public sector (state & local government)
Scientific & research institutions
Investing in combating climate change
AgroTech investors
AgroTech Companies
Shared data repository
Facilitated by digital technology
Facilitated by digital technology
Shareholders-farmers
Agents / intermediaries
Facilitated by legal mechanisms

Source: The Global Partnership on Artificial Intelligence (GPAI) (2021) Enabling Data sharing for Social Benefit Through Data Trusts: Data Trusts in Climate

Fig. 8: 5 factors that have changed the most during the workshops



WE PROPOSE TO BASE THE PILOT ON THE FOLLOWING STEPS:

1. Designation of pilot datasets and a sharing standard

Since quality and volume of data is fundamental to the success of the project, the creators of the shared data repository must identify specific data sets that are available, processed and have features that make them relatively simpler to share than sets requiring intensive cleanup or patching. Such data sets may be more industrial or agricultural, they may function on the border and may come from various types of entities. A functional key is more important than a strict definition of a sectoral boundary. Further, a standard for sharing this data should be defined so as to allow for interoperability of the systems.

2. Selection of a legal formula, objectives and strategy of the shared data repository in a transparent industry dialogue

The recommended legal formulas are primarily a data cooperative (as a trusted entity), designation of a public agency as a trusted entity or contractual cooperation. However, building trust in data sharing will be successful only if the choice of a form of the virtual shared data repository, goals, values, methods of achieving them, and the expected results is made in a dialogue of the most open nature. Regardless of whether the shared data repository starts with industrial or agricultural data sets, the consultation should also take into account future participants in the sharing and potential data users (science, business analytics, innovation). The involvement of industry associations would make it possible to develop conditions under which cooperation in the exchange of data between companies should take place. How competition rules should be respected (in particular as regards membership requirements, type of data shared, third party access). How intellectual property frameworks affect the allocation of data rights in trusted data spaces. The success in implementing new concepts in rural areas is closely correlated with how the proposed solutions are perceived by the local community. In the case of shared repositories of agricultural data, it would be reasonable to involve various entities active in rural areas – not only potential beneficiaries of the shared data repository (farmers), but also socio-professional organizations (e.g. farmer societies, country housewives' clubs, agricultural trade associations) and representatives of local churches.

3. Development of access rules and data protection methods

Those developing the shared data repository must decide what methods of data protection they want to use and they need to prepare access rules (e.g. in the form of use polices or licenses). Since industrial and agricultural data types are not sensitive within the meaning of the GDPR, and their protection is most often based on trade secrets, we recommend the use of a federated (distributed) data architecture, potentially using federated machine learning in edge processing, and the use of blockchain technology to track use. The basic questions, however, are who can give consent and whether human verification of users is required. If not, consent may be inferred automatically based on an ID or on acceptance of the license (e.g. using the gateway in the API). If consent is required each time, the relevant process must be included in the pilot plans.

4. Developing a business model with mutual benefits

Since the participants in the virtual shared data repository are business entities, the business model must provide for clearly defined common benefits. This is why it seems to be a good idea to provide positive incentives for data sharing by offering a promise in profits derived from manufactured products or services (e.g. applications based on data made accessible by a group of interested farmers). As an example of incentives, the workshop participants proposed showcasing samples of agricultural products obtained, or new innovative technical infrastructure built, based on data provided via the shared data repository. The ability to cover own costs and a convincing offer will increase the likelihood of new members joining and developing the project beyond the pilot.

5. Infrastructure preparation and assignment of responsibility for operation

The last stage before launching the pilot is the preparation of the technical infrastructure (both clouds on the back-end and access interfaces on the front-end) and appointment of entities responsible for services (such as administration or technical assistance) and for costs at the first stage. Not all entities will be able to contribute to the shared data repository from the start, so it will be beneficial to identify technical leaders at this stage.

6.2. Pilot of the virtual shared repository of health data

The process of digitization of the health care system, which accelerated in recent years, has contributed to the mass production of health data sets. According to statistics kept by the e-Health Center, the number of accounts in the e-health system has increased from approx. 600 thousand to more than 10 million since 2019 (Torchała, 2021). The important role of digitization of the public medical sector and the introduction of adequate changes is also evidenced by the “National Transformation Plan for 2022-2026” developed by the Ministry of Health, the authors of which wrote about the need to implement appropriate tools supporting the analysis of the patient’s health, the development of artificial intelligence algorithms and the construction of a central repository of medical data, as well as further digitization of medical records and building an ecosystem for their exchange (Kościelniak, 2021). At the same time, more than half of Poles use private facilities where the role of investing in the security of IT infrastructure and collecting patient data in such a way that it is organized and fit for further use is also emphasized (Pawlak, 2021). There is also an increase in interest in digital tools designed, among others, to independently monitor one’s own health, control the course of treatment or accelerate the process of providing medical advice. According to the report on “162 mobile health applications” developed by the National Health Care System, a few years ago there were over 260,000 mobile health applications on the market and the number of their downloads was almost 3 billion (Patient’s Voice, 2017).

Despite the increasing use of electronic databases to store health information, these systems (public, private and related to the functioning of health applications) are not prepared to communicate with one another and to share resources. It is therefore necessary to ensure strong interoperability between systems by establishing uniform standards for the collection of data and the selection of compatible systems for its storage. At the same time, given the particularly sensitive nature of health data, the progressive building of public trust by ensuring the co-management of resources and the possibility to decide on their use in the public interest should be an indispensable element in the creation of sharing mechanisms. Only in this way, through simultaneous regulatory and social initiatives, it will become possible to effectively use the potential of health data for the common good.

During the workshops we identified 10 key success factors that determine the main challenges to be solved during the pilot. Attention is also drawn to factors that in the course of the workshops significantly gained / lost on importance in the eyes of the stakeholders.

Fig. 9: 10 key success factors for the shared repository of health data

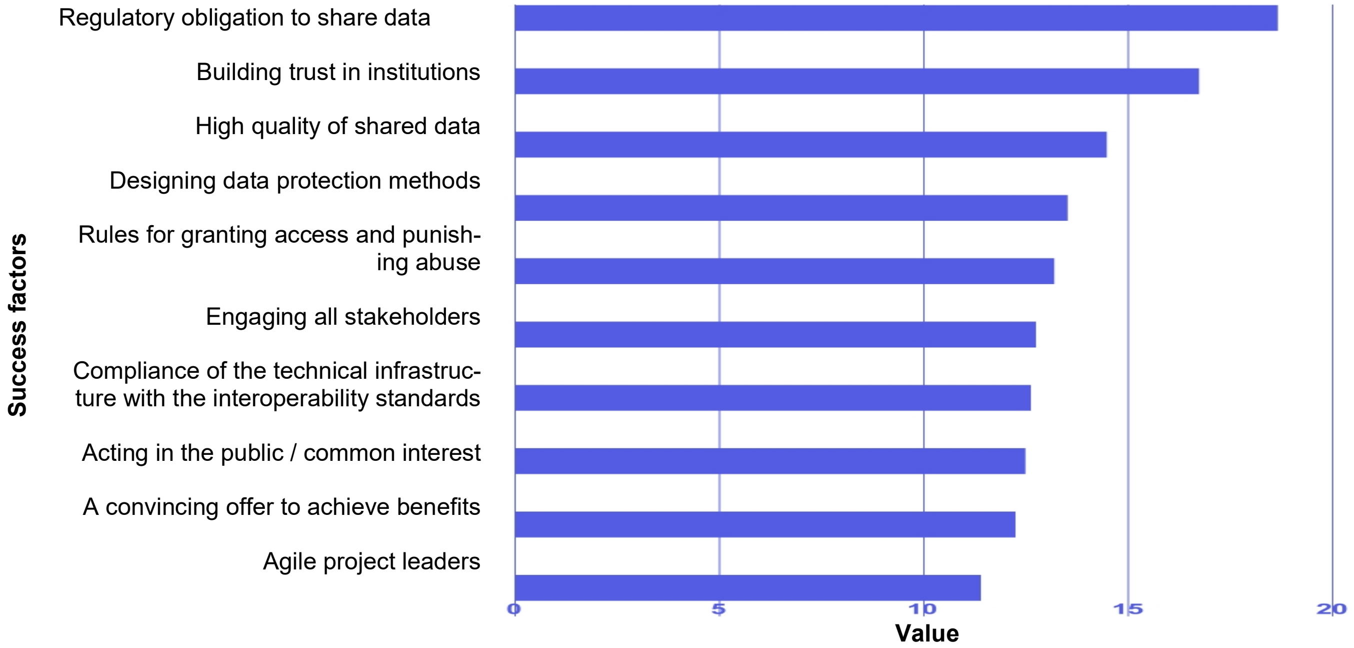
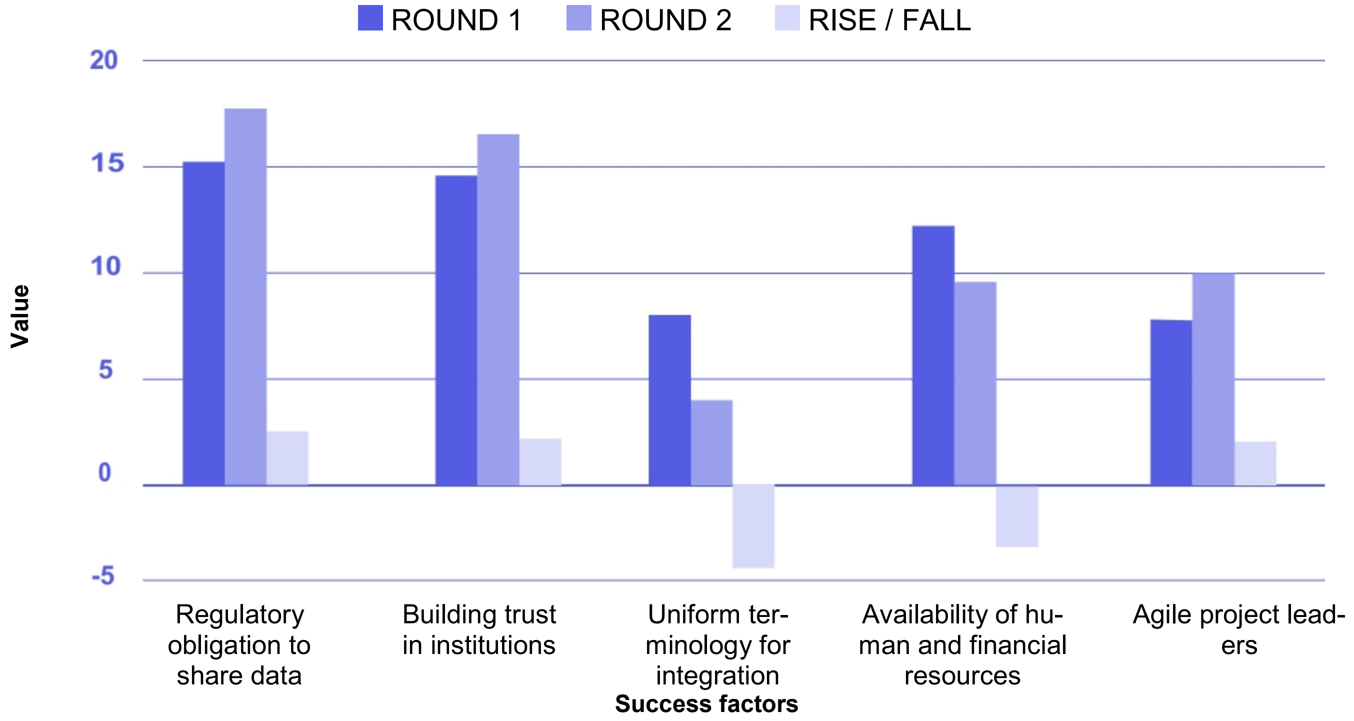


Fig. 10: 5 factors that have changed the most during the workshops



WE PROPOSE TO BASE THE PILOT ON THE FOLLOWING STEPS:

1. Designation of pilot datasets and a sharing standard

Since quality and volume of data is fundamental to the success of the project, the creators of the shared data repository must identify specific data sets that are available, processed and have features that make them relatively simpler to share than sets requiring intensive cleanup or patching. In particular, such data sets include medical images, electronic health records, blood counts, data from diabetic devices, data from pulse measuring devices. It is worth remembering that the adoption of an appropriate law introducing an opt-out system and/or ordering B2G sharing may increase the range of available collections. Further, a standard for sharing this data should be defined so as to allow for interoperability of the systems.

2. Establishing democratic data co-management mechanisms

The public nature of shared data repositories requires maximum inclusivity in the process of creating them. The reluctance to the idea of sharing information is due to the lack of a previous practice of data sharing and a strong concern about the possible loss of privacy on the Web. The most important part of shared data repositories is the community that shares its data for the sake of achieving a multidimensional benefit. Building social trust should become one of the most important elements of ethical design of shared data repositories. Since it is to serve the common good, all parties involved in its creation should be involved in the discussion on the principles of functioning of the shared data repository, already at the design stage. For this reason, the role of the creators should focus on coordinating the process of establishing social supervisory boards or other forms of democratic control (e.g. citizens' panels equipped with decision-making tools). They should make decisions regarding the functioning of shared data repositories – in the name of the principle of "from decision-maker to decision-taker", from the lowest levels – which will build public trust, increase transparency of the process and, from the beginning, involve stakeholders in further development of the project. Representativeness can be ensured by combining key stakeholders (Patient Ombudsman, health care workers' self-governments, patient organizations, legal and digital economy organizations) with an element of random representation. We recommend that the creation of such democratic oversight be a step ahead of further binding decisions in the pilot. Given also the potential of international data sharing practices and Polish participation in multilateral partnerships and global organizations, the role of the state should be to support stakeholders in creating new cross-border links and building international cooperation in the safe and innovative sharing of data.

3. Selection of a legal formula, objectives and strategy of the shared data repository through public consultations

The recommended legal formula is a co-managed public institution (most likely the eHealth Center) or a co-managed scientific center (a state research institute or a university with a high level of technological competence). However, building trust in data sharing will be successful only if the choice of the form of the shared data repository, goals, values, methods of achieving them and the expected results is made in a dialogue of the most open nature. Although this argument may sound repetitive, in the case of sensitive data it cannot be overestimated. The consultation should start from the democratic supervision of the shared data repository, but also take into account future sharing participants (apart from pilot data sets), potential data users (science, healthcare, innovation) and patients. The authors even recommend conducting research in this regard, such as a repetition of the DCE study for health data sharing (Johansson et al., 2021), followed by an information campaign to ensure that the implementation of the pilot will be perceived by society as an element of positive social progress. Social communication should include as many technical examples as possible to demonstrate security and technological resilience of shared data repositories against possible data leaks. In addition, thanks to the organization of workshops explaining in a clear and accessible way the functioning of shared data repositories and their business model, it will be possible to gradually restore public trust in data sharing, as well as to break the dichotomy of the public interest and the interest of the individual, perceived as two completely opposing interests.

4. Development of data protection methods, access rules and a business model serving the public interest

An important task in the process of creating the shared data repository will be to clearly define the possibilities of using data of Polish users and patients collected in public shared repositories of health data. The allocation of value in this model and the issue of commercialization and public sharing of research results require determination. It will also be crucial to determine whether the shared data repository's resources will be accessible to entities outside Poland or the EU and, if so, on what terms. However, the most important element seems to be the development of appropriate data protection methods. In this respect, the authors recommend solutions such as homomorphic encryption, storage of synthetic data, or the use of specialized data loss prevention systems.

→ See section 5.4.2

5. Infrastructure preparation and selection and training of project leaders

To guarantee full operability of the shared data repository, it is necessary to design and implement a trustworthy technical infrastructure. In this respect, it is recommended to develop IT systems for shared data repositories in cooperation with European and national IT solution providers. Infrastructure building tasks could be outsourced through public procurement procedures – in the form of a competition or a restricted tender. This approach would allow selection of the contractor on a fair and competitive basis, thanks to which it would be possible to provide the highest quality technology. Restricting the tender only to offers submitted by EU entities would make it possible to guarantee data security and (thanks to the independence of these entities from third-country corporations) the sovereignty of the systems.

In addition, the process of building shared data repositories should be accompanied by recruitment of competent project leaders. It is advisable that they should be people with not only technical knowledge in the field of infrastructure but also having highly educated soft skills such as creativity, team management, multitasking. These coordinators could be selected through an open call for innovation leaders (GovTech) profiled to search for specialists with adequate experience.

6.3. Selected recommendations for the state

Investments in a long-term digital upskilling program

In order to fully exploit the potential of the shared data repository, it is necessary to equip its future users with appropriate digital competences. Therefore, the state should invest in long-term activities aimed at improving quality of IT teaching at every stage of education. Primary education institutions should develop curricula that meet the needs of the digitized world. In addition, it would be a good idea to integrate digitization into vocational education and to provide specific training programs dedicated to specific sectors (e.g. agri-food, energy, medical). These programs could focus not only on basic knowledge and practical understanding of digital tools, but also on promoting innovative solutions in institutions or private companies.

Regulations for data standardization and strong interoperability

Since the shared data repository is a completely new project, awareness and new habits need to be built in the sectors specific to the pilots, starting with a sanctioned obligation to use established formats. Thus, it is recommended to introduce a regulatory obligation to place data in the shared data repository. At the same time, it would be advisable to promote RegTech tools for effective enforcement of applicable regulations through the use of new technologies verifying compliance with regulatory requirements and compliance in individual industries. In addition, it is necessary to establish appropriate standards. An entity designated by the government, acting as a separate institution (e.g. an executive agency) could be responsible for determining sector-specific formats, certifying those who operate or join the shared data repository and enforcing the shared data repository rules. Consideration should also be given to isolating and classifying domain-specific data representations and sector-specific standards (e.g. health or manufacturing industry) (European Commission, 2019).

Joint purchases of certified software

To strengthen security of data going to the shared data repository, it is worth making systems resistant to possible cyber attacks by selecting the best possible software for individual public institutions in a joint tender. Thanks to the uniform valuation of the desired services, entities using critical infrastructure (i.e. hospitals, the energy sector, the agri-food industry) will be able to adopt a single, coherent purchasing policy.

Act on the re-use of health data in the public interest

Following the example of Finland, we recommend the development of a law enabling the re-use of health data for the purposes of developing health care on an opt-out basis (default re-use). Due to the already existing organizational and technical potential of the e-Health Center, outsourcing the task of creating shared data repositories within universities or some institutes could inhibit the process of data sharing.* However, it is worth to consider expanding the list of entities to which documentation can be made accessible based on the provisions contained in the act and establishing rules based on which such access would be granted for the benefit of the public interest. Importantly, due to the unclear nature of “scientific purposes”, it is recommended to define this concept and expand the catalogue of possibilities for using health data also for other legitimate purposes in the public interest, also for private institutes.

* In accordance with art. 26(4) of the act of Nov. 6, 2008, on the rights of the patient and the Patient Ombudsman, medical documentation may also be disclosed to a university or research institute for use for scientific purposes, without revealing names and other data enabling the identification of persons to whom the documentation relates.

Implementing a uniform definition of health data

In addition to data that is clearly related to health (e.g. information from the Electronic Medical Records), there is a number of data types that, in combination with others, can form the basis for drawing conclusions about someone's health condition. This difficulty in demarcation applies in particular to fitbit bands and sports applications, which can measure not only a person's fitness but also their heart rate or sleep quality. There are three options for deciding whether specific data will be subject to the GDPR's regime of data of a special nature:

- Contextual approach: The context in which the data operates, i.e. the interests, conditions and consequences of the processing, are taken into account.
- Teleological approach: Focused only on clearly defined processing intentions that the data controller had.
- Mixed approach: The objective is of paramount importance but if the aim was not to draw conclusions of a sensitive nature, the context must still be taken into account.

In the context of data collection and processing in the shared data repository, it is recommended to adopt the so-called "mixed approach" that provides the broadest protection for the data used.

Carefully selected sources of financing

Undoubtedly, significant financial outlays are necessary to build technical infrastructures and maintain specialized staff capable of handling innovative solutions. It is proposed that within the framework of public funds allocated to individual ministries, an appropriate mechanism should be developed to provide financing (e.g. by way of a competition, public procurement, technological credit) intended for the erection of the data exchange space dedicated to individual areas. In addition, crowdfunding – popular in Western countries, consisting in engaging private investors in specific projects – could be a good source of funds for building shared data repositories (e.g. for business data). However, this would require an adequate regulatory environment conducive to investment crowdfunding (Council of Ministers, 2019).

7. Conclusion

It is necessary to make appropriate changes consisting in decommodification of data and treating them as a common good managed on behalf of the owners by non-commercial and non-profit public institutions. In the world of advancing digitization, data is our most valuable common resource. The overarching goal should be to create a shared and public value in such a way as to support the holistic development of society by sharing data. We need to use our resources more consciously, without letting them be wasted by being confined to the silos of organizations or technology providers. However, due to the different degrees of sensitivity of such data, it is necessary to pay attention not only to possible models but also to what data we are dealing with in a particular case – such information, combined with knowledge of the management model, may allow selection of the most advantageous sharing model.

The most important task for the development of Poland is to part with the distorted model of cognitive capitalism – that is, the new accumulation regime occurring in today's economy, in which the allocation of value is no longer based on physical labor and the machine system but on the exploitation of knowledge and creativity of people acting as free employees of technology corporations (Zygmuntowski, 2020b). Similarly, institutions should be consistently built to allow the paradigm of the information society as a source of information to be abandoned in favor of building a knowledge society. In the world around us, both private and professional activities are based on cooperation with intelligent autonomous machines. Thus, without building a learning society capable of adapting to new conditions, it will be impossible to implement innovative solutions or rebuild organizations (both private and state) (Ministry of Digitization, 2019).

Therefore, in order to be able to talk about any "sharing" at all, it is necessary to involve entities and experts from as many sectors as possible – so as to lead to the development of the most effective solutions for all interested parties through conversation.

References

Alemanno A. (2018) Big Data for Good: Unlocking Privately-Held Data to the Benefit of the Many [online], European Journal of Risk REgulation, [accessed on 11.05.2022], https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/big-data-for-good-unlocking-privatelyheld-data-to-the-benefit-of-the-many/C739E1DE223088FD3D761466DCDA2EFE

Artyushina, A. (2021) The future of data trusts and the global race to dominate AI [online], Bennett Institute For Public Policy, [accessed on 08.04.2022], https://www.bennettinstitute.cam.ac.uk/blog/data-trusts1/

Baron, J., Contreras, J. L., Husovec, M., Larouche, P., Thumm, N. (2019) Making the Rules: The Governance of Standard Development Organisations and their Policies on Intellectual Property Rights, JRC Science for Policy Report, EUR 29655 EN (March 2019); ISBN 978-92- 76-00023-5 , University of Utah College of Law Research Paper No. 308, TILEC Discussion Paper No. 2019-021, [accessed on 08.04.2022], https://ssrn.com/abstract=3364722

Bartol, A., Herbst, J., Pierścińska, A., (2021), Wykluczenie społeczno-cyfrowe w Polsce 2021.

Bayamlioglu, E. (2021) Data cooperative: a new intermediary on the horizon [online], KU Leuven, [accessed on 07.04.2022], https://www.law.kuleuven.be/citip/blog/data-cooperative-a-new-intermediary-on-the-horizon/

Bechtel, M., Buchholz, S., Deloitte (2022), Tech Trends 2022 [accessed on 27.05.2022],https://www2.deloitte.com/content/dam/Deloitte/pt/Documents/tech-trends/tech-trends-2022/DI_Tech-trends-2022.pdf

Big Data Value Association (2019) Towards a European Data Sharing Space. Enabling data exchange and unlocking AI potential, BDVA Position Paper, April 2019

Bitdefender (2022) Co to jest szyfrowanie danych i kiedy warto je stosować? [online], Bitdefender, [accessed on 27.05.2022], https://bitdefender.pl/co-to-jest-szyfrowanie-danych-i-kiedy-warto-je-stosowac/

Borowik, M., Maśniak, L., Kroplewski, R., Romaniec, H. (2017) Przemysł + Gospodarka oparta o dane, Ministerstwo Cyfryzacji, [accessed on 14.05.2022], Gospodarka oparta o dane – Gov.plhttps://www.gov.pl › documents › Gospodarka+O…

Bożykowski, M., Chłoń-Domińczak A., Jasiński, M., Zając, T. (2019) Dane publiczne – nowy impuls do rozwoju Polski, Polski Instytut Ekonomiczny, Policy Paper 8/2019

Data Collaboratives (2021) Data Collaboratives [online], GovLab, [accessed on 08.04.2022], https://datacollaboratives.org/

Datatilsynet (2022) Guidance on the use of the cloud, Datatilsynet, March 2022

Data Trust Initiative (2021) Data trusts: international perspectives on the development of data institutions, DTA, Working Paper 2

De Groot, J. (2020) What is Data Loss Prevention (DLP)? A Definition of Data Loss Prevention [online], Datainsider: Digital Guardian's Blog, [accessed on 30.05.2022], https:// digitalguardian.com/blog/what-data-loss-prevention-dlp-definition-data-loss-prevention

Delacroix, S., Lawrence, N. D. (2019) Bottom-up data Trusts: distributing the 'one size fits all' approach to data governane, International Data Privacy Law, Volume 9, Issue 4, 236-252

Domeyer, A., Hieronimus, S., Klier, J., Weber, T. (2021) Government data management for the digital age [online], McKinsey & Company, [accessed on 07.04.2022], https://www. mckinsey.com/industries/public-and-social-sector/our-insights/government-data-management-for-the-digital-age

Dymek, D., Komnata, W., Kotulski, L., Federacyjna hurtownia danych w dostępie do informacji poufnej, Akademia Górniczo-Hutnicza w Krakowie, dostęp: http://rocznikikae.sgh.waw.pl/p/ roczniki_kae_z33_08.pdf

Edelman Trust Barometer (2022) Wyniki najnowszego badania zaufania Edelman Trust Barometer 2022 [online], publicrelations.pl, [accessed on 15.05.2022], https://publicrelations.pl/wyniki-najnowszego-badania-zaufania-edelman-trust-barometer-2022/

Empirica (2022) MonitorEHR [online], Empirica, [accessed on 12.05.2022], https:// empirica.com/project/details/?projectid=291

ENISA (2022) Data Protection Engineering [online], ENISA, [accessed on 27.05.2022], https://www.enisa.europa.eu/publications/data-protection-engineering

Ernst & Young (2021) 57% polskich firm przyspieszyło transformację cyfrową w czasie pandemii, a jeden na pięciu uważa, że w ich firmach transformacja jest zaawansowana [online], EY Polska, [accessed on 14.05.2022], https://www.ey.com/pl_pl/news/2021/03/bada- nie-ey-transformacja-cyfrowa

European Data Protection Board (2020) Wytyczne w sprawie przetwarzania danych dotyczących zdrowia do celów badań naukowych w kontekście pandemii COVID-19 [online], EDPB, [accessed on 14.05.2022], https://edpb.europa.eu/sites/default/files/files/file1/ edpb_guidelines_202003_healthdatascientificresearchcovid19_pl.pdf

Fajgielski, P. (2021) Komentarz do rozporządzenia nr 2016/679 w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy

95/46/WE (ogólne rozporządzenie o ochronie danych), [w:] Ogólne rozporządzenie o ochronie danych. Ustawa o ochronie danych osobowych. Komentarz, wyd. II

Foroohar R., (2019) Don't Be Evil: How Big Tech Betrayed Its Founding Principles – – and All of Us, Penguin Books Ltd, ISBN 13: 9781984824004

The Global Partnership on Artificial Intelligence (GPAI) (2021) Enabling Data Sharing for Social Benefit Through Data Trusts: Data Trusts in Climate, [accessed on 27.05.2022], https://gpai.ai/projects/data-governance/data-trusts-in-climate-interim-report.pdf

Głos Pacjenta, (2017) Aplikacje mobilne szturmują rynek zdrowia [online], Głos Pacjenta, [accessed on 31.05.2022], Dostęp w: https://glospacjenta.pl/aktywnosci/przydatne/305,aplikacje-mobilne-szturmuja-rynek-zdrowia

Goasduff, L. (2019) Top Trends on the Garnter Hype Cycle for Artificial Intelligence, 2019 [online], Gartner, [accessed on 13.05.2022], https://www.gartner.com/smarterwith- gartner/top-trends-on-the-gartner-hype-cycle-for-artificial-intelligence-2019

GovTech Polska (2020) Polityka rozwoju AI w Polsce przyjęta przez Radę Ministrów – co dalej? [online], gov.pl, [accessed on 14.05.2022], https://www.gov.pl/web/govtech/ polityka-rozwoju-ai-w-polsce-przyjeta-przez-rade-ministrow--co-dalej

Grzeszak, J., Łukasik, K., Święcicki, I. (2021) Ile warte są nasze dane?, Polski Instytut Ekonomiczny, Warszawa

Hardjono, T., Pentland, S. (2018) Open Algorithms for Identity Federation, Proc IEEE Future of Information and Communication Conference, Singapur, Kwiecień 2018, https://arxiv.org/ pdf/1705.10880.pdf

Hardjono T., Pentland S., (2019) Data Cooperatives: Towards a Foundation for Decentralized Personal Data Management, MIT Connection Science, Massachusetts Institute of Technology

https://doi.org/10.48550/arXiv.1905.08819

Janssen, H., Singh, J. (2022) Data intermediary, Internet Policy Review, 11(1) https://doi. org/10.14763/2022.1.1644

Jemielniak, D., Przegalińska, A. (2020) Społeczeństwo współpracy, Wydawnictwo Naukowe Scholar, Warszawa, ISBN: 978-83-66470-04 – 0

Jessop, B. (2007) State Power: A Strategic-Relational Approach, Polity, Cambridge

Kaczmarek, A. (2022) Inżynieria ochrony danych wg ENISA [online], TKP, [accessed on 16.05.2022], https://www.traple.pl/2022/04/06/inzynieria-ochrony-danych-wg-enisa/

Kaplan, A., Haenlein, M. (2019) Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence, Business Horizons, Vol. 62 Issue 1, January – February 2019, 15-25

Kawalec, J. (2021) Sztuczna inteligencja – wyścig o naszą wolność [online], Pomorski Przegląd Gospodarczy, [accessed on 14.05.2022], https://ppg.ibngr.pl/pomorski-przeglad-gospodarczy/sztuczna-inteligencja-wyscig-o-nasza-wolnosc

Kerber, W., A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis (October 24, 2016). Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil (GRUR Int), 11/2016, 989-999

B. Kiełbasa, J. Puchała, Innowacyjność młodych rolników i ich postawy wobec zmian na przykładzie gospodarstw rolnych położonych w regionie rozdrobnionego rolnictwa, „Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu” 2015, t. 17, z. 1, s. 107-111.

Koloch G., Grobelna K., Zakrzewska-Szlichtyng K., Kamiński B., Kaszyński D. (2017). Intensywność wykorzystania danych w gospodarce a jej rozwój. Analiza diagnostyczna. [accessed on 07.06.2022], https: /mc.bip.gov.pl/rok-2017/analiza-diagnostyczna-intesywnosc-wykorzystania-danych-w-gospodarce-a-jejrozwoj.html

Komisja Europejska (2018) Staff Working Document – Guidance on sharing private sector data in the European data economy [online], Komisja Europejska, [Dostęp: 13.05.2022], https: /digital-strategy.ec.europa.eu/en/news/staff-working-document-guidance-sharing-private-sector-data-european-data-economy

Komisja Europejska (2020a) Europejska strategia w zakresie danych [online], Komisja Europejska, [Dostęp: 11.05.2022], https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_pl

Komisja Europejska (2020b) Rozporządzenie Parlamentu Europejskiego i Rady 2020/0340 z dnia 25 listopada 2020 r w sprawie europejskiego zarządzania danymi (akt w sprawie zarządzania danymi) stanowiące uzupełnienie Dyrektywy Parlamentu Europejskiego i Rady (UE) 2019/1024 z dnia 20 czerwca 2019 r. w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego

Komisja Europejska (2020c) Horizon 2020, Work Programme 2018-2020 Information and Communication Technologies, European Commission Decision C(2020)4029, 17 czerwca 2020

Komisja Europejska (2021) The Digital Economy and Society Index [online], Komisja Europejska, [Dostęp: 12.05.2022], https://digital-strategy.ec.europa.eu/en/policies/desi

Komisja Europejska (2022) Cyfrowe dane i usługi dotyczące zdrowia – europejska przestrzeń danych dotyczących zdrowia [online], Komisja Europejska, [Dostęp: 11.05.2022], https://ec.europa.eu/info/law/better-regulation/haveyour-say/initiatives/12663-Cyfrowe-dane-i-us%C5%82ugi-dotyczace-zdrowia-europejska-przestrzen-danych-dotyczacych-zdrowia_pl

Komisja Europejska (2019) High-level Expert Group on AI; Policy and Investment Recommendations for Trustworthy AI, [online], Komisja Europejska, [Dostęp: 27.05.2022], https://www.europarl.europa.eu/italy/resource/static/files/import/intelligenza_artificiale_30_ aprile/ai-hleg_policy-and-investment-recommendations.pdf

Kościelniak P., (2021) E-zdrowie w planie transformacji na lata 2022-2026, [online], Info. eZdrowie, [dostęp: 31.05.2022], http://forumezdrowia.pl/info/aktualnosci/e-zdrowie-w-planie-transformacji-na-lata-2022-2026/

Kroplewski, R. (2020), Protokół z XXXVII posiedzenia Rady do Spraw Cyfryzacji, KPRM, [dostęp: 06.06.2022],Dostępny w: https://www.gov.pl/attachment/af46cdbe-5ead-44cf-a8ca-a- 91e9ac052a1

Kroplewski, R., (2021), W stronę społeczeństwa wiedzy, Przegląd Techniczny Gazeta Inżynierska [Dostęp: 05.05.2022], http://przeglad-techniczny.pl/artykuly?id=2874

Mayer-Schönberger V., Ramge T. (2022) Access Rules: Freeing Data from Big Tech for a Better Future, University of California Press; First edition (April 26, 2022) ISBN-13: 978-0520387737

Małobęcka-Szwast, I. (2021) Data Governance Act – o krok bliżej do łatwiejszego dzielenia się danymi [online], newtech.law, [Dostęp: 11.05.2022], https://newtech.law/pl/ data-governance-act-o-krok-blizej-do-latwiejszego-dzielenia-sie-danymi/

Małobęcka-Szwast, I. (2022) Projekt Aktu w sprawie danych (Data Act) – kolejne ułatwienia w zakresie dzielenia się danymi [online], newtech.aw [Dostęp: 11.05.2022], https://newtech.law/pl/projekt-aktu-w-sprawie-danych-data-act-kolejne-ulatwienia-w-zakresie-dzielenia-sie-danymi/

Mehta, S., Dawande, M., Mookerjee, V. (2021) Can data cooperatives sustain themselves? [online], LSE, [Dostęp: 14.05.2022], https://blogs.lse.ac.uk/businessreview/2021/08/02/can-data-cooperatives-sustain-themselves/

Mehta, S., Dawande, M., Mu, L. (2022) The key to designing sustainable data cooperatives [online], Światowe Forum Ekonomiczne, [Dostęp: 14.05.2022], https://www. weforum.org/agenda/2022/02/the-key-to-designing-sustainable-data-cooperatives/

Ministerstwo Cyfryzacji (2016) Program Otwierania Danych Publicznych, Załącznik do uchwały nr 107/2016 Rady Ministrów z dnia 20 września 2016 r.

Ministerstwo Cyfryzacji, Polityka Rozwoju Sztucznej Inteligencji w Polsce na lata 2019 – 2027; Godna Zaufania Sztuczna Inteligencja autonomia i konkurencja +PL, Ciesielski, M., Flakiewicz,P., Jarzewski, A., Kroszczyńska, E., Lubos, B., Podgórska, A., Pukaluk, M., Pytko, T., Romaniec, H., Wancio, A., Stefaniak, S., Zaczek, A. (2019), [Dostęp: 06.06.2022), https://www.gov.pl/attachment/0aa51cd5-b934-4bcb-8660-bfecb20ea2a9.

Ministerstwo Cyfryzacji (2019) Polityka Rozwoju Sztucznej Inteligencji w Polsce na lata 2019 – 2027, Godna Zaufania Sztuczna Inteligencja autonomia i konkurencja +PL [Dostęp 07.06.2022], https://www.gov.pl/web/cyfryzacja/konsultacje-spoleczne-pro- jektu-polityki-rozwoju-sztucznej-inteligencji-w-polsce-na lata-2019-2027

Ministry of Economy, Trade and Industry of Japan (2018) Contract Guidelines on Utilization of AI and Data” (on account of amendments to the Unfair Competition Prevention Act of 2018

Ministerstwo Gospodarki (2013) Strategia innowacyjności i efektywności gospodarki “Dynamiczna Polska 2020”, Załącznik do uchwały nr 7 Rady Ministrów z dnia 15 stycznia 2013 r.

Ministerstwo Rozwoju, Diagnoza do Strategii Produktywności 2030 (2020) [Dostęp: 07.06.2022], https: /www.gov.pl/attachment/65c9d9ab-57e2-44dd-bf09-0ce- a82426ccf

Minister Zdrowia (2022) Odpowiedź na interpelację nr 3092 Posła Roberta Kwiatkowskiego w sprawie dostępu do informacji dotyczących cyfryzacji służby zdrowia [online], Ministerstwo Zdrowia, [Dostęp: 15.05.2022], https://interpelacje.sejm.gov.pl/interpelacje9. nsf/0/E8019F514173502EC12587EC0040E718/%24File/ODP_K9INT30932.pdf

Miniszewski, M. (2021), Dwie dekady rozwoju polskiego rolnictwa. Innowacyjność sektora rolnego w XXI wieku, Kutwa, K. (współpr.), Polski Instytut Ekonomiczny, Warszawa.

Nagel, L., Lycklama D. (2021) Design Principles for Data Spaces. Position Paper. Version 1.0. Berlin

Najbuk, P., Pachocki, J., Kruczyk-Gonciarz, A., Kaźmierczyk, P. Lorent, R. (2020). Wykorzystanie danych medycznych w celu rozwoju AI w Polsce i w celu prowadzenia badań naukowych. Raport Regulacyjny, DZP, Warszawa

Nayyar, A., & Puri, V. (2016, September) Smart farming: IoT based smart sensors agriculture stick for live temperature and moisture monitoring using Arduino, cloud computing & solar technology. In Proc. of The International Conference on Communication and Computing Systems (ICCCS-2016) (pp. 9781315364094-121). [Dostęp:14.05.2022], https: /www.researchgate.net/profile/Anand-Nayyar/publication/313804002_Smart_farming_IoT_based_smart_sensors_agriculture_stick_for_live_temperature_and_moisture_monitoring_using_Arduino_cloud_computing_solar_technology/links/59d9f67c0f7e9b12b36d66f8/ Smart-farming-IoT-based-smart-sensors-agriculture-stick-for-live-temperature-and-moisture-monitoring-using-Arduino-cloud-computing-solar-technology.pdf.

Nowoczesna Polska, Lekcja – Cyfrowy świat [online], Edukacja medialna, [Dostęp: 10.05.2022], https://edukacjamedialna.edu.pl/lekcje/cyfrowy-slad/

OECD (2019) Recommendation of the Council on Artificial Intelligence [online], OECD Legal Instruments, [Dostęp: 7.06.2022], https: /legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OVH (2018) Chmura prywatna ochroni dane wrażliwe [online], virtual-it.pl, [Dostęp: 27.05.2022], https://www.virtual-it.pl/8436-chmura-prywatna-ochroni-dane-wrazliwe.html

PAP (2021) Rejestr ciąż. Ministerstwo Zdrowia: "Chodzi o względy medyczne" [online], Dziennik Gazeta Prawna, [Dostęp: 14.05.2022], https://serwisy.gazetaprawna.pl/zdrowie/artykuly/8299619,rejestr-ciaz-ministerstwo-zdrowia-chodzi-o-wzgledy-medyczne.html

Paszcza, B. (2022) Dwa wielkie wyzwania e-gospodarki: kontrola nad danymi i legislacja interoperacyjności [online], Klub Jagielloński, [Dostęp: 13.05.2022], https:// klubjagiellonski.pl/2022/04/22/dwa-wielkie-wyzwania-e-gospodarki-kontrola-nad-danymi-i-legislacja-interoperacyjnosci/

Pawlak, M. 2021. Już prawie połowa Polaków leczy się prywatnie [online], Rzeczpospolita, [Dostęp: 31.05.2022], https://pieniadze.rp.pl/ubezpieczenia-zycia/art18940831-juz-prawie-polowa-polakow-leczy-sie-prywatnie

Petland, A., Hardjono, T. (2020) Data Cooperatives [online], Work in Progress MIT, [Dostęp: 15.05.2022], https://wip.mitpress.mit.edu/pub/pnxgvubq/release/2

Rada Unii Europejskiej i Rada Europejska (2021) EU looks to make data sharing easier: Council Agrees position on Data Governance Act [online], Rada UE i Rada Europejska, Press Release, [Dostęp: 14.05.2022], https://www.consilium.europa.eu/en/press/press-rele- ases/2021/10/01/eu-looks-to-make-data-sharing-easier-council-agrees-position-on-data-governance-act/

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych), OJ L 119, 4.5.2016, p. 1–88, art. 9.

Schneider, G., Health Data Pools under European Policy and Data Protection Law: Research as a New Efficiency Defence?, 11 (2020) JIPITEC 49 para 1.

Schubert, S., Harari Dayan, F. (2020) When is data pooling anticompetitive? [online], Freshfields Bruckhaus Deringer, [Dostęp: 13.05.2022], https://technologyquotient. freshfields.com/post/102glxx/when-is-data-pooling-anticompetitive

Swant. M (2019), People Are Becoming More Reluctant To Share Personal Data, Survey Reveals [online], Forbes, [Dostęp: 10.05.2022], https://www.forbes.com/sites/marty- swant/2019/08/15/people-are-becoming-more-reluctant-to-share-personal-data-survey-reveals/?sh=66b3889b1ed1

The Ministry of Electronics & Information Technology, Government of India (2020) Report by the Committee of Experts on Non-Personal Data Governance Framework, 111972/2020/ CL&ES

Torchała, K. 2021. Internetowe konto pacjenta posiada już ponad 10 mln osób [online], Bankier.pl, [Dostęp: 31.05.2022], https://www.bankier.pl/wiadomosc/Internetowe-Konto-Pacjenta-posiada-juz-ponad-10-mln-osob-8150520.html

Uchwała nr 196 Rady Ministrów z dnia 28 grudnia 2020 r. w sprawie ustanowienia "Polityki dla rozwoju sztucznej inteligencji w Polsce od roku 2020"

Ustawa z dnia 29 sierpnia 1997 r. – Prawo bankowe, Dz.U. 1997 nr 140 poz. 939

Ustawa z dnia 6 listopada 2008 r. o prawach pacjenta i Rzeczniku Praw Pacjenta, Dz. U 2009 nr 52 poz. 417

Wawrzyniak, B., Iwanowski D. (2021). Cyfrowy monopol Nadużycia, których dopuszczają się największe korporacje technologiczne. Instrat Policy Paper 02/2021.

Wawrzyniak, B., Zygmuntowski, J. J., Lamański, F. (2020) Polska suwerenna cyfrowo. Regulacje na rzecz sprawiedliwej i konkurencyjnej gospodarki cyfrowej. Instrat Policy Paper 06/2020

Wojciechowski, M. (1998) Zalety i wady architektury rozproszonej wykorzystującej migawki tylko do odczytu, Materiały IV konferencji PLOUG, Zakopane

Van Hesteren, D., Van Knippenberg, L., Weyzen, R., Huyer, E., Cecconi, G. (2021), Open Data Maturity Report 2021, data.europa.eu, Publications Office of the European Union

Velotio Technologies, 2019, [Dostęp: 27.05.2022], https: /medium.com/velotio-perspectives/a-beginners-guide-to-edge-computing-6cfea853aa11

Verhlust, S., Young, A., Srinivasan, P. (2022) An Introduction to Data Cooperatives [online], GovLab, [Dostęp: 08.04.2022], https://datacollaboratives.org/introduction. html#section1

VMware (2017) Can private cloud be cheaper than public cloud? 41% said yes, and the survey reveals how. VMware, 451 Research, Advisory

Zygmuntowski, J. J. (2020a) Wspólnice danych: Alternatywny model zarządzania danymi, Raport projektu: SpołTech, Centrum Cyfrowe

Zygmuntowski, J. J. (2020b) Kapitalizm Sieci, Stowarzyszenie Rozruch, ISBN: 978-83-957- 6720-3

Zygmuntowski, J. J., Zoboli, L., Nemitz, P. F. (2021). Embedding European values in data governance: a case for public data commons. Internet Policy Review, 10(3). https://doi. org/10.14763/2021.3.1572

Żyrek, A. (2022) Big Data cz. 1, Big Data a prawo autorskie i ochrona sui generis baz danych [online], B&K, [Dostęp: 14.05.2022], https://bartakalinski.pl/artykuly/big-data-cz-i-big-data-a-prawo-autorskie-i-ochrona-sui-generis-baz-danych/

Enclosure to the report

List of all workshop participants

- Representative of Open Future
- Representative of the National Center for Agricultural Support
- Representative of the National Center for Agricultural Support
- Representative of the Instrat Foundation
- Representative of the National Center for Agricultural Support
- Representative of the National Center for Agricultural Support
- Representative of QuantLabs
- Representative of the e-Health Center / National Health Fund
- Representative of the National Center for Agricultural Support
- Representative of the Innovation Department of NCAS
- Representative of the Polish Economic Institute
- Representative of the National Center for Agricultural Support
- Representative of the National Center for Agricultural Support
- Representative of Jutromedical
- Representative of the Instrat Foundation
- Representative of the National Center for Agricultural Support
- Representative of Leon Kozminski University
- Representative of the Polish Economic Institute
- Representative of the Warsaw City Hall
- Representative of the Innovation Department of NCAS
- Representative of HTA
- Representative of COT Łukasiewicz
- Representative of the Instrat Foundation
- Representative of the National Center for Agricultural Support
- Representative of the National Center for Agricultural Support
- Representative of Aida Diagnostics
- Representative of the Ministry of Health
- Representative of alxd
- Representative of Zhiva
- Representative of the e-Health Center
- Representative of the Ministry of Development and Technology
- Representative of the Chancellery of the Prime Minister
- Representative of the Chancellery of the Prime Minister
- Representative of the Chancellery of the Prime Minister
- Representative of the Chancellery of the Prime Minister
- Representative of the Instrat Foundation

Speakers

- Representative of the Wroclaw University of Technology
- Representative of Findat
- Representative of the Open Data Institute
- Representative of OVHCloud